

# On the Approximability of DAG Edge Deletion

Nathan Klein  
Advisor: Tom Wexler

## Abstract

The DAG EDGE DELETION problem of  $k$ , or  $\text{DED}(k)$ , is to delete the minimum weight set of edges from a directed acyclic graph such that the remaining graph has no path of length  $k$ . In 2012, Svensson showed it was hard to approximate the vertex deletion version with a ratio better than  $k$  assuming the Unique Games Conjecture is true [14], and in 2015 Kenkre et al. introduced the  $\text{DED}(k)$  problem and used this result to show that  $\text{DED}(k)$  is UGC-hard to approximate better than  $\lfloor \frac{k}{2} \rfloor$  for  $k \geq 4$  [6]. They motivate it as the minimization version of the problem  $\text{MAX-}k\text{-ORDERING}$  on DAGs, which is to assign a labeling  $\ell: V \rightarrow [k]$  that maximizes the number of edges  $(i, j)$  with  $\ell(i) < \ell(j)$ . This in turn has applications to scheduling jobs with soft precedence constraints. However, the best known approximation algorithm for  $\text{DED}(k)$ , given by Kenkre et al. [6], has a ratio of  $k$ . In this work we tighten this gap by giving a  $(\frac{2}{3}k + O(1))$ -approximation for  $\text{DED}(k)$ , as well as a 1.375-approximation for  $\text{DED}(2)$ , a 2-approximation for  $\text{DED}(3)$ , and a  $\frac{3}{4}k$ -approximation for small  $k$ .

We then address a combinatorial problem of finding a class of DAGs for which the  $\text{MAX-}k\text{-ORDERING}$  is as small as possible. We show that a DAG given by Alon et al. [1] that was shown to have a maximum directed cut of size at most  $(\frac{1}{4} + o(1))|E|$  in fact has a  $\text{MAX-}k\text{-ORDERING}$  of at most  $(\frac{k-1}{2k} + o(1))|E|$ . From this result follows a  $\frac{k+1}{2}$  integrality gap for Kenkre et al.'s LP, nearly matching the UGC-hardness factor. We state finding either a  $\frac{k+1}{2}$  rounding for this LP or an improved integrality gap as an open problem.

## 1 Introduction

The DAG VERTEX DELETION (DVD) problem is as follows: given a parameter  $k$  and a DAG  $G = (V, E)$ , find the smallest set  $S \subseteq V$  such that the induced graph on  $V \setminus S$  has no paths of length  $k$ . In this problem, the length of a path is measured by the number of nodes it contains. It was initially introduced by Paik et al. [12] in 1994. They gave efficient algorithms for special cases of DAGs and proved that on general DAGs the problem was  $\mathcal{NP}$ -Complete. They motivated the problem with VSLI design, as circuits are often modeled as DAGs. To ensure the loss of electric signal when traversing a circuit does

not exceed a threshold, we may desire to delete (or upgrade) the minimum number of components so that no signal travels through more than  $k$  lossy nodes; this is the DVD( $k$ ) problem.

The DAG EDGE DELETION (DED) problem is identical except it asks for the minimal set  $S \subseteq E$  such that  $G' = (V, E \setminus S)$  has no paths of length  $k$ . Here the length of a path is the number of edges it contains. DED( $k$ ) is the minimization version of the MAX- $k$ -ORDERING problem (finding a labeling  $\ell: V \rightarrow [k]$  such that the number of edges from a smaller label to larger label is maximized) on *acyclic* graphs, as shown by Kenkre et al. [6].

We will denote these problems with parameter  $k$  as DVD( $k$ ) and DED( $k$ ). We will focus on the DED( $k$ ) problem. It is motivated by scheduling jobs with soft precedence constraints. In this application, jobs are vertices, and an edge from vertex  $i$  to  $j$  corresponds to the constraint that job  $i$  is completed before we can process job  $j$ . However, in order to complete all jobs in  $k$  time steps it may not be possible to satisfy all constraints. In this setting, the DED( $k$ ) problem models completing all jobs within  $k$  steps while violating the fewest number of constraints. In the VSLI domain, this would correspond to deleting the minimum number of connections between components to ensure that no path through the circuit takes more than  $k$  steps.

