Graph Densification

Moritz Hardt* Nikhil Srivastava† Madhur Tulsiani‡

October 24, 2011

Abstract

We initiate a principled study of *graph densification*. Given a graph G the goal of graph densification is to come up with another graph H that has significantly more edges than G but nevertheless approximates G well with respect to some set of test functions. In this paper we focus on the case of cut and spectral approximations.

As it turns out graph densification exhibits rich connections to a set of interesting and sometimes seemingly unrelated questions in graph theory and metric embeddings. In particular we show the following results:

- A graph G has a multiplicative cut approximation with an asymptotically increased density if and only if it does *not* embed into ℓ_1 under a weak notion of embeddability. We demonstrate that all planar graphs as well as random geometric graphs possess such embeddings and thus do not have densifiers. On the other hand, expanders do have densifiers (namely, the complete graph) and as a result do not embed into ℓ_1 even under our weak notion of embedding.
- An analogous characterization is true for multiplicative spectral approximations where the embedding is into ℓ_2^2. Using this characterization we expose a surprisingly close connection between multiplicative spectral and multiplicative cut densifiers.
- We also consider additive cut and spectral approximations. We exhibit graphs that do not possess non-trivial additive densifiers.

Our results are mainly based on linear and semidefinite programs (and their duals) for computing the maximum weight densifier of a given graph. This also leads to efficient algorithms in the case of spectral densifiers and additive cut densifiers.

*Center for Computational Intractability, Department of Computer Science, Princeton University. Supported by NSF grants CCF-0426582 and CCF-0832797. Email: `mhardt@cs.princeton.edu`.

†Institute for Advanced Study. Supported by NSF grants CCF 0832797 and DMS 0835373 Email: `nikhils@math.ias.edu`.

‡Center for Computational Intractability, Department of Computer Science, Princeton University. Email: `tulsiani@cs.princeton.edu`.

1 Introduction

Understanding the structural differences between sparse graphs and dense graphs is a natural problem with many algorithmic applications. The complexity of various combinatorial optimization problems on graphs depends strongly on the density of the underlying graph.

On the one hand, several problems can be solved more efficiently on sparse graphs than on dense graphs. As a result, much effort has been devoted to graph sparsification. Broadly speaking, graph sparsification is the problem of removing as many edges from a graph as possible while preserving the structure of the graph. Benczur and Karger [BK96] introduced the notion of a *cut sparsifier*, that is, a sparse graph which preserves the fractional value of all cuts up a multiplicative approximation error. They showed that every graph has a cut sparsifiers with $O(n \log n)$ edges. Spielman and Teng [ST04] strengthened this notion to that of a *spectral sparsifier* by requiring that the eigenvalues of the graph Laplacian be preserved under sparsifcation. There has been much work on sparsification since, e.g., [SS08, BSS09, FHHP11].

On the other hand, dense graphs enjoy structural properties that in general do not hold for sparse graphs. For example, Max Cut has a PTAS on dense graphs owing to the existence of low width cut decompositions [FK96]—in general no PTAS exists for Max-Cut unless P = NP [PY91, ALM+98, Hås01]. In fact on a graph with n vertices and average degree D, Max Cut can be solved up to a small additive error in time $2^{O(n/D)}\mathrm{poly}(n)$ thus allowing for non-trivial improvements over brute-force on any graph of super-constant average degree.[1] Similarly, many property testing algorithms and sublinear time approximation algorithms rely on structural properties of dense graphs, e.g., [GGR98].

Any such theorem about dense graphs that depends only on the cut (or spectral) profile of the graph automatically extends to the possibly much larger class of graphs that are well approximated by dense graphs with respect to cuts (or spectral distinguishers). Surprisingly, very little is known about this class of graphs. In particular, the following questions are wide open:

- What is the class of *intrinsically sparse* graphs? Can we characterize which graphs do not possess dense approximations?
- Is there an inverse procedure to graph sparsification? Can we compute dense approximations *efficiently* when they exist?

In this work we address these questions under the umbrella term of *graph densification*. Graph densification is the natural inverse question of graph sparsification. Some striking differences between the two problems are immediate. While it was shown that every graph has a $(1+\varepsilon)$-multiplicative spectral sparsifier (and hence also a multiplicative cut sparsifier) with $O(n)$ edges [BSS09], we cannot hope that every graph has a densifier with $\Omega(n^2)$ edges as this would contradict the previously mentioned hardness results for Max Cut. In fact, we will later exhibit families of graphs with $m = O(n)$ edges that do not have any constant factor multiplicative cut densifier with $\omega(m)$ edges.

Understanding graph densification nevertheless leads to structural insight into graph sparsification as well. Indeed, an intrinsically sparse graph in the above sense can never be the output of any graph sparsification algorithm. Hence graph densification can lead to a better understanding of what sparsifiers of dense graphs may and may not look like.

[1] This follows from the work of [GGR98] using the quantitative improvements of [FS02, BHHS11].

1.1 Our results

We begin with the case of multiplicative cut approximations. We assume that all graphs are undirected. We say a graph H is a one-sided C-multiplicative cut approximation to G if the value of any cut in H (the fraction of edges that cross the cut) is at most C times the value of the same cut in G. We assume that both graphs are defined on the same vertex set. Note that here we only consider a one-sided approximation requirement. It turns out that the one-sided condition will already be enough for us to rule out densifiers in a strong sense.

We will show that G has a one-sided C-multiplicative cut approximation if and only if it does not admit a certain weak embedding into ℓ_1. We call this type of embedding (α, C)-humble. Informally, an (α, C)-humble embedding separates at least $(1-\alpha)n^2$ pairs of vertices (or non-edges) by a distance that is bounded away from C times the average distance taken over edges in G. This means that distances in the graph do not completely collapse under the embedding.

Informal Theorem 1.1. *A graph G has a one-sided C-multiplicative cut approximation with more than αn^2 edges if and only if G does not have an (α, C)-humble embedding into ℓ_1.*

To rule out densifiers with $\Omega(n^2)$ edges the previous theorem states that it is enough to exhibit a $(o(1), C)$-humble embedding.

Another way to think of a humble embedding is as a collection of non-overlapping sparse cuts. Indeed every ℓ_1-metric can be written as a non-negative combination of cut metrics. Since a humble embedding stretches many pairs signifcantly more than the average edge in the graph, it must be the sum of sparse cuts—separating many pairs while separating few edges. Viewed in this light, one noteworthy aspect of the theorem is that it asserts a kind of dichotomy for graphs: informally speaking, either a graph has a densifier or it contains many non-overlapping sparse cuts. This result is similar to the basic fact that either a graph is an expander or there must be a sparse cut. Indeed, expanders are well-approximated by the complete graph with respect to cuts and thus have a densifier.

As it turns out we are often able to give humble embeddings where $\alpha = O(1/n)$ thus ruling out densifiers with $\omega(n)$ edges. One class of graphs for which this is possible are planar graphs.

Informal Theorem 1.2. *Planar graphs have $(O(1/n), O(1))$-humble embeddings.*

While all planar graphs have $O(n)$ edges, we also present for all $m \in [O(n), o(n^2)]$ a family of graphs with m edges that do not have densifiers with $O(mn^{\varepsilon})$ edges These graphs are so-called *random geometric graphs* as studied, for example, by Feige and Schechtman [FS02]. Informally speaking, random geometric graphs are obtained by sampling sufficiently many points on a high-dimensional sphere uniformly at random and connecting points that are within a certain radius. Similar results are possible for other random geometric graph models [Pen03]. We omit the theorem statement from this overview.

Spectral approximations. We extend our characterization to spectral approximations of graphs, as well. A graph H is a one-sided C-multiplicative spectral approximation to a graph G if $x^T L_H x \leqslant C x^T L_G x$ for some constant C and all vectors $x \in \mathbb{R}^n$. Here, L_G (L_H) denotes the graph Laplacian of G (H). As in the case of cuts, here we only consider a one-sided condition. This is a strictly stronger condition as the cut approximation which corresponds to the case where $x \in \{0,1\}^n$.

Informal Theorem 1.3. *A graph G has a one-sided C-multiplicative spectral approximation with more than αn^2 edges if and only if G does not have an (α, C)-humble embedding into ℓ_2^2.*

We mainly use this characterization to exhibit a surprisingly close connection between one-side cut and spectral densifiers as stated next.

Informal Theorem 1.4. *If a graph G has a one-sided C-multiplicative densifier with m edges, then for every constant $\varepsilon > 0$, G also has a one-sided $(1+\varepsilon)C^2$-multiplicative spectral densifier with $\Omega_\varepsilon(m)$ edges.*

We believe this theorem makes a significant step towards proving such a connection for two-sided densifiers which we leave as an interesting open problem.

Based on semidefinite programming we can compute an optimal two-sided and one-sided spectral densifier efficiently. By Theorem 1.4, this also implies a non-trivial efficient approximation algorithm for the best one-sided C-multiplicative cut densifier—a task that might have appeared difficult a priori due to the 2^n cut constraints.

Expanders and spectral gap. We remark that expanders are well-approximated by the complete graph (even in the spectral multiplicative sense) and hence do have densifiers. By Theorem 1.1, this implies new non-embeddability results for expanders. Since our notion of embeddability can be seen to be weaker than the usual notion of having small distortion, this also implies a different proof of the fact that expanders do not embed into ℓ_1 with distortion $o(\log n)$ [LLR95].

