On set expansion problems and the Small Set expansion Conjecture

Rajiv Gandhi [*] and Guy Kortsarz [**]

Department of Computer Science, Rutgers University-Camden, Camden, NJ 08102.
E-mail: {`rajivg, guyk`}`@camden.rutgers.edu`.

Abstract. We consider problems related to the The Small Set expansion conjecture (Small set Expansion Conjecture) [14]. In the MWEC problem, we are given an undirected simple graph $G = (V, E)$ with integral vertex weights. The goal is to select a set $U \subseteq V$ of maximum weight so that the number of edges with at least one endpoint in U is at most m'. Goldschmidt and Hochbaum [8] show that the problem is NP-hard and they give a 3-approximation algorithm for the problem. We present a polynomial time approximation algorithm with ratio 2 algorithm for, MWEC improving the bound of 3 of [8]. Interestingly, we show that a $2 - \epsilon$ ratio for MWEC for any constant $\epsilon > 0$ implies that the Small set Expansion Conjecture [14] fails. Thus under the Small set Expansion Conjecture , the ratio is for MWEC , is tight. To the best of our knowledge, this is the first time that the Small set Expansion Conjecture is shown to be related to breaking the threshold of some approximation problem. The $2 - \epsilon$ inapproximability considerably improves the NPC result of [8]. In the FCEC problem, we are given a vertex weighted graph, a bound k, and our goal is to find a subset of vertices U of total weight at least k such that the number of edges with at least one endpoint in in U is minimized. The NPC result in [8] carries over to this problem as well. The best known ratio for the problem is $2(1 + \epsilon)$ by Carnes and Shmoys [3]. We give a polynomial time ratio 2 approximation algorithm for FCEC improving [3] and show that a $2 - \epsilon$ polynomial time, approximation ratio algorithm, for any constant $\epsilon > 0$, refutes the Small set Expansion Conjecture . Thus under the Small set Expansion Conjecture our ratio is tight. This largely improves the NPC result that follows from [8].

We show that the natural LP for FCEC has an integrality gap of $2 - o(1)$. We show that for any constant $\rho > 1$, an approximation guarantee of ρ for the FCEC problem implies a $\rho(1 + o(1))$ approximation for MWEC . Finally, we define the Degrees density augmentation problem which is the density version of the FCEC problem. In this problem we are given an undirected graph $G = (V, E)$ and a set $U \subseteq V$. The objective is to find a set W so that $(e(W) + e(U, W))/deg(W)$ is maximum. This problem admits an LP-based exact solution [4]. We give a combinatorial algorithm for this problem.

1 Introduction

The Small set Expansion Conjecture is the following. The expansion of a set S is defined as

$$\phi(S) = \frac{e(S, \bar{S})}{deg(S)}.$$

Here $e(S, \bar{S})$ are the number of edges with one vertex in S and one in $V - S$ and $deg(S)$ is the sum of degrees of the vertices in S.

Given some $\delta \leq 1/2$ and a d-regular graph F, consider all subsets $\mathcal{S}$ of V of size at most $\delta \cdot S$. Let

$$\phi_G(\delta) = \min_{S \in \mathcal{S}} \frac{e(S, \bar{S})}{deg(S)}.$$

The Small set Expansion Conjecture : Given any constant $\mu > 0$, it is hard to distinguish between the cases that:

1. $\phi_G(\delta) \geq 1 - \mu$ and
2. $\phi_G(\delta) \leq \mu$

We note that the Small set Expansion Conjecture [14], is easier than the Unique Game conjecture of Khot [9]. Namely, proving the Small set Expansion Conjecture gives a proof for Unique Game conjecture. And an algorithm that refutes the Unique game Conjecture, refutes the Small set Expansion Conjecture .

The purpose of this paper is to relate the Small set Expansion Conjecture [14] to the approximability of two other edge expansion problems. We say that an edge e is *touched* by a set of vertices U or that e *touches* the set of vertices U, if at least one of e's endpoints is in U. The Maximum weight m'-edge cover (MWEC) problem that we study was first introduced by Goldschmidt and Hochbaum [8]. In this problem, we are given an undirected simple graph $G = (V, E)$ with integral vertex weights. The goal is to select a subset $U \subseteq V$ of maximum weight so that the number of edges touching U is at most m'. This problem is motivated by application in loading of semi-conductor components to be assembled into products [8].

We also study the closely related Fixed cost minimum edge cover (FCEC) problem in which given a graph $G = (V, E)$ with vertex weights and a number W, our goal is to find $U \subseteq V$ of weight at least W such that the number of edges touching U is minimized.

Finally, we study the Degrees density augmentation problem which is the density version of the FCEC problem. In the Degrees density augmentation problem, we are given an undirected graph graph $G = (V, E)$ and a set $U \subseteq V$ and our goal is to find a set W with maximum augmenting density i.e., a set W that maximizes $(e(W) + e(U, W))/deg(W)$.