The best known approximation algorithm for DVD identifies paths and removes all involved vertices until no paths of length  $k$  remain (or, in the weighted case, running an LP and rounding with threshold  $\frac{1}{k}$ ). Both algorithms clearly have an approximation ratio of  $k$  (in fact, it is a special case of  $k$ -UNIFORM HYPERGRAPH COVER, for which there is a  $k$ -approximation). Similarly, the best known approximation for DED( $k$ ) from Kenkre et al. [6] also achieves ratio  $k$  by removing every edge in an identified path, or performing a  $\frac{1}{k}$  LP threshold rounding for the weighted case. Kenkre et al. also give a combinatorial local ratio algorithm with ratio  $k$  for the weighted case.

In this work, we first prove that the natural LP for any strict  $k$ -CSP that takes the OR of its variables pays at least  $\frac{1}{k}|L|$ , where  $L$  is the set of variables given non-zero value in any basic feasible solution. We use this to show a  $(\frac{2}{3}k + \frac{2}{3})$ -approximation algorithm for general  $k$  by giving an improved rounding for the linear program from Kenkre et al. We then give a 1.375-approximation for DED(2); for DED(3), a 2-approximation, and for  $4 \leq k \leq 7$  a  $\frac{3}{4}k$ -approximation. We also study NP and UGC-Hardness results for DED(2).

Subsequently, we demonstrate that a graph given by Alon et al. [1] that was shown to have a maximum directed cut of size at most  $(\frac{1}{4} + o(1))|E|$  in fact has a MAX- $k$ -ORDERING of at most  $(\frac{k-1}{2k} + o(1))|E|$ , using techniques similar to those from Alon et al. This immediately implies an integrality gap of size  $\frac{k+1}{2}$  for the LP from Kenkre et al. We state closing this gap as an open problem.

## 1.1 Related Work

Kenkre et al. [6] defined the DAG EDGE DELETION problem in 2015, and so it is not surprising that we have found no other work on this problem. As noted above, they give a  $k$ -approximation for the problem. They motivated it as the minimization of the MAX- $k$ -ORDERING problem, for which they gave a 2-approximation. However, the DAG VERTEX DELETION problem has been studied previously. In 2012, Svensson gave a  $k$ -hardness result for DVD( $k$ ) assuming the UGC [14], and Kenkre et al. used this result to show that DED( $k$ ) is  $\lfloor \frac{k}{2} \rfloor$ -UGC hard [6].

Svensson also demonstrated that showing the unconditional  $\mathcal{NP}$ -Hardness of DVD for any factor growing with  $k$  would in fact show that the DISCRETE TIME-COST TRADEOFF PROBLEM (DTCT) has no constant approximation ratio [14]. The natural reduction from Kenkre et al. shows that this is true of DED as well. This suggests progress in showing the  $\mathcal{NP}$ -Hardness of approximating DED( $k$ ) may be difficult, as DTCT has been studied extensively, yet no unconditional hardness result better than 1.36 has been found. The 1.36-hardness of DVD( $k$ ) when  $k \geq 2$  follows from the fact that VERTEX COVER is a special case of this problem for  $k = 2$ : construct a DAG from the given graph by giving the vertices a strict order and directing every edge from low to high. VC has been shown to have unconditional hardness of 1.36 [5], immediately giving the result for DVD. Kenkre's gadget reduction from DVD( $2k$ ) to DED( $k$ ) gives 1.36-hardness for DED( $k$ ) when  $k \geq 4$ .

We note that the acyclicity of the graph is needed when  $k$  is given as input, as any  $\alpha$ -approximation for the minimization version of MAX- $k$ -ORDERING on general digraphs could decide the DIRECTED HAMILTONIAN PATH problem, while finding paths of any length in a DAG admits a polynomial time algorithm. However, when  $k$  is fixed, the vertex deletion version on undirected graphs has been studied, as there are  $O(n^k)$  vertex collections to check for paths (in fact, an  $O(2^k \cdot \text{poly}(n, k))$  algorithm exists for this problem [18]). This problem on undirected paths is known as  $k$ -PATH COVER and was introduced by Brešar et al. in 2011 [3], who gave bounds for cover sizes, as well as noted that the problem has a linear-time solution on graphs with bounded treewidth. They mentioned the natural  $k$ -approximation. Subsequently, Tu and Zhou gave a 2-approximation for 3-PATH COVER [16], but since it appears no other approximation results have been given.

A similar result from Paik et al. from 1994 shows that DVD is solvable in polynomial time on rooted trees and series parallel graphs [12]. We can construct a simple reduction from DED to DVD that preserves series parallel and tree properties, so that DED has efficient algorithms on these graphs as well. Finally, it is useful to notice that DED( $k$ ) is fixed parameter tractable. If  $d$  is the query of the decision problem (i.e. is there a DED( $k$ ) solution of size  $d$ ?), then there is an exact recursive algorithm that runs in roughly  $O(k^d)$  time that branches on each edge of an identified path.

