Hardness of Vertex Deletion and Project Scheduling*

Ola Svensson

EPFL, Switzerland
ola.svensson@epfl.ch

Abstract. Assuming the Unique Games Conjecture, we show strong inapproximability results for two natural vertex deletion problems on directed graphs: for any integer $k \geq 2$ and arbitrary small $\epsilon > 0$, the Feedback Vertex Set problem and the DAG Vertex Deletion problem are inapproximable within a factor $k-\epsilon$ even on graphs where the vertices can be almost partitioned into k solutions. This gives a more structured and therefore stronger UGC-based hardness result for the Feedback Vertex Set problem that is also simpler (albeit using the "It Ain't Over Till It's Over" theorem) than the previous hardness result.

In comparison to the classical Feedback Vertex Set problem, the DAG Vertex Deletion problem has received little attention and, although we think it is a natural and interesting problem, the main motivation for our inapproximability result stems from its relationship with the classical Discrete Time-Cost Tradeoff Problem. More specifically, our results imply that the deadline version is NP-hard to approximate within any constant assuming the Unique Games Conjecture. This explains the difficulty in obtaining good approximation algorithms for that problem and further motivates previous alternative approaches such as bicriteria approximations.

1 Introduction

Many interesting problems can be formulated as that of finding a large induced subgraph satisfying a desired property of a given (directed) graph. One of the most well studied such problems is the *Feedback Vertex Set (FVS)* problem where the property is acyclicity, i.e., given a directed graph $G = (V, E)$ we wish to delete the minimum number of vertices so that the resulting graph is acyclic. Another example is the *DAG Vertex Deletion (DVD)* problem, where we are given an integer k and a directed acyclic graph and we wish to delete the minimum number of vertices so that the resulting graph has no path of length[1] k.

* This research was supported by Grant 228021-ECCSciEng of the European Research Council.

[1] For notational convenience, we shall measure the length of a path in terms of the number of vertices it contains instead of the number of edges.

A. Gupta et al. (Eds.): APPROX/RANDOM 2012, LNCS 7408, pp. 301–312, 2012.

The FVS problem and the related Feedback Arc Set problem was shown to be NP-complete already in Karp's seminal paper [9] and there is a long history of approximation algorithms for these problems. Leighton and Rao [13] first gave a $O(\log^2 |V|)$-approximation algorithm. Seymour [16] improved the approximation guarantee by showing that a certain linear program approximates the value within a factor $O(\log |V| \log \log |V|)$. Seymour's arguments were then generalized by Even et al. [5] to obtain the best known approximation algorithms achieving a factor $O(\log |V| \log \log |V|)$ even in weighted graphs.

Motivated by certain VLSI design and communication problems, Paik et al. [15] considered the DVD problem and showed it to be NP-complete on general graphs and polynomial time solvable on series-parallel graphs. One can also see that DVD for a fixed k is a special case of the Vertex Cover problem on k-uniform hypergraphs and has a fairly straightforward k-approximation algorithm.

In comparison to FVS, the DVD problem has received little attention and, although we think it is a natural problem, our main motivation for studying its approximability comes from its relationship (that we prove in Section 4) with the classical deadline version of the project scheduling problem known as the *Discrete Time-Cost Tradeoff problem*. Informally (see Section 4 for a formal definition of the Deadline problem), this is the problem where we are given a deadline and a project consisting of tasks related by precedence constraints, and the time it takes to execute each task depends, by a given cost function, on how much we pay for it. The objective is to minimize the cost of executing all the tasks in compliance with the precedence constraints so that they all finish within the given deadline. Due to its obvious practical relevance, the problem has been studied in various contexts over the last 50 years (see the paper [11] by Kelly and Walker for an early reference). Fulkerson [6] and Kelley [10] obtained polynomial time algorithms if all cost functions are linear. In contrast, the problem becomes NP-hard for arbitrary cost functions [3] and there is even no known constant factor approximation algorithm in the general case. However, better (approximation) algorithms have been obtained for special cases. For example, Grigoriev and Woeginger [7] gave polynomial time algorithms for special classes of precedence constraints and one of several algorithms by Skutella [17] is a bicriteria approximation that, for any $\mu \in (0, 1)$, approximates the Deadline problem within a factor $1/(1-\mu)$ if the deadline is allowed to be violated by a factor $1/\mu$.