1.1 Related Work

Goldschmidt and Hochbaum [8] introduced the MWEC problem. They show that the problem is NP-complete and give algorithms that yield 2-approximate and 3-approximate algorithm for the unweighted and the weighted versions of the problem, respectively. Their NP-hardness proof carries over to FCEC as well. A ratio of $2 + \epsilon$ for FCEC follows from the work of Carnes and Shmoys [3].

A class of related problems are the density problems. In density problems we are to find a subgraph and the objective function considers the ratio of the total number or weight of edges in the subgraph to the number of vertices in the subgraph. A well known problem in this class is the Dense k-Subgraph problem (DkS) in which we want to find a subset of vertices U of size k such that the total number of edges in the subgraph induced by U is maximized. The best ratio known for the problem is $n^{1/4+\epsilon}$ [5, 2], which is an improvement over the bound of $O(n^{1/3-\epsilon})$, for ϵ close to $1/60$ [5]. The Dense k-Subgraph problem is APX-hard under the assumption that NP problems can not be solved in subexponential time [10]. If there is no bound on the the size of U then the problem is called The Densest Subgraph problem and can be solved in polynomial time [12, 7].

Variations of the Dense k-Subgraph problem in which the size of U is at least k ($Dalk$) and the size of U is at most k ($Damk$) have been studied [1, 11]. In [1, 11], they give evidence that $Damk$ is just as hard as DkS. They also give 2-approximate solutions to the $Dalk$ problem. In [11], they also consider the density versions of the problems in directed graphs. Gajewar and Sarma [6] consider a generalization in which we are give a partition of vertices $U_1, U_2, \ldots, U_t$, and non-negative integers $r_1, r_2, \ldots, r_t$. the goal is to find a densest subgraph such that partition U_i contributes at least r_i vertices to the densest subgraph. They give a 3-approximation for the problem, which was improved to 2 by Chakravarthy et al. [4], who also consider other generalizations. They also show using linear programming that the Degrees density augmentation problem can be solved optimally.

A problem parameterized by k is Fixed Parameter Tractable [13], if it admits an exact algorithm with running time of $f(k) \cdot n^{O(1)}$. The function f can be exponential in k or larger. Proving that a problem is W[1]-hard (with respect to parameter k) is a strong indication that it has no FPT

algorithm with parameter k (similar to NP-hardness implying the likelihood of no polynomial time algorithm). The FCEC problem parameterized by k is W[1] hard but admits a $f(k, \epsilon) \cdot n^{O(1)}$ time, $(1 + \epsilon)$-approximation, for any constant $\epsilon > 0$ [13]. This is in contrast to our result that shows that it is highly unlikely that FCEC admits a polynomial time approximation scheme (PTAS), if the running time is bounded by a polynomial in k.

1.2 Preliminaries

The set $E(U, W)$ is the set of edges with one endpoint in U and one in W. Let $e(U, W) = |E(U, W)|$. The set of edges with both endpoints in U, also called *internal* edges of U, is denoted by $E(U)$. We denote $e(U) = |E(U)|$. The input is an undirected simple graph $G = (V, E)$ and vertex weights are given by $w(\cdot)$. Let $n = |V|$ and $m = |E|$. For any subset $S \subseteq V$, let $\overline{S} = V \setminus S$. Let $deg(S)$ denote the sum of degrees of all vertices in S, i.e., $deg(S) = \sum_{v \in S} deg(v)$. Let $deg_H(v)$ denote the number of neighbors of v among the vertices in H. Let $deg_H(S)$ denote the quantity $\sum_{v \in S} deg_H(v)$. Let $T(U)$ be the collection of all edges with at least one endpoint in U. Namely, is the set of edges touching U. We denote $t(U) = |T(U)|$.

We use OPT to denote an optimal solution as well as the cost of an optimal solution. The meaning will be clear from the context in which it is used.

We observe that the unweighted FCEC problem admits a simple 2-approximate that just outputs the k least degree vertices.

Claim. For every set U, $t(U) = deg(U) - e(U)$

Proof. Consider separately the edges $E(U, V \setminus U)$ and $E(U)$. Note that the edges $E(U, V \setminus U)$ are counted once in the sum of degrees, but edges in $E(U)$ are counted twice. Thus in order to get the number of edges touching U, we need to subtract $e(U)$ from $deg(U)$.

1.3 Our results

Our contributions in this paper are as follows.

Theorem 1. *The* MWEC *problem admits a ratio* 2 *polynomial algorithm. Further, under the* Small set Expansion Conjecture *, the problem admits no* $2 - \epsilon$ *ratio for any constant* $\epsilon > 0$.