$k$	Approximation	UGC-Hardness	$\mathcal{NP}$ -Hardness	Integrality Gap (LP)
2	1.375	APX-Hard	APX-Hard	$\frac{3}{2}$
3	2	APX-Hard	APX-Hard	2
4-7	$\frac{3}{4}k$	$\lfloor \frac{k}{2} \rfloor$	1.36	$\frac{k+1}{2}$
$\geq 8$	$\frac{2}{3}k + \frac{2}{3}$	$\lfloor \frac{k}{2} \rfloor$	1.36	$\frac{k+1}{2}$

Table 1: Results

We summarize our results in Table 1. Note that the 1.375-approximation for DED(2) is able to beat the integrality gap as it uses an SDP-based result.

## 2 Preliminaries

Here we give a few definitions, as well as show that DED( $k$ ) is the minimization version of MAX- $k$ -ORDERING.

**Approximation Ratio:** For a minimization problem, we say a polynomial time algorithm  $A$  achieves an approximation ratio of  $\alpha$  if for all instances the ratio of the solution produced by  $A$  is at most  $\alpha \cdot OPT$ . For a maximization problem,  $A$  achieves ratio  $\alpha$  if for all instances the solution produced by  $A$  is at least  $\alpha \cdot OPT$ .

**Inapproximability Ratios and the Unique Games Conjecture:** A problem is *hard to approximate* better than a ratio  $\alpha$  if it can be shown that the existence of an approximation algorithm with ratio better than  $\alpha$  implies  $\mathcal{P} = \mathcal{NP}$ . A problem is UGC-hard to approximate better than a ratio  $\alpha$  if such an algorithm does not exist unless the Unique Games Conjecture is false. This conjecture was stated by Subhash Khot in 2002 [7]. It is often the case that when satisfactory  $\mathcal{NP}$ -Hardness of approximation results cannot be found, the UGC is used instead. It implies tight approximation ratios for a range of problems such as VERTEX COVER and MAX CUT.

**$k$ -CSP:** A  $k$ -CSP is a Constraint Satisfaction Problem in which every constraint consists of  $k$  variables. A *strict* CSP will involve satisfying all constraints while minimizing a quantity (as in DED( $k$ ) or VERTEX COVER). Other CSPs, however, will involve maximizing the number of satisfied constraints (such as MAX 3-SAT).

**Threshold Rounding:** A threshold rounding of  $c$  of a linear program for a minimization problem takes all variables in a basic feasible solution with value at least  $c$  and sets them to 1. This preserves a ratio of  $\frac{1}{c}$ .

**DED( $k$ ) is the minimization version of MAX- $k$ -ORDERING on DAGs:** It is useful to consider this problem in both contexts, so we reproduce this claim from Kenkre et al. [6]

here.

*Proof.* Given an MAX- $k$ -ORDERING instance on a DAG  $G$ , run DED( $k$ ) on  $G$ . If DED( $k$ ) has a solution of size  $c$ , then we can delete  $c$  edges to remove all paths of length  $k$  in the graph. To construct a MAX- $k$ -ORDERING solution, we will keep all the edges not removed. This is of size  $|E| - c$ . It only remains to show that we can assign vertices a labeling  $\ell: V \rightarrow [k]$  such that each edge goes forward. To do so, let  $\ell(v)$  be equal to the length of the longest path from any source to it. Since  $G$  is acyclic, all edges  $(u, v)$  have  $\ell(u) < \ell(v)$ , and since no paths exceed length  $k$  we have used at most  $k$  labels.

For the other direction, simply remove all edges not in the MAX- $k$ -ORDERING solution. As there are at most  $k$  labels and each path goes from a smaller label to a higher label, clearly no path can exceed length  $k$ .  $\square$

### 3 Graphs with small MAX $k$ -ORDERING

in 2007 Alon et al. found a class of DAGs whose maximum directed cut was  $\frac{|E|}{4} + O(|E|^{4/5})$  [1]. This implies that the best DED(2) solution tends to  $\frac{3}{4}|E|$  on this class. Here (following the analysis closely used in Alon et al.'s paper) we will show that Alon et al.'s graph in fact has a MAXIMUM  $k$ -ORDERING solution of size at most  $(\frac{k-1}{2k} + o(1))|E|$  for all  $k$  (not just  $k = 2$ ), and thus the best DED( $k$ ) solution is of size at least  $(\frac{k+1}{2k} - o(1))|E|$ .