One might suspect that the property of having a spectral multiplicative densifier is related to the spectral gap of a graph. However, it can be shown that there are graphs with a spectral gap of $O(1/n)$ that do have multiplicative spectral densifiers. One example we give is the dumbell graph, i.e., two expanders joined by a single edge.

Additive Approximations. While not the main focus of our paper, we also consider densifiers that preserve the fractional value of all cuts up to an additive error of ε. We observe that an optimal additive densifier of a given graph can be computed efficiently using the Alon-Naor SDP relaxation for the cut-norm as a separation oracle [AN04].

Informal Theorem 1.5. *We can compute an optimal additive cut densifier for a given graph G in polynomial time up to a small loss in the approximation error and negligible loss in density.*

We then show that the cycle does not have additive cut densifiers with $\omega(n)$ edges.

Informal Theorem 1.6. *Let $\varepsilon < 1$. Then, the cycle on n vertices does not have an ε-additive cut densifier with more than $2n$ edges.*

Our proof establishes the claim in two steps. We first rule out *spectral* additive densifiers using arguments specific to the fact that the cycle is a Cayley graph of the group $\mathbb{Z}/n\mathbb{Z}$. We then transfer this result to cut densifiers by showing that additive spectral and additive cut approximations are equivalent for vertex transitive graphs. This may be of independent interest.

Informal Theorem 1.7. *Let G be a vertex transitive graph. Then G has a spectral additive densifier if and only if G has an additive cut densifier.*

1.2 Proof overview

In Section 2, we illustrate the connection to ℓ_1-embeddings with a short combinatorial argument showing that if the graph metric of a bounded degree graph G admits a non-contractive embedding into ℓ_1 with constant average distortion, then G does not have a multiplicative cut densifier with a superlinear number of edges. For simplicity we present this argument only for unweighted graphs. We also give the analogous argument for spectral densifiers and ℓ_2^2-embeddings.

Before we can give a tight characterization, we need the notion of a weighted densifier. We will think of a graphs as specifying a *measure* over pairs of vertices $\{x_{uv}\}_{u,v\in V}$, i.e., $x_{uv} \in [0,1]$. The edge weight of a graph (which we seek to maximize) is then defined as $\sum_{u,v\in V} x_{u,v} \in [0, n^2]$.[2]

The basic approach. With the above definition of a weighted graph, we can express the optimization problem of finding an optimal densifier for a given graph as linear program (or semidefinite program in the case of spectral approximation). Indeed, given a graph $G = (V, E)$ the linear program $\max \sum_{u,v\in V} x_{u,v}$ subject to $x_{u,v} \in [0, 1]$ and the natural 2^n cut constraints can be expressed as a linear program. These 2^n cut constraints correspond to 2^n dual variables which we interpret as specifying an ℓ_1-embedding. Recall that every ℓ_1-metric ρ can be written as a non-negative combination of cut metrics $\rho = \sum_{S\subseteq[n]} \lambda_S \delta_S$.

Humble embeddings. After simplifying the exact dual program, we arrive at the notion of an (α, C)-humble embedding. Intuitively, a humble embedding asks for a non-trivial separation between edges and non-edges. Note that in a bounded degree graph most pairs of vertices are far apart according to the shortest path metric in the graph. A humble embedding promises that distances do not completely collapse in the embedding.

When constructing humble embeddings a useful building block are *sparse* cuts. A sparse cut has the property that it cuts many pairs (non-edges) while cutting only few edges in the graph. Intuitively, a humble embedding must consist of a collection of such sparse cuts which separates almost all pairs. Using this intuition we construct humble embeddings for planar graphs by recursively applying (a variant of) the planar separation theorem [LT80] which provides us with sufficiently sparse cuts. Similarly, for random geometric graphs we obtain a humble embedding by choosing a collection of random hyperplane cuts in the graph.

Connection between spectral and cut densifiers. The notion of a an (α, C)-humble embedding naturally extends to the spectral case using semidefinite programming duality. In this case a humble embedding is collection of vectors $z_1, \ldots, z_n$ such that $\|z_u - z_v\|^2$ is bounded away from $C \cdot \mathbb{E}_{(u,v)\in G} \|z_u - z_v\|^2$ for $(1-\alpha)n^2$ pairs. We call this a humble embedding into ℓ_2^2. To give a connection between the spectral and cut case we argue in the contrapositive: Given an (α, C^2)-humble embedding into ℓ_2^2 (ruling out spectral densifiers) we exhibit an (α, C)-humble embedding into ℓ_1 (ruling out cut densifiers). To do so we first pass from squared Euclidean norms to Euclidean norms while quantifying the loss (in terms of C) involved in this step. Once we have a separation between non-edges and edges in the Euclidean norm, we appeal to the fact that ℓ_2 embeds isometrically into ℓ_1. Hence, we obtain a separation in ℓ_1 which yields the desired humble embedding.

Altogether this implies that graphs which have non-trivial cut densifiers must also have spectral densifiers.

Additive Case. As in the multiplicative case, the optimization problem of finding a maximum weight densifier of a given graph can be phrased as an optimziation problem. Unlike in the multiplicative cut case, we can solve this problem efficiently. This follows from solving the linear

[2] This is view is equivalent to thinking of the graph as a distribution over pairs and taking the min-entropy of the distribution to be the measure of density.

program with an efficient separation oracle that uses the Alon-Naor SDP relaxation for the Cut-Norm [AN04].

The linear program also leads to a dual characterization. However, we were unable to give tight non-densification results using this characterization even for the cycle. Instead we focus on a specific example and give a direct proof that the n-cycle does not have any non-trivial spectral additive densifiers. The proof uses the fact that the cycle is a Cayley graph of the group $\mathbb{Z}/n\mathbb{Z}$. Moreover, we argue that a densifier has to be invariant under cyclic permutations and hence must assign the same weight to pairs at the same distance in the cycle. This allows us to directly compare the eigenvectors of the cycle with the eigenvectors of a densifier. We then transfer the result to additive cut densifier using an equivalence between spectral additive and spectral cut approximation for vertex transitive graphs. This result is almost a direct consequence of the Grothendieck factorization, see e.g. [Pis86, Tro09].

Organization. The paper is organized as follows. We state our definitions formally in Section 1.4. Then for the sake of exposition we present a simple combinatorial connection between densifiers and embeddings in Section 2. Our tight dual characterization and the notion of a humble embedding follows in Section 3. We present humble embeddings for various families of graphs in Section 4. Additive densifiers are discussed in Section 5. We conclude with open problems in Section 6.

1.3 Related work

We are not aware of any prior work that addresses the general question of graph densification especially in the multiplicative case. In the additive case, a closely related line of research—that inspired our work—revolves around the dense model theorem for graphs [RTTV08]. The general dense model theorem [GT08] asks broadly speaking which sparse structures have dense models. Reingold et al. [RTTV08] gave a dense model theorem for graphs which shows that if a graph G has constant density inside an expander graph, then G has an additive cut approximation with $\Omega(n^2)$ edges. Due to the loss in parameters, their theorem does not answer the question of which graphs have densifiers with, say, $\Omega(n^{1.99})$ edges.

The work of Coja-Oghlan et al. [COCF10] shows that a class of sparse graphs they call *bounded* is well approximated in the additive sense by constant width cut decompositions.[3] An analogous statement was shown previously for dense graphs by Frieze and Kannan [FK96]. Technically, the boundedness condition of [COCF10] involves two parameters that affect the width of the cut decomposition. We omit these parameters from this comparison for simplicity. It is, however, not hard to show that constant width cut decompositions for a graph G such as those constructed by [COCF10] also give an additive cut densifier for G with with $\Omega(n^2)$ edges.[4] An alternative algorithmic approach for obtaining cut decompositions for sparse graphs thus suggested by our work is to use the efficient algorithm for computing a maximum weight additive densifier (Theorem 1.5) and then apply the Frieze-Kannan decomposition to the resulting densifier. It follows from [COCF10] and the previous observation that this approach produces a low-width cut decomposition for bounded graphs.

[3] A cut decomposition of a graph G approximates the the adjacency matrix of G as a sum of so-called *cut matrices*, i.e., rank-1 matrices.

[4] This follows from the fact that the cut matrices used in [COCF10] correspond to large cuts and their coefficients are in $[0, 1]$.

1.4 Preliminaries and definitions

Weighted graphs and density. We identify a weighted graph G on a vertex set V with a set of weights $G = \{x_{uv}\}_{u,v \in V}$ such that $0 \leqslant x_{uv} \leqslant 1$ for all $u, v \in V$. We define the *edge weight* of a weighted graph as $m(G) \stackrel{\text{def}}{=} \sum_{u,v} x_{uv}$. We let $E_G(A,B) = \sum_{u\in A, v \in B} x_{uv}$ and define its normalized variant (called the normalized density of the cut) as

$$e_G(A,B) \stackrel{\text{def}}{=} \frac{E_G(A,B)}{m(G)} \,.$$

When G is unweighted $m(G)$ is just the number of edges and $e_G(A,b)$ is the fraction of edges with one endpoint in A and one in B.

Cut approximations A graph H is a *C-multiplicative cut approximation* of G if for every $S \subseteq V$,

$$\frac{1}{C} \leqslant \frac{e_G(S,S^c)}{e_H(S,S^c)} \leqslant C \,. \tag{1}$$

We define H to be an *ε-additive aditive approximation* of G if the normalized densities for every cut (S,S^c) are within an additive ε of each other, i.e., for every $S \subseteq V$,

$$|e_G(S,S^c) - e_H(S,S^c)| \;\leqslant\; \varepsilon \tag{2}$$

Equivalence with (S,T) cuts and $\infty \to 1$ norms. The notion of cut approximation used in some of the related work [RTTV08, COCF10] asks that $|e_G(S,T) - e_H(S,T)| \leqslant \varepsilon$ for all $S, T \subseteq [n]$. This differs from our notion in that we take $T = S^c$. The next lemma shows that this is without loss of generality up to a small loss in ε.