In summary, there are no known constant approximation algorithms for FVS, DVD, and the Deadline problem although few strong inapproximability results are known. The best known NP-hardness of approximation results follow from the fact that they are all as hard to approximate as Vertex Cover which is NP-hard to approximate within a factor 1.3606 [4]. It is indeed easy to see that Vertex Cover is a special case of FVS and DVD, and Grigoriev and Woeginger [7] gave an approximation-preserving reduction from Vertex Cover to the Deadline problem. If we assume the Unique Games Conjecture (UGC) [12], our understanding of the approximability of FVS becomes significantly better: the hardness of approximation result for Maximum Acyclic Subgraph by Guruswami et al. [8] implies that it is NP-hard to approximate FVS within any constant factor assuming the UGC. However, the results in [8] use very sophisticated techniques that are not known to imply a similar hardness for DVD and the Deadline problem.

Even though the starting motivation of this work was to better understand the approximability of the Deadline problem (and DVD), the techniques that we develop also lead to a stronger UGC-based hardness result for FVS: similar to the recent results for Vertex Cover on k-uniform hypergraphs by Bansal and Khot [1,2], we show that, for any integer $k \geq 2$ and arbitrarily small $\epsilon > 0$, there is no $k - \epsilon$-approximation algorithm for FVS even on graphs where the vertices can be almost partitioned into k feedback vertex sets. Our reduction is also much simpler than the one in [8] (albeit using the "It Ain't Over Till It's Over" theorem) but is tailored for FVS and does not yield any inapproximability result for the Maximum Acyclic Subgraph problem. More importantly, our techniques also lead to an analogous result for the DVD problem (and thereby the Deadline problem). Formally, our results for the considered vertex deletion problems can be stated as follows.

Theorem 1. *Assuming the Unique Games Conjecture, for any integer $k \geq 2$ and arbitrary constant $\epsilon > 0$, the following problems are* NP*-hard:*

FVS: *Given a graph $G(V, E)$, distinguish between the following cases:*
- (Completeness): *there exist disjoint subsets $V_1, \dots, V_k \subset V$ satisfying $|V_i| \geq \frac{1-\epsilon}{k}|V|$ and such that a subgraph induced by all but one of these subsets is acyclic.*
- (Soundness): *every feedback vertex set has size at least $(1 - \epsilon)|V|$.*

DVD: *Given a DAG $G(V, E)$, distinguish between the following cases:*
- (Completeness): *there exist disjoint subsets $V_1, \dots, V_k \subset V$ satisfying $|V_i| \geq \frac{1-\epsilon}{k}|V|$ and such that a subgraph induced by all but one of these subsets has no path of length k.*
- (Soundness): *every induced subgraph of $\epsilon|V|$ vertices has a path of length $|V|^{1-\epsilon}$.*

Note that in the completeness cases, letting $V' = V \setminus (V_1 \cup \dots \cup V_k)$, the sets $V' \cup V_i$ for $i = 1, \dots, k$ are almost disjoint solutions of size at most $(\frac{1}{k} + \epsilon)|V|$ each. In contrast, any solution basically needs to delete all vertices in the soundness case (even to avoid paths of length $|V|^{1-\epsilon}$ for DVD).

When proving UGC-based inapproximability results, the main task is usually to design "gadgets" of the considered problems that simulate a so-called dictatorship test. Once we have such "dictatorship gadgets", the process of obtaining UGC-based hardness results often follows from (by now) fairly standard arguments. In particular, the main ideas needed for our reductions leading to Theorem 1 are already present in the design of the gadgets. We have therefore chosen to present those gadget constructions with less cumbersome notation in the conference version (Section 3) and the reductions from Unique Games can be found in the full version of the paper.

As alluded to above, our main interest in DVD stems from its relationship with the Deadline problem. More specifically, in Section 4, we give an approximation-preserving reduction from DVD to the Deadline problem that combined with Theorem 1 yields:

Theorem 2. *Conditioned on the Unique Games Conjecture, for every $C > 0$, it is* NP*-hard to find a C-approximation to the Deadline problem.*

This explains the difficulty in obtaining good approximation algorithms for the Deadline problem and also further motivates alternative approaches such as the bicriteria approach by Skutella [17] that approximates the Deadline problem within a constant if the deadline is allowed to be violated by a constant factor.

2 Preliminaries

2.1 Low Degree Influence and "It Ain't over Till It's over" Theorem

Let $[k] = \{0, 1, \ldots, k-1\}$. When analyzing our hardness reductions, we shall use known properties regarding the behavior of functions of the form $f : [k]^R \mapsto \{0, 1\}$ depending on whether they have influential co-ordinates. Similar to [14, Section 3], we define the influence of the i-th co-ordinate by

$$\mathrm{Infl}_i(f) = \mathbb{E}_x[\mathrm{Var}(f)|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_R].$$

We note that if $f : \{-1, 1\}^R \mapsto \{-1, 1\}$ then this definition coincides with the intuitive expression $\Pr_x[f(x_1, \ldots, x_i, \ldots, x_R) \neq f(x_1, \ldots, -x_i, \ldots, x_R)]$.