The construction we state is identical to Alon et al.'s [1]. We reproduce this here.

However, first note that fundamentally we want to design a Eulerian tournament that loses few edges to become a DAG. It follows that a Eulerian tournament  $G$  produced from a  $K_n$  has a small MAX  $k$ -ORDERING since by Turan's theorem [17] the best ordering on the (undirected) underlying graph produces a solution of size  $\binom{k}{2} \binom{n/k}{2}$ , and since  $G$  is Eulerian the optimal ordering on the directed graph is at most half of this, as every cut  $(A, B)$  must have an equal number of edges going from  $A$  to  $B$  as go from  $B$  to  $A$ . The ratio of these cuts tend to  $\frac{|E|}{4}$ . Thus, the goal is to design a similar graph that can easily be made a DAG. This is Alon et al.'s construction:

By a theorem of Singer from [13], there exists a set  $A$  of  $r$  natural numbers such that for all  $a, b \in A$  with  $a \neq b$ , we have  $a - b$  distinct. In addition, the maximal element of  $A$  is of size at most  $(1 + o(1))r^2$ . Take any  $n \geq 1$  and let  $r = \lfloor n^{1/3} \rfloor$ . Now, define a directed graph  $G$  with the vertex set  $\mathbb{Z}_n$  and edge set  $E$ . For all  $i \in \mathbb{Z}_n$  and all  $a, b \in A$  with  $a < b$ , create a directed edge  $(i + a \pmod{n}, i + b \pmod{n})$ . Since  $A$  has all differences distinct, it must be the case that there are no multi-edges. It follows that  $|E| = n \binom{r}{2}$  since  $A$  is of size  $r$ .

First note that the underlying undirected graph  $G'$  consists of the union of  $n$  copies of  $K_r$  (i.e.,  $n$  copies which overlap on at most one vertex), as we can pick  $n$  out with the vertex sets  $\{i + 1 \mid a \in A\}$ . Therefore, we can bound above the largest  $k$ -ordering of  $G'$  by  $n \binom{k}{2} \binom{r/k}{2}$  since by Turan's theorem [17] the largest cut takes  $k$  components of size differing by at most 1. This is at most:

$$n \frac{k(k-1)}{2} \frac{r^2}{k^2} = nr \frac{k-1}{2k}$$

$G$  itself is Eulerian by construction, as each vertex has  $\binom{r}{2}$  edges going out and  $\binom{r}{2}$  going in and  $G$  is strongly connected. Therefore any cut in  $G'$  is a cut of half the size in  $G$ , as discussed above. So we can upper bound any  $k$ -ordering in  $G$  by  $nr^2 \frac{k-1}{4k}$ . By choice of  $r$ , this is equal to  $\frac{k-1}{4k} n^{5/3}$ .

Alon et al. then describe how to modify  $G$  to yield a DAG: delete all edges  $(i, j)$  where  $i > j$  (recall that the vertex set is given by  $\mathbb{Z}_n$ ). As we have the maximum element of  $A$  upper bounded by  $(1 + o(1))r^2$ , the only vertices with edges going backwards are the largest  $(1 + o(1))r^2$  of them (ordering vertices with  $\mathbb{Z}_n$ ). Each vertex has  $\binom{r}{2}$  edges leaving it, and in the worst case we delete all of them. Therefore, we have deleted at most  $(1 + o(1))r^2 \binom{r}{2} \leq (1 + o(1)) \frac{r^4}{4}$  edges, and now  $G$  must be acyclic since all edges go from a smaller vertex to a larger one. Therefore, since we chose  $r = \lfloor n^{1/3} \rfloor$ :

$$|E| \geq n \binom{r}{2} - (1 + o(1)) \frac{r^4}{4} = n \binom{n^{1/3}}{2} - (1 + o(1)) \frac{(n^{1/3})^4}{4}$$

So we can determine that  $|E| = (1 - o(1)) \frac{n^{5/3}}{2}$ . And so as a fraction, we have a  $k$ -ordering of size at most:

$$\frac{\frac{k-1}{4k} n^{5/3}}{(1 - o(1)) \frac{n^{5/3}}{2}} = (1 + o(1)) \frac{k-1}{2k}$$

Which is what we wanted to show, giving the following theorem:

**Theorem 1.** *There exists a class of DAGs whose largest MAX  $k$ -ORDERING is of size  $(1 + o(1)) \frac{k-1}{2k} |E|$ .*