Lemma 1.1. *Let $\varepsilon \geqslant 0$. Suppose G, H are graphs such that for every $S \subseteq [n]$, $|e_G(S,S^c) - e_H(S,S^c)| \leqslant \varepsilon$. Then, for every $S, T \subseteq [n]$, $|e_G(S,T) - e_H(S,T)| \leqslant 6\varepsilon$.*

Proof. Note that for every *disjoint* S, T we have $e_G(S,T) = \frac{1}{2}(e_G(S,S^c) + e_G(T,T^c) - e_G(S \cup T, (S \cup T)^c))$, which gives $|e_G(S,T) - e_H(S,T)| \leqslant \frac{3}{2}\varepsilon$.

Next, consider the case when $S = T$. Note that if we randomly partition S into two equal and disjoint pieces S_1 and S_2, then we have $\mathbb{E}\, e_G(S_1,S_2) = \frac{1}{2} e_G(S,S)$. Thus,

$$|e_G(S,S) - e_H(S,S)| \;=\; 2 \cdot |\mathbb{E}\, e_G(S_1,S_2) - \mathbb{E}\, e_H(S_1,S_2)| \;\leqslant\; 2 \cdot \mathbb{E}|e_G(S_1,S_2) - e_H(S_1,S_2)| \;=\; 3\varepsilon \,.$$

Finally, for arbitrary S and T, we note that

$$e_G(S,T) \;=\; e_G(S \setminus T, T) + e_G(S \cap T, T \setminus S) + e_G(S \cap T, S \cap T) \,,$$

which completes the proof. ■

We will also use the well-known fact that the condition $|e_G(S,T) - e_H(S,T)| \leqslant \varepsilon$ for all $S, T \subset [n]$ is equivalent upto a constant factor of 4 in the error ε to

$$\left\| \frac{A_G}{m(G)} - \frac{A_H}{m(H)} \right\|_{\infty \to 1} \leqslant \varepsilon \qquad \text{where} \qquad \|A\|_{\infty \to 1} \stackrel{\text{def}}{=} \max_{x,y \in \{-1,1\}^n} x^T A y \tag{3}$$

and A_G is the adjacency matrix. This was mentioned, for instance, in [AN04]. It will be convenient to use the $\infty \to 1$ norm formulation in several places since it satisfies the triangle inequality.

Multiplicative spectral approximations. The notion of cut approximations is strengthened by that of spectral approximations. For a cut (S, S^c), consider a vector z with $z_u = 1$ if $u \in S$ and 0 otherwise. The quantity $e_G(S, S^c)$ can be written as

$$\frac{\sum_{u,v \in V} x_{uv} \cdot (z_u - z_v)^2}{m(G)} = \frac{z^T L_G z}{m(G)}$$

where L_G is the Laplacian matrix for the graph G with $(L_G)_{uv} = -x_{uv}$ if $u \neq v$ and $\sum_w x_{uw}$ if $u = v$. This can be thought of as the average length of an edge according to the vector z. Cut approximations require that this quantity be preserved for all vectors z with 0/1 coordinates. Spielman and Teng [ST04] strengthened this notion to that of *spectral approximations* which requires the above quantity to be preserved for all real-valued vectors z.

We say that a graph H is a (two-sided) *C-multiplicative spectral approximation* of a graph G if the above quantity is approximated within a factor C for any real-valued vector z i.e.

$$\frac{1}{C} \cdot \frac{z^T L_H z}{m(H)} \leqslant \frac{z^T L_G z}{m(G)} \leqslant C \cdot \frac{z^T L_H z}{m(H)}. \tag{4}$$

Additive Spectral Approximations. It will be convenient for us to define an *ε-additive spectral approximation* by the condition that

$$\left\| \frac{A_G}{m(G)} - \frac{A_H}{m(H)} \right\|_2 \leqslant \frac{\varepsilon}{n}, \tag{5}$$

where A_G denotes the adjacency matrix of G. A common alternative definition would require that

$$\left\| \frac{L_G}{m(G)} - \frac{L_H}{m(H)} \right\|_2 \leqslant \frac{\varepsilon}{n} \quad \Longleftrightarrow \quad \forall z,\ \|z\| \leqslant n\colon \ \left| \frac{z^T L_G z}{m(G)} - \frac{z^T L_H z}{m(H)} \right| \leqslant \varepsilon\,. \tag{6}$$

These definitions are equivalent for regular graphs. Using (5) instead of (6) is more standard in the literature [BN04, BL06] and will be easier to interpret and relate to its cut analogue (3).

Densifiers. We will freely use the term *densifier* in place of cut approximation or spectral approximation. For example, a C-multiplicative cut densifier of a graph G is simply a graph H that is a C-multiplicative cut approximation of G. We typically prefer to use the word *densifier* if we want to indicate that H has larger edge weight than G. Note that we do not require H to contain G.

Embeddings. For a set $S \subseteq V$, the *cut metric* δ_S is defined by $\delta_S(u, v) = 1$ if $|S \cap \{u, v\}| = 1$ and $\delta_S(u, v) = 0$ otherwise. We will consider embeddings of $G = (V, d_G)$ into ℓ_p (for $p = 1, 2$.) An embedding $f\colon G \hookrightarrow \ell_1$ induces a metric ρ on V that can be written as a non-negative linear combination of cut metrics, i.e., $\rho = \sum_S \lambda_S \delta_S$ where $\lambda_S \geqslant 0$.

2 Warm-up: combinatorial connection to embeddability

In this section we give short and direct proofs demonstrating that certain embeddings are basic obstructions for the existence of densifiers. While these characterizations are not tight, we find it instructive to present them first. We say an embedding $f\colon V \to \ell_p^q$ of a graph $G = (V, E)$ is *non-contractive* if for every $u, v \in V$ we have

$$d_G(u, v) \leqslant \|f(u) - f(v)\|_p^q\,, \tag{7}$$

where $d_G(u, v)$ denotes the shortest-path distance between u and v in the graph G.

2.1 Multiplicative cut densifiers and embeddings into ℓ_1

Below we rule out the existence of cut densifiers for graphs that embed into ℓ_1 with small average expansion. The *average expansion* of an embedding f of a unweighted graph $G = (V, E)$ into ℓ_1 is denoted by $\alpha_1(f) \stackrel{\text{def}}{=} \mathbb{E}_{(u,v)\in E} \|f(u) - f(v)\|_1$. We let $\alpha_1(G) \stackrel{\text{def}}{=} \min \alpha_1(f)$ where the minimum is taken over all non-contractive embeddings $f\colon V \to \ell_1$.

Theorem 2.1. *Let $G = (V, E)$ be an unweighted graph of maximum degree Δ and let $H = (V, E_H)$ be an unweighted C-multiplicative densifier of G. Then,*

$$|E_H| \leqslant O\left(\Delta^{2C\cdot\alpha_1(G)} \cdot n\right) .$$

In particular, if $\alpha_1(G)$ and Δ are constant, then G does not have a densifier with $\omega(n)$ edges which gives a constant multiplicative approximation.

Proof. Let $H = (V, E_H)$ be an unweighted (one-sided) C-multiplicative cut densifier for G Further, let $\rho = \sum_S \lambda_S \delta_S$ denote a non-contractive embedding of G that achieves $\alpha_1(G)$. We will bound the number of edges in H as follows. Indeed,

$$\begin{aligned}
\mathop{\mathbb{E}}_{(u,v)\in E_H} d_G(u,v) &\leqslant \mathop{\mathbb{E}}_{(u,v)\in E_H} \rho(u,v) && (\rho \text{ is non-contractive})\\
&= \mathop{\mathbb{E}}_{(u,v)\in E_H} \sum_S \lambda_S \delta_S(u,v)\\
&= \sum_S \lambda_S \mathop{\mathbb{E}}_{(u,v)\in E_H} \delta_S(u,v)\\
&= \sum_S \lambda_S e_H(S, S^c)\\
&\leqslant C \sum_S \lambda_S e_G(S, S^c) && (H \text{ is a cut densifier})\\
&= C \mathop{\mathbb{E}}_{(u,v)\in E} \rho(u,v)\\
&\leqslant C \cdot \alpha_1(G) . && (\text{average distortion of } \rho)
\end{aligned}$$

By Markov's inequality, at least half the edges (u, v) of H must satisfy $d_G(u, v) \leqslant 2C\alpha_1(G)$. But,

$$|\{v \in V : d_G(u,v) \leqslant 2k\}| \leqslant 1 + \Delta + \Delta(\Delta-1) + \cdots + \Delta(\Delta-1)^{2k-1} \leqslant \Delta^{2k} + 1 .$$

Hence, $|E_H| \leqslant 2(\Delta^{2C\alpha_1(G)} + 1)n$ which is what we wanted to show. ∎

We can draw some immediate conclusions from the previous theorem.

Expander graphs. It is not difficult to show that the complete graph is a constant factor multiplicative cut approximation of any expander graph. For simplicity we take this to be the definition of an expander graph. We say that G is an expander graph if it is a (two-sided) $O(1)$-multiplicative cut approximation of the complete graph.