It is well known that if we let $f = \sum_\Phi \hat{f}(\phi) X_\phi$ be the multi-linear representation of f (where, analogous of the standard Fourier representation, the characters $(X_\phi)_{\phi \in [k]^R}$ define an orthonormal basis of the vector space of all functions $[k]^n \mapsto \mathbb{R}$) then the influence can also be expressed as

$$\mathrm{Infl}_i(f) = \sum_{\phi:\phi_i \neq 0} \hat{f}^2(\phi),$$

which motivates the following definition of the degree d-influence of the i-th co-ordinate:

$$\mathrm{Infl}_i^d(f) = \sum_{\phi:\phi_i \neq 0, |\phi| \leq d} \hat{f}^2(\phi).$$

As we shall not work directly with these definitions or with the multi-linear representation, we refer the reader to [14] for the precise definitions and cut the discussion short by mentioning the property of low degree influence that shall be crucial to us (which follows from that $\sum_\phi \hat{f}^2(\phi) = \mathbb{E}_x[f(x)^2] \leq 1$).

Observation 3. *For a boolean function $f : \{0, 1\}^R \mapsto \{0, 1\}$, the sum of all degree d-influences is at most d.*

We shall now introduce a simplified version of the "It Ain't Over Till It's Over" theorem that is sufficient for the applications in this paper. The first proof was given by Mossel et al. [14] and a more combinatorial proof of a simplified version (very similar to the one used here) was given by Bansal and Khot [1] who used it to prove tight inapproximability results for Vertex Cover and a classical single machine scheduling problem. In fact many of our ideas are inspired from [1]. For $x \in [k]^R$ and a subsequence $S_\epsilon = (i_1, \ldots, i_{\epsilon R})$ of ϵR not necessarily distinct indexes in $[R]$, let

$$C_{x,S_\epsilon} = \{z \in [k]^R : z_j = x_j \ \forall j \notin S_\epsilon\}$$

denote the sub-cube defined by fixing the co-ordinates not in S_ϵ according to x. Let also $f(C_{x,S_\epsilon}) \equiv 0$ denote the expression that f is identical to 0 on the sub-cube C_{x,S_ϵ}.

Theorem 4. *For every $\epsilon, \delta > 0$ and integer k, there exists $\eta > 0$ and integer d such that any $f : [k]^R \mapsto \{0,1\}$ that satisfies*

$$\mathbb{E}[f] \geq \delta \qquad \textit{and} \qquad \forall i \in [R], \mathrm{Infl}_i^d(f) \leq \eta,$$

has

$$\Pr_{x,S_\epsilon}[f(C_{x,S_\epsilon}) \equiv 0] \leq \delta.$$

Here and throughout the paper, the probability over x, S_ϵ is such that x and S_ϵ are taken independently and uniformly at random. When ϵ is clear from the context we often also abbreviate S_ϵ by S. Note that the theorem says that a reasonably balanced function with no low degree influential co-ordinates has very low probability to be identical to 0 over the random choice of sub-cubes. In contrast, it is easy to see that a dictatorship function (on the boolean domain) $f(x) = x_s$, for some s, has $\Pr_{x,S_\epsilon}[f(C_{x,S_\epsilon}) \equiv 0] = \Pr_{x,S_\epsilon}[f(C_{x,S_\epsilon}) \equiv 1] \geq 1/2 - \epsilon$. It is this drastic difference that we will exploit in our hardness reductions.

2.2 Unique Games Conjecture

An instance of Unique Games $\mathcal{L} = (G(V,W,E), [R], \{\pi_{v,w}\}_{(v,w)})$ consists of a regular bipartite graph $G(V,W,E)$ and a set $[R]$ of labels. For each edge $(v,w) \in E$ there is a constraint specified by a permutation $\pi_{v,w} : [R] \mapsto [R]$. The goal is to find a labeling $\rho : (V \cup W) \mapsto [R]$ so as to maximize $val(\rho) := \Pr_{e \in E}[\rho \text{ satisfies } e]$, where a labeling ρ is said to satisfy an edge $e = (v,w)$ if $\rho(v) = \pi_{v,w}(\rho(w))$. For a Unique Games instance $\mathcal{L}$, we let $OPT(\mathcal{L}) = \max_{\rho : V \cup W \mapsto [R]} val(\rho)$. The now famous Unique Games Conjecture that has been extensively used to prove strong hardness of approximation results can be stated as follows.