## 4 Improved Approximation and Integrality Gap

The LP given by Kenkre admits a  $k$ -approximation on weighted instances. The LP  $z$  is as follows:

$$\begin{array}{ll}
 \text{Minimize} & \sum_{e \in E} w(e)x_e \\
 \text{Subject to} & \sum_{e \in P} x_e \geq 1 \quad \text{For all paths } P \text{ of length } k \\
 & x_e \geq 0 \quad \forall e \in E
 \end{array}$$

Although there are potentially exponentially many constraints, a separation oracle can be implemented in polynomial time since the input graph is a DAG. Therefore this LP can be solved in polynomial time.

The trivial rounding with threshold  $\frac{1}{k}$  gives a  $k$ -approximation. However, no integrality gap was shown by Kenkre et al. We give one here:

**Theorem 2.** *This LP has an integrality gap of at least  $\frac{k+1}{2}$ .*

*Proof.* Theorem 1 shows the existence of a class of graph whose MAX  $k$ -ORDERING is a  $(1 + o(1))\frac{k-1}{2k}$  fraction of the edges. Therefore, the optimal DED( $k$ ) solution is of size  $(1 - o(1))\frac{k+1}{2k}|E|$  on this graph. Yet a fractional solution can give value  $\frac{1}{k}$  to all edges, immediately giving an integrality gap of size  $\frac{k+1}{2}$ .  $\square$

Note that the natural random assignment of labels to vertices produces a MAX- $k$ -ORDERING solution of size at most  $\frac{k-1}{2k}|E|$ , even in the weighted case. This follows from the fact that the probability an edge  $(i, j)$  has  $i < j$  is equal to  $\frac{k-1}{2k}$ , as  $(i, j)$  is kept in this case and cut otherwise. This implies every weighted DAG has a DED( $k$ ) solution of size at most  $\frac{k+1}{2k}(\sum_{e \in E} w_e)$ .

### 4.1 $\frac{2}{3}k$ -approximation

Our failure to find a better integrality gap suggested the existence of a better rounding for this LP. Here we show that the problem has at least a  $\frac{2}{3}k + \frac{2}{3}$ -approximation. First we need a theorem lower bounding the weight of the valued component.

**Theorem 3.** *In the linear relaxation of any strict  $k$ -CSP which takes the OR of its variables, the non-zero variables in the LP solution average at least  $\frac{1}{k}$ .*

*Proof.* By way of contradiction, suppose otherwise. Let  $\epsilon$  be the minimum non-zero value given by the LP. If  $\epsilon \geq \frac{1}{k}$ , we are done; otherwise, set each edge's value to the following, where  $e$  is its value in the LP, and  $c = \frac{\epsilon}{\epsilon - \frac{1}{k}}$ :

$$e = e - c(e - \frac{1}{k})$$

Notice that if  $e = \epsilon$ , we get  $\epsilon - \frac{\epsilon}{\epsilon - \frac{1}{k}}(\epsilon - \frac{1}{k}) = 0$ , i.e. we set the smallest value to 0.

The first claim is that this is actually a feasible solution. To show this, take any path consisting of  $k$  edges. Denote the edges with value from the LP in this constraint as  $e_1 \dots e_m$ .

First note that  $\sum_{i=1}^m (e_i - \frac{1}{k})$  is at least 0, since the sum of  $e_i$  is at least 1 and  $m \leq k$ . It follows that  $-c(\sum_{i=1}^m (e_i - \frac{1}{k})) \geq 0$  since  $c < 0$ . Then:

$$\begin{aligned} \sum_{i=1}^m e_i &\geq 1 && \text{Since } e_i \text{ form the values of a path} \\ \sum_{i=1}^m e_i - c(\sum_{i=1}^m e_i - \frac{1}{k}) &\geq 1 && \text{By our observation} \\ \sum_{i=1}^m e_i - c(e_i - \frac{1}{k}) &\geq 1 && \text{Combining terms} \end{aligned}$$

And therefore our solution is feasible. Note that we have no negative values, since  $e$  is the minimal edge. Any values that exceed 1 can be given value 1, as any constraint with 1-valued components are clearly satisfied.