Corollary 2.2. *Every bounded degree expander graph G must have $\alpha_1(G) \geqslant \Omega(\log n)$.*

Planar graphs. It is known that planar graphs embed into ℓ_1 with distortion $O(\sqrt{\log n})$ under the usual notion of distortion [Rao99]. Since this is a stronger notion than the one above this rules out certain densifiers.

Corollary 2.3. *A bounded degree planar graph does not have a constant factor multiplicative cut densifier with* $\omega\left(n2^{\sqrt{\log n}}\right)$ *edges.*

The planar embedding conjecture states that every planar graph in fact embeds into ℓ_1 with constant distortion [GNRS04]. This would rule out densifiers with a superlinear number of edges. In Section 4 we will give a direct proof of the latter statement.

2.2 Spectral densifiers and embeddings into ℓ_2^2

The next theorem shows that a graph cannot have a spectral densifier if it embeds into ℓ_2^2 with small average expansion. The *average expansion* of an embedding f of a unweighted graph $G = (V, E)$ into ℓ_2^2 is denoted by $\alpha_2(f) \stackrel{\text{def}}{=} \mathbb{E}_{(u,v)\in E} \|f(u) - f(v)\|_2^2$. We let $\alpha_2(G) \stackrel{\text{def}}{=} \min \alpha_2(f)$ where the minimum is taken over all non-contractive embeddings $f \colon V \to \ell_2^2$. Here, we mean all embeddings even those that do not satisfy triangle inequalities.

Note that ℓ_1 embeddings are a strict subset of ℓ_2^2 embeddings (not necessarily metric) consistent with the fact that spectral densifiers imply cut densifiers.

Theorem 2.4. *Let* $G = (V, E)$ *be a graph of maximum degree* Δ *and let* $H = (V, E_H)$ *be an unweighted* C*-multiplicative spectral densifier. Then,*

$$|E_H| \leqslant O\left(\Delta^{2C\cdot\alpha_2(G)} \cdot n\right) .$$

Proof. The proof is analogous to that of Theorem 2.1. Let $H = (V, E_H)$ be an unweighted C-multiplicative spectral densifier for G. Further, let $f : V \to \ell_2^2$ denote a non-contractive embedding of G that achieves $\alpha_2(G)$. We will bound the number of edges in H as follows. For every C-multiplicative spectral densifier and every set of real numbers $(\xi_u)_{u\in V}$ we have

$$\sum_{(u,v)\in E_H} (\xi_u - \xi_v)^2 \leqslant C \cdot \sum_{(u,v)\in E} (\xi_u - \xi_v)^2 . \tag{8}$$

Hence,

$$\begin{aligned}
\mathop{\mathbb{E}}_{(u,v)\in E_H} d_G(u, v) &\leqslant \mathop{\mathbb{E}}_{(u,v)\in E_H} \|f(u) - f(v)\|_2^2 && (f \text{ is non-contractive}) \\
&= \mathop{\mathbb{E}}_{(u,v)\in E_H} \sum_{i=1}^{m} (f(u)_i - f(v)_i)^2 \\
&= \sum_{i=1}^{m} \mathop{\mathbb{E}}_{(u,v)\in E_H} (f(u)_i - f(v)_i)^2 \\
&\leqslant C \sum_{i=1}^{m} \mathop{\mathbb{E}}_{(u,v)\in E} (f(u)_i - f(v)_i)^2 && (\text{using (8) with } \xi_u = f(u)_i) \\
&= C \mathop{\mathbb{E}}_{(u,v)\in E} \|f(u) - f(v)\|_2^2 \\
&= C \cdot \alpha_2(G) .
\end{aligned}$$

As argued in the proof of Theorem 2.1 this implies that $|E_H| \leqslant 2(\Delta^{2C\alpha_2(G)} + 1)n$ which is what we wanted to show. ■

3 Dual characterization and humble embeddings

In this section we will study the necessary and sufficient conditions for the existence of densifiers. For simplicity, we will focus on the case of *one-sided* C-multiplicative cut approximations, for which we already obtain strong non-densification results (the two-sided case is similar but somewhat more messy). Throughout this section, let the term *C-densifier of G* denote a graph H such that $e_H(S, S^c) \leqslant Ce_G(S, S^c)$ for all sets $S \subseteq [n]$.

As it turns out the existence of C-densifiers is closely related to the following notion of embedding.

Definition 3.1. We say that a graph G has an (α, C)-humble embedding into ℓ_1 if there exists an ℓ_1-metric ρ such that

$$\sum_{u,v \in V} \min\left\{\rho(u,v) - C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \rho(\tilde{u},\tilde{v}), 1\right\} \geqslant (1-\alpha)n^2 \,. \tag{9}$$

Intuitively, the requirement is that $(1-\alpha)n^2$ non-edges are bounded away from the average length of an edge in G. This means the embedding does not completely collapse all distance scales.

Theorem 3.2. *Let G be a graph and let $\alpha > 0, C > 1$. Then, G has an (α, C)-humble embedding into ℓ_1 if and only if G does not have a C-densifier with edge weight more than αn^2.*

Proof. We start with the linear program that computes a C-densifier of G with maximum edge weight.

$$\begin{aligned}
\max \quad & \sum_{u,v} x_{uv} \\
\text{subject to} \quad & \forall u,v \colon\ x_{uv} \leqslant 1 && (\sigma_{uv}) \\
& \forall S \colon \sum_{u\in S, v\in S^c} x_{uv} \leqslant C \cdot e_G(S,S^c) \sum_{u,v} x_{uv} && (\lambda_S) \\
& x_{uv} \geqslant 0
\end{aligned}$$

The dual program can be written as follows:

$$\begin{aligned}
\min_{\rho=\sum \lambda_S \delta_S} \quad & \sum_{u,v} \sigma_{uv} \\
\text{subject to} \quad & \forall u,v \colon \rho(u,v) - C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \rho(\tilde{u},\tilde{v}) \geqslant 1 - \sigma_{uv} \\
& \sigma_{uv}, \lambda_S \geqslant 0
\end{aligned}$$

Note that in an optimal solution we must have

$$\sigma_{uv} = \max\left\{0, C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \rho(\tilde{u},\tilde{v}) + 1 - \rho(u,v)\right\} .$$

Hence, the dual program simplifies to

$$\min_{\rho} \sum_{u,v} \max\left\{0, C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \rho(\tilde{u},\tilde{v}) + 1 - \rho(u,v)\right\} = n^2 - \max_{\rho} \sum_{u,v} \min\left\{1, \rho(u,v) - C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \rho(\tilde{u},\tilde{v})\right\} \tag{10}$$

It follows that a dual solution of value αn^2 is equivalent to the existence of an (α, C)-humble embedding. The theorem now follows by LP duality. ∎

The next lemma is a technical tool that in particular allows to argue that without loss of generality we can assume that $C \cdot \mathbb{E}_{(\tilde{u},\tilde{v})\in E_G}\, \rho(\tilde{u},\tilde{v}) \geqslant 1$ in the above optimization problem. This will be helpful later.

Lemma 3.3 (Rescaling Lemma)**.** *Consider the concave maximization program:*

$$\mathrm{OPT} = \max_{s\geqslant 0} f(s) \stackrel{\text{def}}{=} \max_{s\geqslant 0} \sum_{i\leqslant M} \min\{sx_i - s\beta, 1\} \tag{11}$$

where $\beta, x_i \geqslant 0$ are given as input. Then $\mathrm{OPT} = M - D$, for some integer D, and there is always an $s' \geqslant 1/\beta$ satisfying $f(s') \geqslant M - 2D$.

Proof. Put $y_i(s) = sx_i - s\beta$. Given a value of s (with corresponding $y_i(s)$), partition the y_i into sets $L = \{i : y_i(s) \geqslant 1\}$, $F = \{i : y_i(s) \in [0,1)\}$, and $N = \{i : y_i(s) < 0\}$ and also denote $y(L) = \sum_{i\in L} y_i(s)$, $y(F) = \sum_{i\in F} y_i(s)$, and $y(N) = \sum_{i\in N} y_i(s)$. We can assume the optimum has $s > 0$ since $s = 0$ gives $f(s) = 0$.

Observe that the derivative of $f(s)$ with respect to s is precisely $y(F) + y(N)$, which must be zero at the optimum s^*, y^*, so that

$$f(s^*) = y^*(F) + y^*(N) + y^*(L) = y^*(L) = |L| = M - (|F| + |N|) = M - D,$$

establishing the first claim.

For the second claim, we note that the average value of the negative terms satisfies:

$$\frac{|y^*(N)|}{|N|} \leqslant \max_{i\in N} |y_i^*| \leqslant s^*\beta.$$

Thus, if $|y^*(N)| \geqslant |N|$, then $s^* \geqslant 1/\beta$ and we are done. Else, we take

$$s' = \frac{|N|}{|y^*(N)|} s^*$$

and note that

$$s'\beta = \frac{|N|}{|y^*(N)|} s^*\beta \geqslant 1,$$

by the previous inequality. Moreover, the negative contribution of these terms is at most

$$y^*(N) \cdot \frac{|N|}{|y^*(N)|} = |N| \leqslant D,$$

as desired. Also, the contribution of the positive terms remains at least $M - D$ since $s' \geqslant s^*$. ∎

3.1 Spectral densifiers

Here we extend our results from the previous section to the case of spectral densifiers. The main observation is that the optimal multiplicative spectral densifier can be computed using a suitable semidefinite program.