This improves both the guarantee of 3 given by Goldschmidt and Hochbaum [8] and their NPC proof. And for the first time relates the refutation of the Small set Expansion Conjecture to breaking the approximation threshold of a problem. Feige (in a private communication) stated that giving a $2 - \epsilon$ ratio to the $Dalk$ problem (that admits a ratio of 2 [1, 11]) refutes the Small set Expansion Conjecture and thus $Dalk$ behaves similarly to MWEC and FCEC .

Theorem 2. *The* FCEC *problem admits a ratio* 2 *polynomial time algorithm. Under the* Small set Expansion Conjecture *this ratio is tight, namely, the problem admits no better than* $2 - \epsilon$ *ratio for every constant* $\epsilon > 0$.

This improves the $2(1 + \epsilon)$-ratio of Carnes and Shmoys [3] and the NPC that follows from [8].

Theorem 3. *The natural LP for* FCEC *has an integrality gap of* $2(1 - o(1))$, *even for the unweighted case.*

We also provide the following results:

Theorem 4. *For any constant* $\rho > 1$, *we show that if* FCEC *admits a* ρ-*approximation algorithm then* MWEC *admits a* $\rho(1 + o(1))$-*approximation algorithm.*

Remark: It seems that using a ρ ratio for FCEC give a ρ ratio of MWEC is more complex to prove (if true).

Theorem 5. *The* Degrees density augmentation *problem admits a polynomial time exact algorithm using flow techniques*

2 A tight 2-approximation for Maximum Weight m'-Edge Cover and FCEC

In this section, we prove Theorems 1 an Theorem 2

2.1 A ratio 2 approximation for Maximum Weight m'-Edge Cover problem

In this subsection we give a dynamic programming based solution for the MWEC problem. The idea of using dynamic programming in this context was first proposed by Goldschmidt and Hochbaum [8]. Recall that in the MWEC problem, we are given an undirected simple graph $G = (V, E)$ with integral vertex weights. The goal is to select a subset $U \subseteq V$ of maximum weight so that the number of edges touching U is at most m'.

We will guess the following entities (by trying all possibilities) and for each guess, we use dynamic programming to solve the problem.

1. $H^* = \{v_h\}$, where v_h is the heaviest vertex in an optimal solution.
2. $P_{H^*} = e(H^*, OPT \setminus H^*)$ – the number of neighbors of v_h in the optimal solution. There are at most n possibilities.
3. $D_{H^*} = deg_{\overline{H}^*}(OPT \setminus H^*)$: total degree of vertices in $OPT \setminus H^*$ in the graph induced by vertices in $V \setminus H^*$. There are at most n^2 possibilities.

We will try all combinations of the above entities. Since there are at most polynomial number of possibilities for each entity, we have at most polynomial number of possibilities in total. We define the following subproblems as part of our dynamic programming solution. For a guess H for H^*, let $\{v_1, v_2, \ldots, v_{|\overline{H}|}\}$ be the vertices in $\overline{H}$. Then, any H, we solve the following subproblems.

$A[H, i, P_H, D_H]$ denote the maximum weighted subset $Q \subseteq \{v_1, v_2, \ldots, v_i\}$ such that $e(H, Q) \geq P_H$ and $deg_{\overline{H}}(Q) \leq D_H/2$.

Note that while the natural bound on $deg_{\overline{H}}(Q)$ is D_H, using such a bound will lead to an infeasible solution. For fixed parameters H, P_H, and D_H, we are interested in $A[H, |\overline{H}|, P_H, D_H]$. We use the following recurrence as the basis for our dynamic programming solution: the value of $A[H, i, P_H, D_H] = -\infty$ in any of the following three cases – (i) $i = 0$ and $P_H > 0$, (ii) $i = 0$ and $D_H/2 < 0$, and (iii) $D_H/2 > m' - e(H, \overline{H})$. When $i = 0$, $P_H \leq 0$ and $D_H/2 \geq 0$, the value of $A[H, i, P_H, D_H] = 0$. Otherwise, we have

$$A[H, i, P_H, D_H] = \max\{A[H, i - 1, P_H, D_H], w(v_i) + A[H, i - 1, P'_H, D'_H]\}$$

where, $P'_H = P_H - deg_H(v_i)$ and $D'_H = D_H - 2(deg_{\overline{H}}(v_i))$. Our solution is given by $\max_{H, P_H, D_H}\{w(H) + A[H, |\overline{H}|, P_H, D_H]\}$.