Conjecture 1 ([12]). For any constants $\zeta, \gamma > 0$, there is a sufficiently large integer $R = R(\zeta, \gamma)$ such that, for Unique Games instances $\mathcal{L}$ with label set $[R]$ it is NP-hard to distinguish between:

- *(Completeness): $OPT(\mathcal{L}) \geq 1 - \zeta$.*
- *(Soundness): $OPT(\mathcal{L}) \leq \gamma$.*

3 Dictatorship Gadgets for Vertex Deletion Problems

We give fairly simple gadgets of the considered vertex deletion problems that informally corresponds to a dictatorship test in the following sense: (Completeness:) any dictatorship function $f : [k]^R \mapsto [k]$ (defined by $f(x) = x_s$ for some $s \in [R]$) corresponds to a good solution whereas (Soundness:) any non-trivial solution corresponds to a function $f : [k]^R \mapsto \{0, 1\}$ with a high influence co-ordinate. By fairly standard arguments, these gadgets are then used to obtain analogous hardness results assuming the Unique Games Conjecture (see the full version of the paper for details).

Throughout this section, we fix k to be an integer, $\epsilon, \delta > 0$ to be arbitrarily small constants, and let η and d be as in Theorem 4 (depending on k, ϵ and δ).

3.1 Feedback Vertex Set

We shall here describe a graph $G = (V, E)$ that naturally corresponds to a dictatorship test in the following sense:

– *(Completeness:)* A dictatorship function partitions the vertex set into subsets $V', V_0, \ldots, V_{k-1}$ satisfying $V_j \geq \frac{1-\epsilon}{k}|V|, |V'| \leq \epsilon|V|$, and for $j \in [k]$ the graph obtained by deleting $V' \cup V_j$ is acyclic.
– *(Soundness:)* Any feedback vertex set that deletes less than $(1 - 2\delta)|V|$ vertices corresponds to a function $f : [k]^R \mapsto \{0, 1\}$ with a co-ordinate i so that $\mathrm{Infl}_i^d(f) > \eta$.

Dictatorship Gadget. To make the analysis more intuitive, it will be convenient to first present a gadget that consists of two types of vertices that we refer to as *bit-vertices* and *test-vertices* and all arcs are between bit- and test-vertices:

– There is a bit-vertex b_x of weight ∞ for every $x \in [k]^R$.
– There is a test-vertex $t_{x,S}$ of weight 1 for every $x \in [k]^R$ and every sequence $S = (i_1, \ldots, i_{\epsilon R}) \in [R]^{\epsilon R}$ of ϵR not necessarily distinct indices.
– The arc incident to a test-vertex $t_{x,S}$ are the following. There is an arc $(b_z, t_{x,S})$ if $z \in C_{x,S}$ and an arc $(t_{x,S}, b_z)$ if $z \in C_{x,S}^{\oplus}$, where

$$C_{x,S}^{\oplus} = \{z \oplus 1 : z \in C_{x,S}\}$$

(here $\oplus$ denotes addition mod k).

As the bit-vertices have weight ∞, they will never be deleted in an optimal solution. We can therefore obtain an unweighted graph G of same optimal value by omitting the bit-vertices and having an arc $(t_{x,S}, t_{x',S'})$ between two test vertices if there exists a bit-vertex b_z so that $(t_{x,S}, b_z)$ and $(b_z, t_{x',S'})$. The vertex set of G will therefore correspond to the set T of test-vertices. The analysis of G therefore follows from proving that (completeness:) any dictatorship function partitions the test-vertices as required and (soundness:) that any solution that deletes less than a fraction $1 - 2\delta$ of the test-vertices corresponds to a function with a co-ordinate of high influence.

Completeness. We show that a dictatorship function $f : [k]^R \mapsto [k]$ of index s naturally partitions the test-vertices into subsets $T', T_0, \ldots, T_{k-1}$ satisfying $T_j \geq \frac{1-\epsilon}{k}|T|, |T'| \leq \epsilon|T|$, and such that the sets $T' \cup T_j$ for $j \in [k]$ are almost disjoint feedback vertex sets of size at most $(\frac{1}{k} + \epsilon)|T|$ each.