In the following let $b_{uv} = e_u - e_v$ where e_u denotes the standard basis vector that has a 1 in coordinate u and is 0 elsewhere. Recall that for a graph G we have $L_G = \sum_{uv} w_{uv} b_{uv} b_{uv}^\top$ where $w_{u,v}$ is the weight of the edge (u,v).

The best one-sided multiplicative spectral densifier for a graph G is then given by the following SDP:

$$\begin{aligned} \max \quad & \sum_{u,v} x_{uv} \\ \text{subject to} \quad & \forall u,v\colon\ x_{uv} \leqslant 1 \\ & \sum_{u,v} x_{uv} b_{uv} b_{uv}^\top \preceq \left(C \sum_{u,v} x_{uv} \right) L_G \\ & x_{uv} \geqslant 0 \end{aligned}$$

The dual program is the following:

$$\begin{aligned} \min \quad & \sum \sigma_{uv} \\ \text{subject to} \quad & b_{uv}^\top Z b_{uv} - C \cdot \operatorname{tr}(L_G Z) \geqslant 1 - \sigma_{uv} \\ & Z \succeq 0, \sigma_{uv} \geqslant 0 \end{aligned}$$

Strong duality holds, since the primal program is strictly feasible. With the same reasoning as above, the dual simplifies to

$$n^2 - \max_Z \sum_{u,v} \min\left\{1, b_{uv}^\top Z b_{uv} - C \cdot \operatorname{tr}(L_G Z)\right\} = n^2 - \max_{z_1,\ldots,z_n} \sum_{u,v} \min\left\{1, \|z_u - z_v\|^2 - C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2\right\}, \tag{12}$$

where $z_1, \ldots, z_n$ are vectors in $\mathbb{R}^n$. This leads to the following definition, analogous to Definition 3.1.

Definition 3.4. We say that a graph G has an (α, C)-humble embedding into ℓ_2^2 if there are vectors $z_1, \ldots, z_n \in \mathbb{R}^n$ such that

$$\sum_{u,v\in V} \min\left\{\|z_u - z_v\|^2 - C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2, 1\right\} \geqslant (1-\alpha)n^2 . \tag{13}$$

We note that Lemma 3.3 also applies to (12) so that if G has an (α, C)-humble embedding into we may assume that it also has a $(2\alpha, C)$-humble embedding in which $C \cdot \mathbb{E}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2 \geqslant 1$.

Theorem 3.5. *Let G be a graph and let $\alpha > 0, C > 1$. Then, G has an (α, C)-humble embedding into ℓ_2^2 if and only if G does not have a one-sided C-multiplicative spectral densifier with edge weight more than αn^2.*

We will crucially make use of the dual characterization below proving a close connection between the cut and spectral case.

3.2 Relation between multiplicative cut and spectral densifiers

The next theorem shows that one-sided multiplicative cut and spectral densifiers are surprisingly closely related.

Theorem 3.6. *Let G be a graph and let $\gamma > 0$. Suppose that G has a one-sided C-multiplicative cut densifier with αn^2 edges. Then, G also has a one-sided $(1+\gamma)^2 C^2$-multiplicative spectral densifier with $\Omega(\frac{\gamma}{1+\gamma}\alpha n^2)$ edges.*

Proof. Put $C^* = (1+\gamma)^2 C^2$. We will prove the theorem in the contrapositive. Suppose G does not have a C^*-multiplicative spectral densifier with αn^2 edge weight. We will show that G admits a $(2(1+\gamma)\alpha/\gamma, C)$-humble embedding into ℓ_1. This implies, by Theorem 3.2, that there does not exist C-multiplicative cut densifier for G of edge weight more than $2\frac{1+\gamma}{\gamma}\alpha n^2$ for sufficiently large constant $c > 0$.

By our assumption and Theorem 3.5, we can find a vectors $z_1, \ldots, z_n \in \mathbb{R}^n$ with

$$\sum_{u,v} \min\left\{1, \|z_u - z_v\|^2 - C^* \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2\right\} \geqslant (1-2\alpha)n^2 . \tag{14}$$

such that, furthermore,

$$C^* \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2 \geqslant 1 . \tag{15}$$

The latter assumption is without loss of generality as shown in Lemma 3.3. Note that this is why we introduced a factor 2 in front of α in (14).

The proof strategy from here on is to first pass from squared Euclidean norms in (14) to Euclidean norms while quantifying the loss involved. Once we have a Euclidean metric we can use the fact ℓ_2 embeds isometrically into ℓ_1 to obtain a humble embedding into ℓ_1. For this purpose, we will need the following two lemmas.

Lemma 3.7. *Suppose that*

$$\|z_u - z_v\|^2 \geqslant C^* \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2 .$$

Then,

$$\|z_u - z_v\| \geqslant C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\| + \frac{\gamma}{1+\gamma} .$$

Proof. Taking square roots on both sides of the assumption, it follows that

$$\|z_u - z_v\| \geqslant (1+\gamma)C \cdot \sqrt{\mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2} .$$

Noting that by (15), the RHS is at least 1, this implies that

$$\|z_u - z_v\| \geqslant C \cdot \sqrt{\mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2} + \frac{\gamma}{1+\gamma} .$$

Finally, by Jensen's inequality,

$$\sqrt{\mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2} \geqslant \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|$$

which concludes the proof. ■

Lemma 3.8. *Suppose that* $\|z_u - z_v\|^2 < C^* \cdot \mathbb{E}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2$. *Then,*

$$\|z_u - z_v\| - C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\| \geqslant \|z_u - z_v\|^2 - C^* \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v})\in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2$$

Proof. First observe that for any two nonnegative numbers a, b such that $0 \leqslant a < b$ and $b \geqslant 1$ we have $\sqrt{a} - \sqrt{b} \geqslant a - b$. (Indeed, this is equivalent to $b - a \leqslant b^2 - a^2 = (b+a)(b-a)$ which is in turn equivalent to $b + a \geqslant 1$ and hence follows from the assumption that $b \geqslant 1$ and $a \geqslant 0$.) This fact implies that

$$\|z_u - z_v\| - \sqrt{C^* \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v}) \in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2} \geqslant \|z_u - z_v\|^2 - C^* \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v}) \in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2 ,$$

where we used (15) again. But note that by Jensen's inequality,

$$\sqrt{C^* \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v}) \in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2} \geqslant \sqrt{C^*} \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v}) \in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\| \geqslant C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v}) \in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\| .$$

Hence, the claim follows. ■

Let

$$P = \left\{ (u,v) \in V \times V \colon \ \|z_u - z_v\|^2 \geqslant C^* \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v}) \in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2 \right\}$$

and put

$$d(u,v) = \frac{1+\gamma}{\gamma} \|z_u - z_v\| .$$

Note that for every $(u,v) \in P$, it follows from Lemma 3.7 that

$$\min \left\{ 1, d(u,v) - C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v}) \in E_G} d(\tilde{u},\tilde{v}) \right\} \geqslant 1 .$$

On the other hand, for every $(u,v) \notin P$, it follows from Lemma 3.8 that

$$\min \left\{ 1, d(u,v) - C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v}) \in E_G} d(\tilde{u},\tilde{v}) \right\} \geqslant \frac{1+\gamma}{\gamma} \left(\|z_u - z_v\|^2 - C^* \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v}) \in E_G} \|z_{\tilde{u}} - z_{\tilde{v}}\|^2 \right) .$$

We conclude from the previous two lemmas that

$$\sum_{u,v} \min \left\{ 1, d(u,v) - C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v}) \in E_G} d(\tilde{u},\tilde{v}) \right\} \geqslant n^2 - \frac{1+\gamma}{\gamma} 2\alpha n^2 .$$

Finally, note that $d \in \ell_2 \subseteq \ell_1$. Hence, there is an ℓ_1-metric ρ such that

$$\sum_{u,v} \min \left\{ 1, \rho(u,v) - C \cdot \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v}) \in E_G} \rho(\tilde{u},\tilde{v}) \right\} \geqslant n^2 \left(1 - \frac{1+\gamma}{\gamma} 2\alpha \right)$$

This means that G has a $(2(1+\gamma)\alpha/\gamma, C)$-humble embedding into ℓ_1 which is what we wanted to show. ■

We conjecture that Theorem 3.6 is also true for two-sided densifiers.

3.3 Spectral densifier for the dumbbell graph

In this section we consider the example of the dumbbell graph (two expanders connected by a single edge) and show that it has good multiplicative spectral densifiers in spite of having a tiny spectral gap. This shows that expansion alone does not characterize the class of graphs which have densifiers.

Formally, the dumbbell G^{ij} of degree d is a disjoint union of two degree d expanders (say Ramanujan expanders for concreteness) E_X and E_Y on vertex sets X and Y of cardinality n, connected by a single edge ij for $i \in X, j \in Y$. The dumbbell has $dn + 1$ edges and is thus sparse when $d \ll n$. We now show that it is well-approximated by two copies of the complete graph joined by a matching.

Proposition 3.9. *Let G^{ij} be a dumbbell with $d \geqslant 3$ and let H consist of two complete graphs on vertex sets X and Y connected by any matching of size n/d. Then*

$$\frac{1}{4}H \preceq (n/d)G^{ij} \preceq 4H,$$

where we have written $H \preceq G$ in place of $L_H \preceq L_G$ to ease notation.