Analysis

Lemma 1. *Our algorithm yields a feasible solution.*

Proof. Let $H' \cup Q'$, where $Q' \subseteq V \setminus H'$, be the set of vertices returned by our solution. The number of edges with at least one endpoint in $H' \cup Q'$, is

$$
\begin{aligned}
&= e(H', \overline{H}') + e(Q', \overline{H}') \\
&\leq e(H', \overline{H}') + deg_{\overline{H}'}(Q') \\
&\leq e(H', \overline{H}') + \frac{D_{H'}}{2} \\
&\leq e(H', \overline{H}') + (m' - e(H', \overline{H}')) \qquad \text{(using the base case)} \\
&= m'
\end{aligned}
$$

Lemma 2. *The above algorithm results in a 2-approximate solution.*

Proof. Recall that H^* consists of the highest degree vertex in the optimal solution. Let Q^* be the remaining vertices in the optimal solution. Consider the scenario when our algorithm makes the correct guess for H^*. Let $Q \subseteq \overline{H^*}$ be the solution returned by the dynamic program in this setting. We know that

$$
deg_{\overline{H}^*}(Q) \leq \frac{deg_{\overline{H}^*}(Q^*)}{2}
$$

We now use ideas from [8] to show that $w(H^* \cup Q) \geq 2w(H^* \cup Q^*)$. Recall that $H' \cup Q'$ be the output of our algorithm. Since $w(H' \cup Q') \geq w(H^* \cup Q^*)$, it follows that our solution is a factor of at most 2 away from OPT.

Consider any arbitrary ordering of vertices $v_1, v_2, \ldots$ in Q^*. Note that the weight of each vertex in Q^* is at most $w(H^*)$. Let Q_r^* denote the the first r vertices in the above ordering of vertices of Q^*. Let p be the first index such that $deg_{\overline{H}^*}(Q_p^*) > deg_{\overline{H}^*}(Q^*)/2$. This implies the following – (i) $deg_{\overline{H}^*}(Q_{p-1}^*) \leq deg_{\overline{H}^*}(Q^*)/2$, and (ii) $deg_{\overline{H}^*}(Q^* \setminus Q_p^*) < deg_{\overline{H}^*}(Q^*)/2$. Note that both the sets Q_{p-1}^* and $Q^* \setminus Q_p^*$ (neither set contains v_p) are feasible candidates for the set Q, the solution returned by our algorithm when the heaviest vertex set was chosen to be H^*. Since $w(Q) \geq w(Q_{p-1}^*)$, $w(Q) \geq w(Q^* \setminus Q_p^*)$, and $w(v_p) \leq w(H^*)$, we have

$$
\begin{aligned}
w(OPT) &\leq w(H^* \cup Q^*) \\
&\leq w(H^*) + w(Q^*) \\
&\leq w(H^*) + w(Q_{p-1}^*) + w(v_p) + w(Q^* \setminus Q_p^*) \\
&\leq w(H^*) + w(Q) + w(H^*) + w(Q) \\
&= 2w(H^* \cup Q) \\
&\leq 2w(H' \cup Q')
\end{aligned}
$$

2.2 A 2-approximation for Fixed Weight Minimum Edge Cover

Recall the FCEC problem: Given a graph $G = (V, E)$ with arbitrary vertex weights and a positive integer W, our objective is to choose a set $S \subseteq V$ of vertices of total weight at least W such that the number of edges with at least one end point in S is minimized.

We will solve the following related problem optimally and then show that an optimal solution to the problem is a 2-approximation to FCEC : we want to find a subset S of vertices such that $deg(S)$ is smallest and $w(S)$ is at least W.

We use the dynamic programming algorithm of the well-known Knapsack problem to find a solution to the above problem. For completeness, we restate the dynamic programming formulation below.

$P[i, D]$: maximum weight of set $Q \subseteq \{v_1, v_2, \ldots, v_i\}$ such that $deg(Q)$ is at most D.

Note that $P[0, D] = 0$, for all values of D is the base case. For all other case, we invoke the following recurrence.

$$P[i, D] = \max\{P[i - 1, D], w(v_i) + P[i - 1, D - w(v_i)]\}$$

After filling the table P using dynamic programming, we scan all entries of the form $P[|V|, D]$ to find the smallest value of D for which $P[|V|, D] \geq W$. Let S be the corresponding set.

Lemma 3. *The is a 2-approximate solution to the Fixed Cost Minimum Edge Cover Problem as follows.*

$$t(S) \leq deg(S) \leq deg(OPT) = 2(deg(OPT)/2) \leq 2OPT$$

2.3 A $2 - \epsilon$ inapproximability for FCEC and under the Small set Expansion Conjecture

Theorem 6. *If unweighted* FCEC *admits better than a* $2 - 6\mu$ *approximation, then the* Small set Expansion Conjecture *fails.*

Proof. Consider the first case that $\phi_G(\delta) \geq 1 - \mu$, and let S be the set achieving this bound. We guess the size of S by trying all possible sizes not larger than $\delta \cdot n$. Let k be the correct size of S.