Now it only remains to show that our solution has in fact gotten better. This is equivalent to showing that the sum of  $-c(e_i - \frac{1}{k})$  over all variables with value  $e_i$  is negative. Since  $c$  is a negative constant, we only have to show that the sum of  $e_i - \frac{1}{k}$  is negative. Yet by our assumption this is true, since our valued edges average below  $\frac{1}{k}$ . Contradiction.  $\square$

Given this theorem, we have the following:

**Theorem 4.** *There is a  $(\frac{2}{3}k + \frac{2}{3})$ -approximation algorithm for DED( $k$ ).*

*Proof.* Run the LP  $z$  from Kenkre et al. given above [6]. Denote the value of this solution as  $z^*$ . Now do a threshold rounding of  $\frac{3}{2k+2}$ , and remove all edges that are no longer part of  $k$ -paths. Every path remaining must include  $\lceil \frac{2}{3}(k+1) \rceil$  edges with value (otherwise the sum would be less than 1). By the Pigeonhole Principle, every path must include a path of



length 3 of valued edges. Therefore, a DED(3) solution on the the valued edges provides a solution.

Take the subgraph induced by the edges  $L$  with non-zero value that remain after the threshold rounding. The random assignment of labels produces a solution to the DED(3) problem on this subgraph of size at most  $\frac{2}{3}|L|$ . Since by Theorem 3,  $z^* \geq \frac{1}{k}|L|$ , we have a ratio of at most  $\frac{2}{3}k$ . Given our rounding, this algorithm produces a  $\frac{2}{3}k + \frac{2}{3}$ -approximation.  $\square$

**Theorem 5.** *There is a  $\frac{3}{4}k$ -approximation algorithm for DED( $k$ ).*

*Proof.* Instead, do a threshold rounding of  $\frac{2}{k+1}$ . Now every path remaining must take  $\lceil \frac{k}{2} \rceil + 1$  edges with value, forcing a path of valued edges of length at least 2. The subgraph induced by the edges  $L$  with non-zero value is now a DED(2) problem. There is a solution to the DED(2) problem on this graph of size at most  $\frac{3}{4}|L|$  by random assignment. By Theorem 3  $z^* \geq \frac{1}{k}|A|$ , so we have a ratio of at most  $\frac{3}{4}k$ .  $\square$

**Theorem 6.** *There is a 2-approximation for DED(3).*

*Proof.* In this case, a threshold rounding of  $\frac{1}{2}$  implies that all edges have value. Then the solution has value at least  $\frac{1}{3}|L|$  by Theorem 3. Apply the  $\frac{2}{3}|L|$  randomized solution to DED(3) and we have our 2-approximation.  $\square$

Note that Theorem 3 can be extended to the weighted case, implying that these ratios hold for the weighted case as well.

## 5 On DED(2)

No hardness result was shown for this problem in Kenkre’s paper. However, in 2008 it was shown that MAX DICUT on DAGs is APX-Hard [9]. Since DED(2) is the minimization version of this problem, it is APX-Hard as well. This implies that a  $\frac{k}{2}$  algorithm would have to have at least have a small additive constant, or apply only to  $k > 2$ .

### 5.1 Algorithms for DED(2)

For DED(2), Theorem 3 lower bounds the valued portion of edges  $L$  as  $\frac{1}{2}|L|$ . Then, if we run the randomized algorithm we cut at most  $\frac{3}{4}$  of the edges, leaving us with a  $\frac{3}{2}$  approximation.<sup>1</sup>

---

<sup>1</sup>Note that a well-known theorem attributed to Nemhauser and Trotter [11] gives us that all basic feasible solutions to the standard VERTEX COVER LP are half-integral, i.e. all variables are given values in the set

We can run MAX DICUT approximation algorithms along with our LP rounding to achieve a slightly better approximation with ratio 1.37534, even though  $\frac{3}{2}$  is the best possible LP rounding given the integrality gap from Alon et al.'s graph [1]. The MAX DICUT SDP from Lewin, Livnat, and Zwick [10] gives an approximation ratio of 0.874. No matching UGC hardness result for MAX DICUT is known, although this ratio has been shown to be optimal for a large class of CSPs [2].

However, from the results on MAX CUT from Khot et al. that gave 0.878-hardness [8] (and using the standard bi-direction transformation) we can yield a 1.0954 UGC-hardness result for DED(2) on general digraphs, as Khot et al. showed that it is UGC-Hard to differentiate between a MAX CUT instance of size  $s = \frac{\arccos \rho}{\pi}$  and one of size  $c = \frac{1}{2} - \frac{1}{2}\rho$  for all  $-1 < \rho < 0$ . Yet if we take the maximum over all such  $\rho$  of  $\frac{1-s}{1-c}$  (corresponding to the ratio DED(2) would distinguish between, as it is the minimization version of this problem) we yield 1.0954, giving us our result.