Proof. The proof relies on the fact that any G^{ij} is well-approximated by any other dumbbell $G^{i',j'}$ with the same partition $V = X \cup Y$. To show this, we first observe that by definition, each expander is well-approximated by a reweighted complete graph:

$$\frac{d - O(\sqrt{d})}{n} K_n(X) \preceq E_X \preceq \frac{d + O(\sqrt{d})}{n} K_n(X). \tag{16}$$

Thus, incurring a multiplicative error of $1 \pm O(\sqrt{d}/n)$ it suffices to consider instead of G^{ij} the graph G_K^{ij} given by two copies of the scaled complete graph $(d/n)K_n(X), (d/n)K_n(Y)$ connected by the edge ij.

Claim 3.10. *For any $i' \in X, j' \in Y$ we have*

$$G_K^{ij} \preceq 3 \cdot G_K^{i'j'}.$$

Proof. Write

$$G_K^{ij} = K_n(X) + K_n(Y) + L_{ij} \quad \text{and} \quad G_K^{i'j'} = K_n(X) + K_n(Y) + L_{i'j'},$$

where $L_{ij} = (e_i - e_j)(e_i - e_j)^T$ is the Laplacian of a single edge ij. We will show that

$$L_{ij} \preceq 2 \cdot G_K^{i'j'} \tag{17}$$

and adding $K_n(X) + K_n(Y)$ to both sides will give the claim.

Recall that for an edge ij and graph G:

$$L_{ij} \preceq L_G \iff (e_i - e_j)^T L_G^+ (e_i - e_j) \leqslant 1 \iff \mathrm{Reff}_G(i,j) \leqslant 1,$$

where Reff denotes effective resistance [SS08]. Since effective resistance is a metric, we have

$$\begin{aligned}\mathrm{Reff}_{G_K^{i'j'}}(i,j) &\leqslant \mathrm{Reff}_{G_K^{i'j'}}(i,i') + \mathrm{Reff}_{G_K^{i'j'}}(i',j') + \mathrm{Reff}_{G_K^{i'j'}}(j',j) \\ &= (n/d)\cdot(2/n) + 1 + (n/d)\cdot(2/n) \quad \text{since the complete graph has } \mathrm{Reff}_{K_n}(i,j) = 2/n \\ &= 1 + 4/d.\end{aligned}$$

Since Reff scales linearly, this implies

$$\mathrm{Reff}_{2G_k^{i'j'}}(i,j) \leqslant \frac{1+4/d}{2} \leqslant 1,$$

establishing (17) as desired. ■

The claim implies that

$$(1/3)G_K^{i'j'} \preceq G_K^{ij} \preceq 3G_K^{i'j'} \tag{18}$$

for any $i'j'$. Let M be any matching of n/d edges between X and Y, and consider the graph $H = \cup_{i'j' \in M} G_K^{i'j'}$. Now H is unweighted and summing (18) over all $i'j' \in M$ yields

$$(1/3)H \preceq (n/d)G_K^{ij} \preceq 3H,$$

which together with (16) finishes the proof. ■

The above proof can easily be adapted to the case of unbalanced dumbbells, where the expanders have different degrees or different sizes.

4 Embeddability results

We are about to show that certain families of graphs naturally exhibit humble embeddings. We start with planar graphs in the following section.

4.1 Planar graphs

We will need the following suitable variant of the planar separator theorem due to [Mil86, DDSV93].

Theorem 4.1. *Every planar graph with maximum degree Δ can be partitioned into two sets A, B each of size at most $n/2+1$ such that $O(\sqrt{\Delta n})$ edges are cut.*

The theorem is usually stated with $2n/3$ replacing $n/2+1$, but see [ST04] for an argument on why this implies the theorem as stated.

Theorem 4.2. *Let $f(n)$ be any function growing with n. Then, for sufficiently large n, every bounded degree n-vertex planar graph G has an $(O(f(n)), O(1))$-humble embedding into ℓ_1.*

Proof. Let C be any constant and $\alpha = O(f(n))$. We will show that G has an (α, C)-humble embedding by recursively applying Theorem 4.1 to G. On the first application we obtain two sets $A, B \subseteq [n]$ each of size at least $n/2-1$ cutting only $O(\sqrt{n})$ edges. We let $G[A], G[B]$ denote the subgraphs induced by A, B respectively. Note that both graphs are still planar. Hence, we can apply Theorem 4.1 to each graph again. We repeat this process recursively $k = \lceil \log(n/f(n)) \rceil$ times. At the end of this process we arrive at a partition of the vertices into sets $A_1, \ldots, A_{2^k}$ such that each set has size at most $O(f(n))$.

Claim 4.3. *The number of edges crossing the partition is $O\left(n/\sqrt{f(n)}\right)$.*

Proof. The number of edges cut at step i is at most $O(2^i\sqrt{n/2^i})$ by Theorem 4.1. Hence, the total number of edges cut is

$$\sum_{i=1}^{k} 2^i \cdot O\left(\sqrt{\frac{n}{2^i}}\right) = O\left(\sqrt{n}\right) \sum_{i=1}^{k} 2^{i/2} = O\left(2^{k/2}\sqrt{n}\right).$$

■

Now, let $\rho = \sum_{i=1}^{p} \delta_{A_i}$ be the ℓ_1-metric corresponding to the cuts $A_1, \ldots, A_p$. The next claim concludes the proof of the theorem.

Claim 4.4. *The metric ρ is an (α, C)-humble embedding of G.*

By Claim 4.3,

$$\mathop{\mathbb{E}}_{G} \rho(u,v) \leqslant \frac{O(1)}{\sqrt{f(n)}} \leqslant \frac{1}{C}\,,$$

for sufficiently large n.

On the other hand, the only way a pair (u,v) is not separated by the partition is if $\{u,v\} \in A_i$ for some i. By our choice of k we have that $|A_i| \leqslant O(f(n))$. The fraction of pairs not crossing the partition is therefore bounded by $O(f(n)/n)$. If on the other hand (u,v) is such that $u \in A_i$ and $v \in A_j$ for $i \neq j$, then $\rho(u,v) \geqslant 2$ and thus $\rho(u,v) \geqslant C\,\mathbb{E}_G\,\rho(u,v) + 1$. Furthermore, $\rho(u,v) - C\,\mathbb{E}_G\,\rho(u,v) \geqslant -1$ for all pairs (u,v). Hence,

$$\sum_{u,v \in V} \min\{1, \rho(u,v) - C\rho(u,v)\} \geqslant (1 - O(f(n))/n)n^2\,.$$

This shows that ρ is indeed an (α, C)-humble embedding. ■

With the previous theorem at hand, the next corollary is an immediate consequence of Theorem 3.2.

Corollary 4.5. *For every $m \in \omega(n)$, every sufficiently large bounded degree planar graph does not have a $O(1)$-multiplicative cut densifier of edge weight m.*

4.2 Random geometric graphs

A random geometric graph is obtained by taking the vertices as random unit vectors in $\mathbb{R}^d$, and connecting two vertices $u, v \in \mathbb{R}^d$ based on their spherical distance. We study the distribution $\mathcal{G}(n, d, \gamma)$, where the vertices are n random unit vectors in $\mathbb{R}^d$, and two vectors are connected by an edge iff $\arccos(\langle u, v\rangle) \leqslant \gamma$. The expected degree $\deg(n, \gamma)$ of these graphs can be easily shown to be $n\mu(\gamma)$, where $\mu(\gamma)$ denotes the spherical measure of a cap of radius γ, i.e., $\mu(\gamma) = \mathrm{Vol}_{d-1}(\{u \in \mathbb{S}^{d-1} \colon \arccos(\langle u, u_0\rangle) \leqslant \gamma\})/\mathrm{Vol}_{d-1}(\mathbb{S}^{d-1})$ where the choice of $u_0 \in \mathbb{S}^{d-1}$ is arbitrary due to spherical symmetry.

Theorem 4.6. *Let $\varepsilon > 0$. Then, for sufficiently large n and every $m \in [1, n^2]$ there is a choice of d, γ such that $G \sim \mathcal{G}(n, d, \gamma)$ has m edges in expectation, but with high probability there does not exists a $O(1)$-multiplicative cut densifier with edge weight $\Omega(n^{\varepsilon} m)$.*

Proof. Let $G = (V, E)$ be a graph sampled from $\mathcal{G}(n, d, \gamma)$ where we put $d = 100/\varepsilon$. We choose γ such that $\deg(n, \gamma) = m/n$. This implies that the expected number of edges is m. Note that $\mu(r) \leqslant r^d \mu(1) = r^d$. Since $d = 100/\varepsilon$ this means that $\mu(n^{-\varepsilon}) < 1/n$ and thus $\gamma > n^{-\varepsilon}$.