As the graph is d regular, we get that:

$$|E(S, V - S)| \geq (1 - \mu)deg(S) = (1 - \mu)k \cdot d.$$

As $t(S) \geq |E(S, V - S)|$ we get that $t(S) \geq (1 - \mu) \cdot k \cdot d$.

The second case is that

$$|E(S, V - S)| \leq \mu \cdot deg(S) = \mu \cdot k \cdot d.$$

Observe that:

$$deg(S) = d \cdot k \leq \mu \cdot k \cdot d + 2e(S)$$

$$\text{Therefore } 2e(S) \geq dk(1 - \mu).$$

$$\text{And } e(S) \geq dk(1 - \mu)/2$$

$$\text{Thus } t(S) = deg(S) - e(S) \leq k \cdot d(1 + \mu)/2$$

The ratio between a yes and a no instance is:

$$\frac{(1 - \mu)k \cdot d}{(1 + \mu)/2 \cdot k \cdot d} = 2 \cdot \frac{1 - \mu}{1 + \mu} \geq 2 \cdot (1 - 3\mu).$$

This means that if FCEC admits an approximation ratio smaller than $2(1 - 3\mu)$ the Small set Expansion Conjecture is disproved. We can pick $\epsilon = \mu/6$ instead of μ and get the $2 - \epsilon$ inapproximability. As we can choose any small ϵ, (the conjecture allows us to chose μ as small constant as we want) the inapproximability can be made close to 2 almost matching the lower bound.

2.4 A $2 - \epsilon$ inapproximability for Maximum Weight m'-Edge Cover under the Small Set Expansion Conjecture

Recall that in our hardness example for FCEC for hardness $2 - \epsilon$, the graph is d regular.

For simplicity, we ignore the small constant ϵ in this section. We will assume 2 inapproximability for FCEC and show that better than 2 approximation for MWEC implies a better than 2 approximation for FCEC . Making the hardness $2 - \epsilon$ is elementary.

Thus for a yes instance we assume that $t(opt) = k \cdot d/2$. And that for a no instance it is kd. Consider a yes instance I for FCEC . Let OPT be the optimum solution. Thus if you have a better than 2 approximation for MWEC it will return $t(opt)$ vertices with *strictly more than $k/2$ vertices*. The amount of vertices still required to be added to transform S to a legal FCEC output is *strictly less than $k - k/2 = k/2$*. We can complete the set S to size k by *any* set S' of $k/2$ vertices. In such case $t(S \cup S') < kd/2 + kd/2 = kd$ and thus we will know that the instance is a yes instance. Thus we can tell between a no and a yes instance of FCEC , contradiction.

3 Integrality gap for Fixed Cost Minimum Edge Cover

Consider the following natural integer linear program for the problem

$$\min \sum_e y_e$$

$$\text{subject to} \quad \sum_{v \in V} x_v \geq k,$$

$$y_e \geq x_u, \qquad\qquad \forall e = (u, v)$$

$$y_e \geq x_v, \qquad\qquad \forall e = (u, v)$$

$$x_v \in \{0, 1\}, \qquad\qquad \forall v \in V$$

$$y_e \in \{0, 1\}, \qquad\qquad \forall e \in E$$

The LP relaxation can be obtained by relaxing the integrality constraints on x_v and y_e to $x_v \geq 0, \forall v \in V$ and $y_e \geq 0, \forall e \in E$.

Theorem 7. *The above LP has an integrality gap of $2(1 - o(1))$.*

Let $k = \lfloor \sqrt{n} \rfloor$. Construct a graph G on n vertices as follows. For each pair of vertices, include an edge between the pair with a probability $1/\lfloor \sqrt{n} \rfloor$. For any vertex v, $\mathbf{E}[deg(v)] = n(1/\lfloor \sqrt{n} \rfloor) \leq \lceil \sqrt{n} \rceil$. Using Chernoff bounds, for $0 < \delta < 1$, we have

$$\sqrt{n}(1 - o(1)) \leq deg(v) \leq \sqrt{n}(1 + o(1))$$

Consider any subset Q of vertices in G such that $|Q| = \lfloor \sqrt{n} \rfloor$. Then we have

$$\mathbf{E}[e(Q)] = \frac{1}{\lfloor \sqrt{n} \rfloor} \binom{Q}{2} = \frac{\lfloor \sqrt{n} \rfloor (\lfloor \sqrt{n} \rfloor - 1)}{2 \lfloor \sqrt{n} \rfloor} = \frac{\lfloor \sqrt{n} \rfloor - 1}{2}$$

Thus, $n \geq 4$, we have $\sqrt{n}/4 \leq \mathbf{E}[e(Q)] < \sqrt{n}/2$. We use the following Chernoff bound to obtain the probability that $e(Q) \geq n^{1-\epsilon}$, for a constant ϵ.