As $f(x) = x_s$, it partitions the bit-vertices in k equal sized sets

$$B_j = \{b_x : f(x) = j\} \qquad \text{for} \qquad j \in [k].$$

We say that a test-vertex $t_{x,S}$ is good if $s \notin S$ and partition the good test-vertices into k equal sized sets

$$T_j = \{t_{x,S} : s \notin S \text{ and } f(x) = j\} \qquad \text{for} \qquad j \in [k].$$

The sets are of equal size since they are partitioned according to x and whether a test-vertex is good only depends on S. Furthermore, as at least a fraction $1-\epsilon$ of the test-vertices are good we have that $|T_j| \geq \frac{1-\epsilon}{k}|T|$ for $j \in [k]$ and therefore the remaining test-vertices in T' are at most $\epsilon|T|$ many.

It remains to show that $T_j \cup T'$ defines a feedback vertex set for any $j \in [k]$. The key observation is that T_j only have incoming edges from bit-vertices in B_j and outgoing edges to bit-vertices in $B_{j\oplus 1}$. Indeed, consider a test-vertex $t_{x,S} \in T_j$ and an arc $(b_z, t_{x,S})$. By definition we have that $z \in C_{x,S}$ and as S is good we have that $f(z) = f(x) = j$, which implies that $z \in B_j$. The exact same argument implies that $t_{x,S}$ only has outgoing edges to $B_{j\oplus 1}$.

The graph obtained by deleting all bad test-vertices and one of the sets $T_0, T_1, \ldots, T_{Q-1}$ is therefore acyclic as required.

Soundness. Let A be the last $1/2$ fraction of the bit-vertices according to a topological sort of the graph. Let f_A be the indicator function of A. Note that a test-vertex $t_{x,S}$ has incoming arcs from all bit-vertices in $C_{x,S}$ and outgoing arcs to all bit-vertices in $C^{\oplus}_{x,S}$. Therefore, if a test-vertex $t_{x,S}$ is not deleted then we must have that either f_A is identical to 0 on $C_{x,S}$ (if $t_{x,S}$ is placed before the last bit-vertex for which f_A evaluates to 0) or identical to 1 on $C^{\oplus}_{x,S}$ (if $t_{x,S}$ is placed after the last bit-vertex for which f_A evaluates to 0) depending on where $t_{x,S}$ is placed according to the topological sort.

As $\mathbb{E}[f_A] = 1/2$, we have by Theorem 4 that if $\mathrm{Infl}_i^d(f_A) \leq \eta$ for all $i \in [R]$ then

$$\Pr_{x,S}[f(C_{x,S}) \equiv 0] \leq \delta$$

and

$$\Pr_{x,S}[f(C^{\oplus}_{x,S}) \equiv 1] = \Pr_{x,S}[f(C_{x,S}) \equiv 1] = \Pr_{x,S}[(1-f)(C_{x,S}) \equiv 0] \leq \delta.$$

Therefore, if the solution does not correspond to a function with a co-ordinate of high low-degree influence it must have deleted at least a fraction $1-2\delta$ of the test-vertices.

3.2 Dag Vertex Deletion Problem

We shall describe a directed acyclic graph (DAG) $G = (V, E)$ that naturally corresponds to dictatorship test in the following sense:

– *(Completeness:)* A dictatorship function partitions the vertex set into subsets $V', V_0, \ldots, V_{k-1}$ satisfying $V_j \geq \frac{1-\epsilon}{k}|V|, |V'| \leq \epsilon|V|$, and such that for $j \in [k]$ the graph obtained by deleting $V' \cup V_j$ has no path of length k.
– *(Soundness:)* Any graph obtained by deleting less than $(1-6\delta)|V|$ vertices either has a path of length $|V|^{1-\delta}$ or corresponds to a function $f : [k]^R \mapsto \{0,1\}$ with a co-ordinate i such that $\mathrm{Infl}_i^d(f) > \eta$.

Dictatorship Gadget. As in Section 3.1, it will be convenient to first present a gadget that consists of two types of vertices that we refer to as *bit-vertices* and *test-vertices*, and all edges will be between bit- and test-vertices:

– The bit-vertices are partitioned into $L+1$ bit-layers (L is selected below). Each bit-layer $\ell = 0, \ldots, L$ contains a bit-vertex b_x^ℓ of weight ∞ for every $x \in [k]^R$.
– Similarly, the test-vertices are partitioned into L test-layers. Each test-layer $\ell = 0, \ldots, L-1$ has a test-vertex $t_{x,S}^\ell$ of weight 1 for every $x \in [k]^R$ and every sequence of indices $S = (i_1, \ldots, i_{\epsilon R}) \in [R]^{\epsilon R}$.
– The arcs are the following: there is an arc $(b_z^\ell, t_{x,S}^{\ell'})$ if $\ell \leq \ell'$ and $z \in C_{x,S}$, and there is an arc $(t_{x,S}^{\ell'}, b_z^\ell)$ if $\ell > \ell'$ and $z \in C_{x,S}^{\oplus}$.
– Finally, L is selected so as $\delta L \geq |T|^{1-\delta}$, where T is the set of test-vertices.