Consider the following weighted family of cuts. For a set $S \subseteq [n]$ we define λ_S to be the probability that a random hyperplane cut on the sphere induces the partition (S, S^c) on G. We let $\rho = \sum_{S \subseteq [n]} \lambda_S \delta_S$ be the corresponding cut metric. Note that $\rho(u,v)$ is equal to the probability that (u,v) is separated by a random hyperplane cut, i.e., $\arccos(\langle u, v\rangle)$. Hence, for every edge (u,v) in the graph G we have

$$\rho(u,v) \leqslant \gamma.$$

On the other hand, put $C = O(1)$ and let $F = \{(u, v) \in V \times V : \arccos(\langle u, v\rangle) \geqslant 10C\gamma\}$. Since the dimension is a constant independent of n we have that $\deg(n, 10C\gamma) = O(\deg(n, \gamma))$ and hence with high probability

$$|F| \geqslant n^2 - O(m).$$

Now, consider $\rho' = \rho/\gamma$. We claim that ρ' is an (α, C)-humble embedding of G with $\alpha = O(mn^{\varepsilon}/n^2)$. Indeed, the expectation of $\rho'(u, v)$ over edges is at most 1, while $\rho(u, v) \geqslant 10C$ for every $(u, v) \in F$. Pairs in F^c can only contribute $-O(Dn/\gamma) = -O(Dn^{1+\varepsilon})$. Hence,

$$\sum_{u,v\in V} \min\{1, \rho'(u, v) - C \mathop{\mathbb{E}}_{G} \rho'(u, v)\} \geqslant n^2 - O(mn^{\varepsilon}) .$$

Rearranging gives the stated bound on α. The theorem then follows from Theorem 3.2. ■

We remark that a very similar proof can be used to establish that random geometric graphs in the plane (or higher dimensional Euclidean space) do not have densifiers. Such graphs are defined by taking the vertex set to be random points in the unit square $[0, 1]^2$ (say), and connecting all pairs of vertices which are within distance r of each other, for some threshold $r \geqslant 0$. When r is subconstant as a function of the number of vertices n, the graph is sparse. It is then easy to check (in a manner similar to Theorem 4.6) that the given embedding of the graph into ℓ_2 is humble, and this rules out any nontrivial cut densification since ℓ_2 embeds isometrically into ℓ_1.

Such random graph models are used heavily in wireless routing and other applications, see for instance [Pen03].

5 Additive densifiers

We begin with the observation that we can efficiently compute a nearly optimal additive cut densifier for a given graph G. This is in contrast to the multiplicative cut case where we don't have an efficient algorithm for this task.

Proposition 5.1. *Given a graph G and $\varepsilon > 0$ we can efficiently compute an 12ε-additive cut densifier whose edge weight is $\mathrm{OPT}_\varepsilon - o(1)$ where OPT_ε denotes the edge weight of an optimal ε-additive cut densifier.*

Proof (Sketch). Consider the linear program $\max_H m(H)$ subject to $|e_G(S, T) - e_H(S, T)| \leqslant 6\varepsilon$ for all $S, T \subseteq [n]$. Here, the maximum is over all weighted graphs H.

By Lemma 1.1, if an optimal ε-additive cut approximation H has weight $m(H)$, then H is a feasible solution to the above LP. Hence, the LP has optimal value at least $m(H)$.

The advantage of using the above LP is that we can implement an efficient approximate separation oracle using the Alon-Naor SDP relaxation for the cut-norm [AN04]. Since the SDP relaxation has an integrality gap of less than a factor 2, all constraints are satisfied up to a factor 2. Hence, the resulting graph must be a 12ε-approximation to G. Since we can efficiently solve the LP up to numerical error $o(1)$, the claim follows. ■

The dual of the additive cut LP is deferred to Appendix A as we will not use the dual in this section. Instead we will give a direct proof that the cycle does not have additive cut densifiers with $\omega(n)$ edges. Our proof establishes the claim in two steps. In Section 5.2 we rule out *spectral* additive densifiers using the fact that the cycle is a Cayley graph of the group $\mathbb{Z}/n\mathbb{Z}$. We then transfer this result in Section 5.1 to cut densifiers by showing that additive spectral and additive cut approximations are equivalent for transitive graphs.

5.1 Transitive Graphs

In this section we attempt to relate the cut and spectral notions of additive densification. The relationship

$$\|A_G/m(G) - A_H/m(H)\|_{\infty\to 1} \leqslant n \cdot \|A_G/m(G) - A_H/m(H)\|_{2\to 2}$$

always holds by Cauchy-Schwarz. We show here that a reverse inequality also holds when G has a lot of symmetry.

Proposition 5.2. *Suppose G is vertex-transitive. Then the best densifier H for G may be assumed to be transitive under the automorphism group of G, and moreover*

$$\|A_G/m(G)-A_H/m(H)\|_{\infty\to 1} \leqslant n\cdot\|A_G/m(G)-A_H/m(H)\|_{2\to 2} \leqslant K_G\cdot\|A_G/m(G)-A_H/m(H)\|_{\infty\to 1},$$

where K_G is the Grothendieck constant.

Proof. Let $\widehat{A_G} = A_G/m(G)$ and $\widehat{A_H} = A_H/m(H)$. Observe that $PA_GP = A_G$ for every permutation P in the (transitive) automorphism group $\mathrm{Aut}(G)$ of G, so that

$$\|\widehat{A_G} - \widehat{A_H}\|_2 = \|P(\widehat{A_G} - \widehat{A_H})P\|_2 = \|\widehat{A_G} - P\widehat{A_H}P\|_2$$

for any such P. By convexity of the norm we conclude that:

$$\left\|\widehat{A_G} - \frac{1}{n}\sum_{P\in\mathrm{Aut}(G)} P\widehat{A_H}P\right\|_2 \leqslant \frac{1}{n}\sum_{P\in\mathrm{Aut}(G)} \|\widehat{A_G} - P\widehat{A_H}P\|_2 = \|\widehat{A_G} - \widehat{A_H}\|_2.$$

Thus we may assume without loss of generality that $\widehat{A_H} = \frac{1}{n}\sum_{P\in\mathrm{Aut}(G)} P\widehat{A_H}P$, establishing the first claim.

The proof of the second claim is almost immediate from the following semidefinite programming characterization of the Grothendieck factorization due to Tropp [Tro09].

Theorem 5.3 (Grothendieck Factorization [Pis86, Tro09])**.** *If A is a symmetric matrix, then there is a nonnegative diagonal matrix D with $\mathrm{tr}(D) = 1$ and*

$$\|A\|_{\infty\to 1} \leqslant \|\sqrt{D^+}A\sqrt{D^+}\|_{2\to 2} \leqslant K_G \cdot \|A\|_{\infty\to 1} \tag{19}$$

where K_G is a numerical constant and D^+ denotes the pseudoinverse.

Moreover, the matrix D can be obtained by solving the following semidefinite program:

$$\min \alpha \quad \textit{subject to} \quad D \geqslant 0 \textit{ diagonal}, \quad \mathrm{tr}(D) = 1, \quad -\alpha D \preceq A \preceq \alpha D. \tag{20}$$

Let $A = \widehat{A_G} - \widehat{A_H}$. For any $\alpha \geqslant 0$ consider the feasible set

$$S_\alpha = \{D \geqslant 0 \text{ diagonal} : \mathrm{tr}(D) = 1 \wedge -\alpha D \preceq A \preceq \alpha D\},$$

and let α^* be the smallest α for which $S_\alpha \neq \varnothing$ (it exists because $\{\alpha : S_\alpha \neq \varnothing\}$ closed). Then the diagonal matrix D promised by the Grothendieck factorization is any member of S_{α^*}. Let P be the permutation matrix corresponding to any automorphism of G; it is immediate that $PAP = A$ so that for any D

$$-\alpha D \preceq A \preceq \alpha D \iff -\alpha D \preceq PAP \preceq \alpha D \iff -\alpha P^{-1}DP^{-1} \preceq A \preceq \alpha P^{-1}DP^{-1}.$$

Thus we must have $P^{-1}DP^{-1} \in S_{\alpha_0}$ for every $P \in \mathrm{Aut}(G)$ and $D \in S_{\alpha^*}$. Since S_{α^*} is convex, we can average over all automorphisms and apply vertex-transitivity to conclude that $\frac{1}{n}I \in S_{\alpha^*}$. Using this as D in (19) finishes the proof. ■

5.2 The Cycle

The following proposition shows that absolutely no nontrivial spectral additive densification is possible for the cycle.

Theorem 5.4. *Let n be prime and $\varepsilon \leqslant 2$. Suppose G is the cycle on n vertices and H is any graph with edge weights bounded by 1. If*

$$\left\| \frac{A_G}{m_G} - \frac{A_H}{m_H} \right\|_2 \leqslant \varepsilon/n \tag{21}$$

then H has total weight at most $\frac{2n}{2-\varepsilon}$.

Proof. Let the length of an edge in H be the shortest path distance between the end points of the corresponding pair in G. As $\mathrm{Aut}(G) = \mathbb{Z}/n\mathbb{Z}$, Proposition 5.2 implies that H is invariant under all cyclic P. Thus it must assign the same weight $\alpha_j \in [0,1]$ to all edges of a given length j for $j = 1, 2, \dots, \lfloor n/2 \rfloor$, and can therefore be represented as

$$A_H = \sum_j \alpha_j A_{S_j}$$

where S_j is the graph containing all edges of length j.

Since n is prime, each S_j is the Cayley graph of $\mathbb{Z}/n\mathbb{Z}$ with generators $\{-j, +j\}$, the $j = 1$ case being the original cycle G. All of the matrices A_{S_j} have exactly the same eigenvectors, which are given by the characters of $\mathbb{Z}/n\mathbb{Z}$, and may be written along with their well-known spectra as follows:

$$\begin{aligned} &v_0(x) = 1 && \text{with eigenvalue} \quad 0 \\ &v_k(x) = \cos(2\pi kx/n) \quad \text{and} \quad w_k(x) = \sin(2\pi kx/n) && \text{with eigenvalue } \lambda^j(k) = 2\cos(2\pi \cdot kj/n). \end{aligned}$$

Notice that multiplication by k is a bijection for n prime, so that the vectors λ^j containing the spectra of the A_{S_j} are simply permutations of each other.