$$\Pr\left[[]e(Q) \geq (1 + \delta)\mathbf{E}[e(Q)]\right] \leq \left(\frac{\exp(\delta)}{(1 + \delta)^{(1+\delta)}}\right)^{\mathbf{E}[e(Q)]}$$

In our case, $2n^{1/2-\epsilon} \leq 1 + \delta \leq 4n^{1/2-\epsilon}$, thus we get

$$\Pr\left[[]e(Q) \geq n^{1-\epsilon}\right] \leq \left(\frac{\exp(4n^{1/2-\epsilon})}{(2n^{1/2-\epsilon})^{2n^{1/2-\epsilon}}}\right)^{\sqrt{n}/4}$$

Let $f(n,\epsilon) = \left(\frac{\exp(n^{1/2-\epsilon})}{(2n^{1/2-\epsilon})^{(n^{1/2-\epsilon}/2)}}\right)^{\sqrt{n}}$. The number of sets of size $\lfloor\sqrt{n}\rfloor$ is given by $\binom{n}{\sqrt{n}} \leq (ne/\lfloor\sqrt{n}\rfloor)^{\sqrt{n}} = (\lceil\sqrt{n}\rceil e)^{\sqrt{n}}$. The probability that there is no subset of size $\lfloor\sqrt{n}\rfloor$ that has at least $n^{1-\epsilon}$ edges is given by the union-bound as follows

$$f(n,\epsilon)\binom{n}{\sqrt{n}} << 1$$

The number of edges with at least one end point in Q is given by

$$t(Q) = deg(Q) - e(Q)$$
$$\geq \lfloor\sqrt{n}\rfloor \cdot \sqrt{n}(1 - o(1)) - n^{1-\epsilon}$$
$$= n(1 - o(1))$$

On the other hand, consider the fractional solution in which $x_v = 1/\sqrt{n}$, for each v and $y_e = 1/\sqrt{n}$, for each $e \in E$. This LP solution is feasible and has a cost of $|E|/\sqrt{n}$. The number of edges $|E| = n\sqrt{n}/2(1 + o(1))$. Thus the cost of the LP solution is at most $n(1 + o(1))/2$, which results in a gap of $2(1 - o(1))$.

4 An approximation for Fixed Cost Minimum Edge Cover implies the same approximation for Maximum Weight m'-Edge Cover

We first transform the input instance for the MWEC problem to one in which the optimum value of the objective function is at most n^5 by paying a very small penalty in the approximation ratio.

Lemma 4. *For the Maximum weight m'-subgraph problem, we can convert the input instance $\langle G, w, m'\rangle$, with an optimal solution denoted by OPT into an instance $\langle G', w', m'\rangle$, with optimal solution OPT'', such that $OPT'' \leq n^5$. Furthermore, if OPT' is the total weight of the vertices in OPT'' under the weight function w, then*

$$OPT' \geq OPT(1 - 1/n)(1 - 1/n^2)$$

Proof. Let $v_1, v_2, \ldots, v_n$ be the vertices in G such that $w(v_1) \geq w(v_2) \geq \cdots \geq w(v_n)$. Let v_p be the last vertex in the ordering such that $w(v_p) \geq w(v_1)/n^2$. In other words, for each j, $p < j \leq n$, $w(v_1) > n^2w(v_j)$. Let G' is the graph induced on vertices $v_1, v_2, \ldots, v_p$. Let OPT_1 be the optimal solution for the instance $\langle G', w, m'\rangle$. Note that OPT may choose some vertices from the set $\{v_{p+1}, v_{p+1}, \ldots, v_n\}$. The error incurred in not considering these vertices is at most $n(w(v_1)/n^2) \leq OPT/n$. Thus we get

$$OPT_1 \geq OPT(1 - 1/n)$$

We now scale the weights of vertices in G' to create an instance $\langle G', w', m'\rangle$, where

$$w'(v_j) = \left\lfloor\left(\frac{w(v_j)}{w(v_p)}\right)n^2\right\rfloor$$

Let OPT'' be an optimal solution to $\langle G', w', m' \rangle$. Clearly, $OPT'' \leq n^5$. Let OPT' be the cost of the solution OPT'' under the weight function w, i.e., $OPT' = \sum_{v \in OPT''} w(v)$. Thus we have

$$OPT' \geq OPT_1 \left(1 - \frac{1}{n^2}\right) \geq OPT \left(1 - \frac{1}{n}\right) \left(1 - \frac{1}{n^2}\right)$$

Theorem 8. *For some constant α, an α approximation guarantee for* FCEC *implies an $\alpha(1 + o(1))$ approximation guarantee for* MWEC *.*

Proof. Suppose that we have an $\alpha > 1$ approximation algorithm for FCEC , for some constant α. Using Lemma 4, we transform the MWEC instance (G, m') with an optimal weight W^* to an instance in which the optimum weight $W^* \leq n^4$. This increase the approximation ratio by a factor of only $(1 + o(1))$. We now consider the modified instance (G', m') as an input to FCEC . We guess the value of W^* by trying all possible integral values between 1 and n^4. For each guess of W^*, we apply the α-approximation algorithm for FCEC to the new instance. When our guess W^* is correct and we apply the algorithm, we obtain a set U of vertices of cost at least W^* and that touch at most $\alpha \cdot m'$ edges.