Using a gadget of Trevisan et al. [15] for MAX DICUT, we can similarly show  $\frac{24}{23}$  hardness on general digraphs. We state finding a gadget with an improved gap (above that as shown by Lampis et al.) that preserves the acyclic structure as an open problem.

**Theorem 7.** *There is a 1.37534-approximation for DED(2).*

*Proof.* Given an LP solution, take the subgraph which includes only edges of weight  $\frac{1}{2}$ . If this subgraph is small, our LP has provided a good approximation. Otherwise, we will apply a MAX DICUT algorithm.

In particular, let  $0 \leq \alpha \leq 1$  be the weight of the LP solution provided from edges with weight  $\frac{1}{2}$ . We will choose some threshold  $\alpha$  to compare the two solutions.

If  $\alpha \leq x$ , then the randomized rounding gives at most  $(1-x)O + \frac{3}{2}xO$ .

Otherwise, we have  $\alpha > x$ . Let  $|E|$  be the number of edges given value 1 by our LP. Remove these from the graph, and all edges which are no longer part of a path of length 2. Now, examine only the portion of the graph  $G'$  with edge value  $\frac{1}{2}$  and  $|E'|$  edges. Here,  $\frac{1}{2}|E'| \leq O$ . A MAX DICUT solution  $D$  gives us within  $\beta = 0.874$  of the MAX DICUT solution. So, where  $O'$  is the optimal solution for MAX DICUT,  $\beta O' \leq D$ . Yet  $O \geq |E'| - O'$ , and thus we can yield a solution  $S$  of size  $|E| + |E'| - D$  to the DED(2) instance.

---

$\{0, \frac{1}{2}, 1\}$ . This is actually a stronger statement than Theorem 3. However, our theorem suffices here. Note that the general claim that  $k$ -CSPs are “ $k$ -integral,” i.e. they fall in the set  $\{1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{k}\}$  is *not* true: in fact, the natural LP for  $k$ -UNIFORM HYPERGRAPH COVER can yield LP solutions with values equal to any rational number [4]. Whether or not the solution structure for DED( $k$ ) is simpler is open. Currently we have found no graphs where there is a value smaller than  $\frac{1}{k}$ .

$O \geq |E'| - O'$  implies  $O' \geq |E'| - O$ . Then  $D \geq \beta O' \geq \beta(|E'| - O)$ , so:

$$\begin{aligned}\beta|E'| - D &\leq \beta O \\ |E'| - D &\leq \beta O + (1 - \beta)|E'| \\ |E| + |E'| - D &\leq \beta O + (1 - \beta)|E'| + |E| \\ S &\leq \beta O + (1 - \beta)|E'| + |E|\end{aligned}$$

Yet we know that  $0.5|E'| + |E| \leq O$ , so  $(1 - \beta)|E'| + 2(1 - \beta)|E| \leq 2(1 - \beta)O$ . Plugging this in above, we yield:

$$S \leq \beta O + 2(1 - \beta)O + (1 - (2(1 - \beta)))|E| = (2 - \beta)O + (2\beta - 1)|E|$$

Plugging in  $\beta$ , we yield  $S \leq 1.126O + 0.748|E|$ . And we know that  $|E|$  is small compared to  $O$ . To find the optimal threshold, set:

$$(1 - x)O + \frac{3}{2}xO = 1.126O + (1 - x)O$$

Which is  $\frac{3}{2}x = 1.126$ , so  $x = 0.75066$ . We can see that now we achieve approximation ratio at most 1.37534 in both cases.  $\square$

## 6 Conclusion

While this work begins to tackle the  $\text{DED}(k)$  problem, it asks more questions than it answers. Most pressing is the question of what the approximation ratio of the LP from Kenkre et al. [6] is. We have shown it is at most  $(\frac{2}{3}k + O(1))$  and at least  $\frac{k+1}{2}$ . If it can be shown to have a ratio equal to this integrality gap, then the problem is (asymptotically) resolved under the UGC for  $k \geq 4$ ; however, for small values of  $k$  the problem would still be open. It is also possible that there is an integrality gap of size tending to  $\frac{2}{3}k$ , in which case it might be useful to attempt to prove that it is UGC-hard to do better than this ratio; alternatively, it might be possible to show that the LP from Kenkre et al. [6] used to yield a 2-approximation to  $\text{MAX } k\text{-ORDERING}$  is in fact a  $\frac{k}{2}$  approximation for  $k \geq 4$ , however this seems somewhat unlikely as it works on general digraphs (for which there is no  $\alpha$  approximation for the minimization version).