Note that, as there are only arcs from a bit-layer ℓ to a test-layer ℓ' if $\ell' \geq \ell$ and only arcs from a test-layer ℓ' to a bit-layer ℓ if $\ell > \ell'$, the constructed graph is acyclic. Similar to the gadget for FVS, the bit-vertices can be omitted to obtain an unweighted graph G (with the set T of test-vertices as vertices) with the same optimal value by having an arc between two test-vertices if there was a path between them through one bit-vertex. Note that a path in G of length k is a path in the gadget that consists of k test-vertices. When arguing about the gadget, we will therefore say that a *path has length k if it consists of k test-vertices.*

Similarly to Section 3.1, the analysis of G follows from proving that (completeness:) any dictatorship function partitions the test-vertices as required and (soundness:) that any solution that deletes less than a fraction $1-6\delta$ of the test-vertices either has a path of length $|T|^{1-\delta}$ or corresponds to a function with a co-ordinate of high influence.

Completeness. We show that a dictatorship function $f : [k]^R \mapsto [k]$ of index s naturally partitions the test-vertices into subsets $T', T_0, \ldots, T_{k-1}$ satisfying $T_j \geq \frac{1-\epsilon}{k}|T|, |T'| \leq \epsilon|T|$, and such that for $j \in [k]$ the graph obtained by deleting $T' \cup T_j$ has no path of length k.

This can be seen by the same arguments as in Section 3.1. Indeed if we "collapse" the different layers by identifying the different copies of bit- and test-vertices then the gadget constructed here is identical to the gagdet in that section. We can therefore (by the arguments of Section 3.1), partition the bit-vertices into k equal sized sets $B_0, B_1, \ldots, B_{k-1}$ and all but an ϵ fraction of the test-vertices into k equal sized sets $T_0, T_1, \ldots, T_{k-1}$ so that any test-vertex in T_j has only incoming arcs from bit-vertices in B_j and outgoing arcs to bit-vertices in $B_{j\oplus 1}$.

Any $j \in [k]$ therefore corresponds to a solution by removing an ϵ fraction of the test-vertices (i.e., the set T') and those test-vertices in T_j.

Soundness. Before proceeding to the analysis it will be convenient to consider a different but equivalent formulation of the problem.

First, note that in any solution to DVD, i.e., a subgraph so that each path contains less than k test-vertices, we can find a coloring χ (using for example depth-first search) that assigns a color in $\{1, 2, \ldots, k\}$ to the bit-vertices with the property that, for each remaining test-vertex, the maximum color assigned to its predecessors is strictly less than the minimum color assigned to its successors. Similarly, any such coloring χ can be turned into a solution to DVD by deleting those test-vertices, for which not all predecessors are assigned lower colors than all its successors. Furthermore, from the construction of the arcs, we can assume w.l.o.g that the coloring satisfies $\chi(b_x^\ell) \le \chi(b_x^{\ell'})$ if $\ell \le \ell'$.

From the above discussion, an equivalent formulation of DVD on the constructed instances is as follows: find a coloring χ that assigns a color in $\{1, 2, \ldots, k\}$ to each bit-vertex satisfying $\chi(b_x^\ell) \le \chi(b_x^{\ell'})$ if $\ell \le \ell'$ so as to minimize the number of unsatisfied test-vertices where a test-vertex $t_{x,S}^\ell$ is said to be satisfied if

$$\max_{z \in C_{x,S}} \chi(b_z^\ell) < \min_{z \in C_{x,S}^\oplus} \chi(b_z^{\ell+1}),$$

that is all its predecessors are assigned lower colors than its successors.

It will also be convenient to consider the following lower bound on the colors assigned to most bit-vertices in each layer: define the color $\chi(\ell)$ of a bit-layer $\ell = 0, 1, \ldots, L$ as the maximum color that satisfies $\Pr_x[\chi(b_x^\ell) \ge \chi(\ell)] \ge 1 - \delta$.

Now, with each test-layer $\ell = 0, 1, \ldots, L-1$ we associate the indicator function $f^\ell : [k]^R \mapsto \{0, 1\}$ defined as follows

$$f^\ell(x) = \begin{cases} 0 & \text{if} \quad \chi(b_x^{\ell+1}) > \chi(\ell), \\ 1 & \text{otherwise.} \end{cases}$$

The key observation for the soundness analysis is the following.