Create a new set B in which every vertex from U is chosen with a probability $1/\alpha$. We say that an edge e is *deleted* if $e \notin E(B)$. Let τ be a constant.

We consider the following "bad" events: (i) $w(B) \leq W^*/((1+\tau)\alpha)$, (ii) $t(B) > m'$.

We first bound the probability that $w(B) \leq W^*/((1+\tau)\alpha)$. The expected cost of B is $w(U)/\alpha = W^*/\alpha$. Consider the expected cost of $U \setminus B$. The expected cost is $W^* - W^*/\alpha$. The event that $w(B) \leq W^*/(\alpha(1+\tau))$ is equivalent to the event $w(U) - w(B) \geq W^* - W^*/(\alpha(1+\tau)) = W^*(1 - 1/(\rho(1+\tau))$. By the Markov's inequality, the last event has probability at most $(1 - 1/\alpha)/(1 - 1/(\alpha(1+\tau)) = 1 - \tau/(\alpha + \alpha \cdot \tau - 1)$

We now bound the probability of the second bad event. The expected number of edges in $E(B)$ is at most $m'(1 - (1 - \frac{1}{\alpha})^2)$. Note that the events that edges are deleted are positively correlated because given that an edge (v, u) is deleted, one of the possibilities that can cause this event, is that v is deleted, and in that case all edges of v are deleted with probability 1. Clearly, we can assume that $m' \geq c$ for any constant c. Otherwise, we can solve the MWEC problem in polynomial time by checking all subsets of edges. By the Chernoff bound, the probability that the number of edges is more than m' is bounded by $exp(-c\delta^2/2)$, for some $\delta < 1$. We can choose a large enough c so that the above probability is at most $\tau/(2(\alpha + \alpha \cdot \tau - 1))$. This would mean that the sum of probabilities of bad events is strictly smaller than 1. This construction can be derandomized by the method of conditional expectations.

5 Exact algorithm for the Degrees Density Augmentation Problem

The Degrees density augmentation problem is as follows: Given a graph $G = (V, E)$ and a subset $U \subseteq V$, the objective is to find a subset $W \subseteq V \setminus U$ such that

$$\rho = \frac{e(W) + e(U, W)}{deg(W)} \text{ is maximized}$$

The Degrees density augmentation problem is related to the FCEC problem in the same way as the Densest subgraph problem is related to the Dense k-subgraph problem. A natural heuristic for the FCEC problem would be to iteratively find a set W with good augmentation degrees density. A polynomial time exact solution for the problem using linear programming is given in [4]. Here we present a combinatorial algorithm.

We solve the Degrees density augmentation problem exactly by finding minimum s-t cut in the flow network constructed as follows. Let $\overline{U}$ denote the set $V \setminus U$. In addition to the source s and the

sink t, the vertex set contains $V_{E'} \cup \overline{U}$, where $V_{E'} = \{v_e \,|\, e \in E$ and both end points of e are in $\overline{U}\}$. There is an edge from s to every vertex in $V_{E'} \cup \overline{U}$. If a is a vertex in $V_{E'}$ then the capacity of the edge (s, a) is 1, otherwise, the capacity of the edge is $deg_U(a)$. For each vertex $v_e \in V_{E'}$, where $e = (p, q)$, there are edges (v_e, p) and (v_e, q). Each such edge has a large capacity of $M = \infty$ (any capacity of at least n^5 would work). Finally, each vertex $p \in \overline{U}$ is connected to t and has a capacity of $\rho \cdot deg(p)$.

5.1   Algorithm

For a particular value of ρ, let $W_s \subseteq \overline{U}$ be the vertices that are on the $s(t)$ side of a minimum s-t cut. Let $V_{E'}^s \subseteq V_{E'} (V_{E'}^t \subseteq V_{E'})$ be the vertices in $V_{E'}$ that are on the $s(t)$ side of the minimum s-t cut. We now state the algorithm.

1. Construct the flow network as shown above.
2. For each value of ρ, compute a minimum s-t cut and find the resulting value of $e(W_s) + e(U, W_s) - \rho deg(W_s)$. Find the largest value of ρ for which the expression is at least 0.
3. Return W_s corresponding to the largest value of ρ.