The question as to whether any constant approximation for  $\text{DED}(k)$  implies  $\mathcal{P} = \mathcal{NP}$  is deeply connected to a host of problems. Progress on this front would constitute a large breakthrough, perhaps even giving the truth of the UGC.

## 7 Acknowledgments

Thanks to Tom Wexler for guiding me through the research process and working to make sense out of my rants each week, and to Alexa Sharp for teaching the amazing class on approximation algorithms that got me interested in the subject. Finally, thanks to Samir Khuller, Ben Eggers, and Katherine Scola for introducing me to this problem during the fantastic CAAR REU program at the University of Maryland.

## References

- [1] Noga Alon, Béla Bollobás, András Gyárfás, Jenő Lehel, and Alex Scott. Maximum directed cuts in acyclic digraphs. *J. Graph Theory*, 55(1):1–13, May 2007.
- [2] Per Austrin. Towards sharp inapproximability for any 2-csp. *SIAM J. Comput.*, 39(6):2430–2463, April 2010.
- [3] Botjan Brear, Frantiek Kardo, Jn Katreni, and Gabriel Semaniin. Minimum  $\ell_1$ -path vertex cover. *Discrete Applied Mathematics*, 159(12):1189 – 1195, 2011.
- [4] F. R. K. Chung, Z. Füredi, M. R. Garey, and R. L. Graham. On the fractional covering number of hypergraphs. *SIAM Journal on Discrete Mathematics*, 1(1):45–49, 1988.
- [5] Samuel Safra Irit Dinur. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, 162(1):439–485, 2005.
- [6] Sreyash Kenkre, Vinayaka Pandit, Manish Purohit, and Rishi Saket. On the approximability of digraph ordering. *CoRR*, abs/1507.00662, 2015.
- [7] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the Thiry-fourth Annual ACM Symposium on Theory of Computing*, STOC '02, pages 767–775, New York, NY, USA, 2002. ACM.
- [8] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan ODonnell. Optimal inapproximability results for maxcut and other 2variable csps? *SIAM Journal on Computing*, 37(1):319–357, 2007.
- [9] Michael Lampis, Georgia Kaouri, and Valia Mitsou. On the algorithmic effectiveness of digraph decompositions and complexity measures. In Seok-Hee Hong, Hiroshi Nagamochi, and Takuro Fukunaga, editors, *Algorithms and Computation*, volume 5369 of *Lecture Notes in Computer Science*, pages 220–231. Springer Berlin Heidelberg, 2008.
- [10] Michael Lewin, Dror Livnat, and Uri Zwick. Improved rounding techniques for the max 2-sat and max di-cut problems. In William J. Cook and Andreas S. Schulz, editors,

*Integer Programming and Combinatorial Optimization*, volume 2337 of *Lecture Notes in Computer Science*, pages 67–82. Springer Berlin Heidelberg, 2002.

- [11] G. L. Nemhauser and L. E. Trotter. Vertex packings: Structural properties and algorithms. *Mathematical Programming*, 8(1):232–248.
- [12] D. Paik, S. Reddy, and S. Sahni. Deleting vertices to bound path length. *IEEE Transactions on Computers*, 43(9):1091–1096, Sep 1994.
- [13] James Singer. A theorem in finite projective geometry and some applications to number theory. *Transactions of the American Mathematical Society*, 43(3):377–385, 1938.
- [14] Ola Svensson. Hardness of vertex deletion and project scheduling. In Anupam Gupta, Klaus Jansen, Jos Rolim, and Rocco Servedio, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 7408 of *Lecture Notes in Computer Science*, pages 301–312. Springer Berlin Heidelberg, 2012.
- [15] Luca Trevisan, Gregory B. Sorkin, Madhu Sudan, and David P. Williamson. Gadgets, approximation, and linear programming. *SIAM Journal on Computing*, 29(6):2074–2097, 2000.
- [16] Jianhua Tu and Wenli Zhou. A factor 2 approximation algorithm for the vertex cover problem. *Information Processing Letters*, 111(14):683 – 686, 2011.
- [17] Paul Turán. On an extremal problem in graph theory. *Mat. Fiz. Lapok*, 48(436-452):137, 1941.
- [18] Ryan Williams. Finding paths of length  $k$  in  $O(2k)$  time. *Inf. Process. Lett.*, 109(6):315–318, February 2009.