Claim. For $\ell = 0, \ldots, L-1$, assuming that $\mathrm{Infl}_i^d(f^\ell) \le \eta$ for all $i \in [R]$: if a fraction 3δ of the test-vertices of test-layer ℓ are satisfied, then $\chi(\ell) < \chi(\ell+1)$.

Proof. As at least a fraction 3δ of the test-vertices of test-layer ℓ are satisfied,

$$\Pr_{x,S}\left[\max_{z \in C_{x,S}} \chi(b_z^\ell) < \min_{z \in C_{x,S}^\oplus} \chi(b_z^{\ell+1})\right] \ge 3\delta.$$

By the definition of $\chi(\ell)$ we have $\Pr_x[\chi(b_x^\ell) \geq \chi(\ell)] \geq 1-\delta$ and therefore

$$\Pr_{x,S}\left[\chi(\ell) < \min_{z \in C^{\oplus}_{x,S}} \chi(b_z^{\ell+1})\right] = \Pr_{x,S}\left[f^\ell(C_{x,S}) \equiv 0\right] \geq 2\delta.$$

As $\mathrm{Infl}_i^d(f^\ell) \leq \eta$ for all $i \in [R]$, Theorem 4 implies that $E[f^\ell] < \delta$ and hence $\chi(\ell+1) > \chi(\ell)$.

If a coloring satisfies more than a fraction 6δ of the test-vertices then at least a 3δ fraction of the test-layers are such that at least a fraction 3δ of the test-vertices of that layer are satisfied, which in turn by the preceding claim implies that either one of them corresponds to a function with a co-ordinate of high influence or $3\delta L$ many colors are needed (or equivalently the graph contains a path consisting of at least $3\delta L - 1 \geq \delta L \geq |T|^{1-\delta}$ test-vertices).

4 Discrete Time-Cost Tradeoff Problem

In the discrete time-cost tradeoff problem we are given a set J of activities together with a partial order $(J, <)$. Any execution of the activities must comply with the partial order, that is, if $j < k$ activity k may not be started until j is completed. The duration of an activity depends on how much resources that are spent on it. This tradeoff between time and cost for each job is described by a nonnegative cost function $c_j : \mathbb{R}_+ \mapsto \mathbb{R}_+ \cup \{\infty\}$, where $c_j(x_j)$ denotes the cost to run j with duration x_j. The project duration $t(x)$ of the realization x is the makespan (length) of the schedule which starts each activity at the earliest point in time obeying the precedence constraints and durations x_j. Given a deadline $T > 0$, the *Deadline problem* is that of finding the cheapest realization x that obeys the deadline, i.e., $t(x) \leq T$.

Theorem 5. *The Deadline problem is as hard to approximate as DVD.*

Proof. We reduce (in polynomial time) the problem of approximating DVD to that of approximating the Deadline problem. Given an instance of DVD, i.e., an integer k and a DAG $G(V, A)$ with the vertices ordered $0, 1, \ldots, n-1$ according to a topological sort, consider the instance of the Deadline problem defined as follows:

- The deadline T is set to n.
- The set J of activities contains three activities l_i, m_i, r_i for each vertex $i \in V = \{0, 1 \ldots, n-1\}$ with precedence constraints $l_i < c_i < r_i$ and cost functions

$$c_{l_i}(x) = \begin{cases} 0, & \text{if } x \geq i \\ \infty, & \text{otherwise} \end{cases} \qquad c_{m_i}(x) = \begin{cases} 0, & \text{if } x \geq 9/10 \\ 1, & \text{otherwise} \end{cases}$$

$$c_{r_i}(x) = \begin{cases} 0, & \text{if } x > n-1-i \\ \infty, & \text{otherwise .} \end{cases}$$

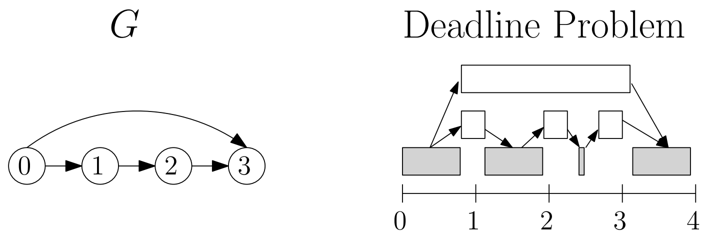

Fig. 1. For each vertex $i \in V$ the activity m_i is depicted in light gray (activities l_i and r_i are omitted). The activities corresponding to arcs are depicted in white. Finally, the depicted solution pays a cost of 1 for running activity m_2 in time 0.