5.2   Analysis

Lemma 5. *Any minimum s-t cut in the above flow network has capacity at most $2n^2$.*

Proof. This follows because the s-t cut $(s, V \setminus \{s\})$ has capacity at most $2n^2$.

Lemma 6. *For any minimum s-t cut C, $|V_{E'}^s| = e(W_s)$.*

Proof. Note that it cannot be the case that $|V_{E'}^s| > e(W_s)$, as this will result in the capacity of the cut C being at least M, which is not possible by Lemma 5. Note that any s-t cut for which $|V_{E'}^s| < e(W_s)$ can be transformed into another s-t cut of a lower capacity in which $|V_{E'}^s| = e(W_s)$ by moving vertices in $V_{E'}^t$ that correspond to edges in W_s to $V_{E'}^s$. Since edges from s to vertices in $V_{E'}$ (vertices in $V_{E'}^t$, in particular) have capacity of 1, the capacity of the cut reduces. The claim follows.

Lemma 7. *The Degrees Density Augmentation problem admits a polynomial time exact solution.*

Proof. We are interested in finding a non-empty set $W_s \subseteq \overline{U}$ such that $\frac{e(W_s) + e(U, W_s)}{deg(W_s)}$ is maximized . Note that there are at most $2n^4$ possible values of ρ that our algorithm needs to try. Indeed, the numerator is an integer between 1 and $2n^2$ and the denominator is an integer between 1 and n^2. Since minimum s-t cut can be computed in polynomial time, our algorithm runs in polynomial time.

For any fixed guess for ρ, the capacity of the min s-t cut is given by

$$\min_{W_s \subseteq \overline{U}} |V_{E'}^t| + deg_U(W_t) + \rho deg(W_s)$$

$$= \min_{W_s \subseteq \overline{U}} |V_{E'}| - |V_{E'}^s| + deg_U(\overline{U}) - deg_U(W_s) + \rho deg(W_s)$$

$$= |V_{E'}| + deg_U(\overline{U}) - \max_{W_s \subseteq \overline{U}} |V_{E'}^s| + deg_U(W_S) - \rho deg(W_s)$$

$$= |V_{E'}| + deg_U(V \setminus U) - \max_{W_s \subseteq \overline{U}} e(W_s) + e(U, W_S) - \rho deg(W_s) (\text{ using Lemma 6})$$

Our algorithm ensures that $\rho deg(W_s) \geq e(W_s) + e(U, W_S)$, which eliminates the possibility of $W_s = \emptyset$. Thus, finding the minimum s-t cut for a fixed ρ in the above flow network is equivalent to finding a set W_s with the largest degree density. Thus we have

$$\frac{e(W_s) + e(U, W_s)}{deg(W_s)} \geq \rho$$

Since our algorithm finds such W_s for each possible fraction that ρ can assume and returns the W_s with the highest degree density, our solution is optimal.

Acknowledgements

References

1. R. Andersen and K. Chellapilla. Finding dense subgraphs with size bounds. In *WAW*, pages 25–36, 2009.
2. Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities: an O($n^{1/4}$) approximation for densest k-subgraph. In *STOC*, pages 201–210, 2010.
3. T. Carnes and D. Shmoys. Primal-dual schema for capacitated covering problems. In *IPCO*, pages 288–302, 2008.
4. V.T̃. Chakravarthy, N. Modani, S.R̃. Natarajan, S. Roy, and Y. Sabharwal. Density functions subject to co-matroid constraint. In *Foundations of Software Technology and Theoretical Computer Science*, pages 236–248, 2012.
5. U. Feige, G. Kortsarz, and D. Peleg. The dense k-subgraph problem. *Algorithmica*, 29(3):410–421, 2001.
6. A. Gajewar and A. Sarma. Multi-skill collaborative teams based on densest subgraphs. In *SDM*, pages 165–176, 2012.
7. A. V. Goldberg. Finding a maximum density subgraph. Technical Report UCB/CSD-84-171, EECS Department, University of California, Berkeley, 1984.
8. Olivier Goldschmidt and Dorit S. Hochbaum. k-edge subgraph problems. *Discrete Applied Mathematics*, 74(2):159–169, 1997.
9. Subhash Khot. On the unique games conjecture. In *FOCS*, page 3, 2005.
10. Subhash Khot. Ruling out ptas for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM J. Comput.*, 36(4):1025–1071, 2006.
11. S. Khuller and B. Saha. On finding dense subgraphs. In *ICALP*, pages 597–608, 2009.
12. Eugene L. Lawler. *Combinatorial optimization: networks and matroids*. Holt, Rinehart and Winston, New York, 1976.
13. Dániel Marx. Parameterized complexity and approximation algorithms. *Comput. J.*, 51(1):60–78, 2008.
14. Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *STOC*, pages 755–764, 2010.