Since all of the relevant matrices A_{S_j} commute, we can rewrite the quantity of interest as:

$$\begin{aligned} \|\widehat{A_G} - \widehat{A_H}\|_2 &= \left\| \frac{A_{S_1}}{m_{S_1}} - \frac{\sum_j \alpha_j A_{S_j}}{\sum_j \alpha_j m_{S_j}} \right\|_2 \\ &= \frac{1}{n} \left\| A_{S_1} - \frac{\sum_j \alpha_j A_{S_j}}{\sum_j \alpha_j} \right\|_2 \quad \text{since } m_{S_j} = n \text{ for all } j \\ &= \frac{1}{n} \left\| \lambda^1 - \frac{\sum_j \alpha_j \lambda^j}{\sum_j \alpha_j} \right\|_\infty \quad \text{where } \lambda^j \text{ are vectors containing the eigenvalues of the } A_{S_j} \\ &= \frac{1}{n} \|\lambda^1 - \sum_j \beta_j \lambda^j\|_\infty \quad \text{for convex coefficients } \beta_j \\ &= \frac{1}{n} \|(1-\beta_1)\lambda^1 - (1-\beta_1)v\|_\infty \quad \text{for some } v \in \mathrm{conv}(\lambda^j : j \geqslant 2) \\ &= \frac{1-\beta_1}{n} \|\lambda^1 - v\|_\infty. \end{aligned}$$

By Lemma 5.5 we know that $\|\lambda^1 - v\|_\infty \geqslant 2$, so in order to satisfy (21) we must have

$$\frac{\sum_{j \geqslant 2} \alpha_j}{\sum_j \alpha_j} = 1 - \beta_1 \leqslant \varepsilon/2,$$

which upon rearranging is just

$$\sum_{j \geqslant 2} \alpha_j \leqslant \frac{\alpha_1}{2/\varepsilon - 1} \leqslant \frac{1}{2/\varepsilon - 1},$$

for $\varepsilon \leqslant 2$. Thus the total weight of edges in H is at most

$$1 + \frac{1}{2/\varepsilon - 1} = \frac{2}{2-\varepsilon}.$$

■

Lemma 5.5. *Let $\lambda^1, \lambda^2, \ldots, \lambda^{\lfloor n/2 \rfloor} \in \mathbb{R}^n$ be given by $\lambda^j(k) = 2\cos(2\pi k \cdot j/n)$. Then*

$$\|\lambda^1 - v\|_\infty \geqslant 2$$

for every $v \in \operatorname{span}(\lambda^j : j \geqslant 2)$.

Proof. From standard facts regarding orthogonality of trigonometric polynomials we obtain

$$\langle \lambda^{j_1}, \lambda^{j_2} \rangle = 4 \sum_{k=0}^{n-1} \cos(2\pi k j_1/n) \cos(2\pi k j_2/n) = \frac{4n}{2\pi} \int_0^{2\pi} \cos(j_1 x) \cos(j_2 x) dx = 4n \mathbf{1}_{j_1 = j_2}.$$

Thus

$$\begin{aligned} \|\lambda^1 - v\|_\infty &\geqslant \frac{\|\lambda^1 - v\|_2}{\sqrt{n}} \\ &\geqslant \frac{\|\lambda^1\|_2}{\sqrt{n}} \quad \text{since } \langle \lambda^1, v \rangle = 0 \\ &= 2, \end{aligned}$$

as desired. ■

6 Conclusion and open problems

Our goal in this work was to initiate a principled study of graph densification. We gave several results characterizing when densifiers exist thus ruling out densifiers for various families of *intrinsically sparse* graphs. In the multiplicative case our characterization was based on a one-side approximation. It would be interesting to understand the dual interpretation of the two-sided condition. In particular, this might lead to a generalization of Theorem 3.6 to the two-sided case. We believe this is an interesting open problem.

Our understanding of the additive case is somewhat less developed than the multiplicative case. In particular, we do not have a satisfying characterization of when additive densifiers exist. We believe that there should be a closer connection between additive densifiers and certain graph partitions. This intuition partly stems from the dual in Appendix A as well as the fact that the existence of densifiers with $\Omega(n^2)$ edges is closely related to the existence of low-width cut decompositions (or weak regularity partitions) as shown by [RTTV08, COCF10].

Acknowledgments

We would like to thank Boaz Barak, Russell Impagliazzo, and James Lee for always insightful discussions on the topic.

References

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.

[AN04] Noga Alon and Assaf Naor. Approximating the cut-norm via Grothendieck's inequality. In *STOC '04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 72–80, New York, NY, USA, 2004. ACM.

[BHHS11] Boaz Barak, Moritz Hardt, Thomas Holenstein, and David Steurer. Subsampling semidefinite programs and average-case complexity. In *Proc. 22nd Symposium on Discrete Algorithms (SODA)*. ACM-SIAM, 2011.

[BK96] András A. Benczúr and David R. Karger. Approximating minimum cuts in $\tilde{O}(n^2)$ time. In *Proc. 28th STOC*, pages 47–55. ACM, 1996.

[BL06] Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap*. *Combinatorica*, 26(5):495–519, 2006.

[BN04] Béla Bollobás and Vladimir Nikiforov. Hermitian matrices and graphs: singular values and discrepancy. *Discrete Mathematics*, 285(1-3):17–32, 2004.

[BSS09] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-Ramanujan sparsifiers. In *Proc. 41st STOC*, pages 255–262. ACM, 2009.

[COCF10] Amin Coja-Oghlan, Colin Cooper, and Alan M. Frieze. An efficient sparse regularity concept. *SIAM J. Discrete Math.*, 23(4):2000–2034, 2010.

[DDSV93] Krzysztof Diks, Hristo Djidjev, Ondrej Sýkora, and Imrich Vrto. Edge separators of planar and outerplanar graphs with applications. *J. Algorithms*, 14(2):258–279, 1993.

[FHHP11] Wai Shing Fung, Ramesh Hariharan, Nicholas J. A. Harvey, and Debmalya Panigrahi. A general framework for graph sparsification. In *Proc. 43st STOC*, pages 71–80. ACM, 2011.

[FK96] Alan M. Frieze and Ravi Kannan. The regularity lemma and approximation schemes for dense problems. In *FOCS*, pages 12–20. IEEE, 1996.

[FS02] Uriel Feige and Gideon Schechtman. On the optimality of the random hyperplane rounding technique for max cut. *Random Struct. Algorithms*, 20(3):403–440, 2002.

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998.

[GNRS04] Anupam Gupta, Ilan Newman, Yuri Rabinovich, and Alistair Sinclair. Cuts, trees and ℓ_1-embeddings of graphs. *Combinatorica*, 24(2):233–269, 2004.

[GT08] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167(2):481–547, 2008.

[Hås01] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.

[LLR95] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[LT80] Richard J. Lipton and Robert Endre Tarjan. Applications of a planar separator theorem. *SIAM J. Comput.*, 9(3):615–627, 1980.

[Mil86] Gary L. Miller. Finding small simple cycle separators for 2-connected planar graphs. *Journal of Computer and System Sciences*, 32(3):265 – 279, 1986.

[Pen03] Mathew Penrose. *Random geometric graphs.* Oxford University Press, 2003.

[Pis86] Gilles Pisier. *Factorization of Linear Operators and Geometry of Banach Spaces.* Number 60 in CBMS Regional Conference Series in Mathematics. Amer. Math. Soc., Providence, RI, 1986.

[PY91] Christos H. Papadimitriou and Mihalis Yannakakis. Optimization, approximation, and complexity classes. *J. Comput. Syst. Sci.*, 43(3):425–440, 1991.

[Rao99] Satish Rao. Small distortion and volume preserving embeddings for planar and euclidean metrics. In *Symposium on Computational Geometry*, pages 300–306, 1999.

[RTTV08] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil P. Vadhan. Dense subsets of pseudorandom sets. In *Proc.* 49*th FOCS*, pages 76–085. IEEE, 2008.

[SS08] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. In *Proc.* 40*th STOC*. ACM, 2008.

[ST04] Daniel A. Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proc.* 36*th STOC*, pages 81–90. ACM, 2004.

[Tro09] Joel A. Tropp. Column subset selection, matrix factorization, and eigenvalue optimization. In *SODA*, pages 978–986, 2009.

A LP duality for the additive densifier

The primal LP for the additive case is as follows:

$$\begin{aligned} \max \quad & \sum_{u,v} x_{uv} \\ \text{subject to} \quad & \sum_{u,v} x_{uv}\delta_S(u,v) \geqslant e_G(S,S^c) - \varepsilon && (\lambda_S) \\ & \sum_{u,v} x_{uv}\delta_S(u,v) \leqslant e_G(S,S^c) + \varepsilon && (\sigma_S) \\ & 0 \leqslant x_{uv} \leqslant 1 \end{aligned}$$

With the same reasoning as before, the dual simplifies to:

$$n^2 - \max \sum_{u,v} \min \left\{ 1, \rho(u,v) - \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v}) \in E_G} \rho(\tilde{u},\tilde{v}) - \varepsilon \sum_S \lambda_S - \mu(u,v) + \mathop{\mathbb{E}}_{(\tilde{u},\tilde{v}) \in E_G} \mu(\tilde{u},\tilde{v}) - \varepsilon \sum_S \sigma_S \right\}, \tag{22}$$

where the minimum is taken over all ℓ_1-metrics $\rho = \sum_S \lambda_S \delta_S$ and $\mu = \sum_S \lambda_S \delta_S$.