In addition, there is an activity $a_{(i,j)}$ for each arc $(i,j) \in A$ with precedence constraints $m_i < a_{(i,j)} < m_j$ and cost function

$$c_{a_{(i,j)}}(x) = \begin{cases} 0, & \text{if } x \geq j - i - \frac{9}{10} + \frac{1}{10(k-1)} \\ \infty, & \text{otherwise.} \end{cases}$$

See Figure 1 for an example of the Deadline problem corresponding to a DVD instance G with $k = 3$.

Note that the cost functions of l_i, m_i, and r_i enforces that activity m_i has to be executed in the interval $[i, i+1)$ and that it will require time 9/10 unless we pay a cost of 1 which allows us to run the activity in 0 time. Furthermore, as an activity $a_{(i,j)}$ always has duration (at least) $j - i - \frac{9}{10} + \frac{1}{10(k-1)}$, the start time s_j of activity m_j must be such that $s_j - j \geq s_i - i + \frac{1}{10(k-1)}$, where s_i is the start time of activity i. Using the fact that an activity m_i must run in the interval $[i, i+1)$ in order to obey the deadline, it follows that we have to pay a cost of 1 for at least one activity corresponding to each path of length k. By similar arguments, it also follows that this is also sufficient for having a realization respecting the deadline. Therefore, any solution to the Deadline problem naturally corresponds to a solution to DVD (and vice versa) by deleting those vertices corresponding to activities with a cost of 1.

References

1. Bansal, N., Khot, S.: Optimal long code test with one free bit. In: Proceedings of the 2009 50th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2009, pp. 453–462. IEEE Computer Society, Washington, DC (2009)
2. Bansal, N., Khot, S.: Inapproximability of Hypergraph Vertex Cover and Applications to Scheduling Problems. In: Abramsky, S., Gavoille, C., Kirchner, C., Meyer auf der Heide, F., Spirakis, P.G. (eds.) ICALP 2010. LNCS, vol. 6198, pp. 250–261. Springer, Heidelberg (2010)

3. De, P., James Dunne, E., Ghosh, J.B., Wells, C.E.: The discrete time-cost tradeoff problem revisited. European Journal of Operational Research 81(2), 225–238 (1995)
4. Dinur, I., Safra, S.: On the hardness of approximating minimum vertex cover. Annals of Mathematics 162, 2005 (2004)
5. Even, G., (Seffi) Naor, J., Schieber, B., Sudan, M.: Approximating minimum feedback sets and multi-cuts in directed graphs. Algorithmica 20, 151–174 (1998)
6. Fulkerson, D.R.: A network flow computation for project cost curves. Management Science 7(2), 167–178 (1961)
7. Grigoriev, A., Woeginger, G.J.: Project scheduling with irregular costs: complexity, approximability, and algorithms. Acta Inf. 41(2-3), 83–97 (2004)
8. Guruswami, V., Håstad, J., Manokaran, R., Raghavendra, P., Charikar, M.: Beating the random ordering is hard: Every ordering CSP is approximation resistant. SIAM J. Comput. 40(3), 878–914 (2011)
9. Karp, R.: Reducibility among combinatorial problems. In: Miller, R., Thatcher, J. (eds.) Complexity of Computer Computations, pp. 85–103. Plenum Press (1972)
10. Kelley, J.E.: Critical-path planning and scheduling: Mathematical basis. Operations Research 9(3), 296–320 (1961)
11. Kelley Jr., J.E., Walker, M.R.: Critical-path planning and scheduling. Papers presented at the December 1-3, Eastern Joint IRE-AIEE-ACM Computer Conference, IRE-AIEE-ACM 1959 (Eastern), pp. 160–173. ACM, New York (1959)
12. Khot, S.: On the power of unique 2-prover 1-round games. In: Reif, J.H. (ed.) STOC, pp. 767–775. ACM (2002)
13. Leighton, T., Rao, S.: An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms. In: Proceedings of the 29th Annual Symposium on Foundations of Computer Science, SFCS 1988, pp. 422–431. IEEE Computer Society, Washington, DC (1988)
14. Mossel, E., O'Donnell, R., Oleszkiewicz, K.: Noise stability of functions with low influences: Invariance and optimality. Annals of Mathematics 171(1) (2010)
15. Paik, D., Reddy, S., Sahni, S.: Deleting vertices to bound path length. IEEE Trans. Comput. 43(9), 1091–1096 (1994)
16. Seymour, P.D.: Packing directed circuits fractionally. Combinatorica 15(2), 281–288 (1995)
17. Skutella, M.: Approximation algorithms for the discrete time-cost tradeoff problem. Mathematics of Operations Research 23(4), 909–929 (1998)