Online Disjoint Set Cover Without Prior Knowledge

Yuval Emek
Technion, Haifa, Israel
yemek@technion.ac.il

Adam Goldbraikh
Technion, Haifa, Israel
sgoadam@campus.technion.ac.il

Erez Kantor
Akamai, Cambridge, Massachusetts, USA
erez.kantor@gmail.com

Abstract

The *disjoint set cover (DSC)* problem is a fundamental combinatorial optimization problem concerned with partitioning the (hyper)edges of a hypergraph into (pairwise disjoint) clusters so that the number of clusters that cover all nodes is maximized. In its *online* version, the edges arrive one-by-one and should be assigned to clusters in an irrevocable fashion without knowing the future edges. This paper investigates the *competitiveness* of online DSC algorithms. Specifically, we develop the first (randomized) online DSC algorithm that guarantees a poly-logarithmic ($O(\log^2 n)$) competitive ratio without prior knowledge of the hypergraph's minimum degree. On the negative side, we prove that the competitive ratio of any randomized online DSC algorithm must be at least $\Omega(\frac{\log n}{\log \log n})$ (even if the online algorithm does know the minimum degree in advance), thus establishing the first lower bound on the competitive ratio of randomized online DSC algorithms.

2012 ACM Subject Classification Theory of computation → Online algorithms

Keywords and phrases disjoint set cover, online algorithms, competitive analysis, competitiveness with high probability

Digital Object Identifier 10.4230/LIPIcs.ESA.2019.44

Funding *Yuval Emek*: The work of Y. Emek was supported in part by an Israeli Science Foundation grant number 1016/17.

1 Introduction

1.1 Model and Problem Statement

A *hypergraph* $G = (V, E)$ consists of a set V of *nodes* and a multiset E of *hyperedges* (or simply *edges*), where each edge is a non-empty subset of V.[1] Unless stated otherwise, we denote $n = |V|$ and $m = |E|$.

The input to the *disjoint set cover (DSC)* problem is a hypergraph $G = (V, E)$ and the output is a *color* assignment $C : E \to \mathbb{Z}_{>0}$ to the edges in E. The objective is to maximize the number of colors $c \in \mathbb{Z}_{>0}$ that *cover* V, where color c is said to cover V (a.k.a. a *covering* color) if the union over all edges $e \in E$ with color $C(e) = c$ equals V. (The DSC problem

[1] The problem we address in this paper is often defined in terms of the equivalent *set system* terminology, where the nodes in V are identified with the elements of some abstract universe and the edges in E are simply referred to as sets or subsets.

27th Annual European Symposium on Algorithms (ESA 2019).
Editors: Michael A. Bender, Ola Svensson, and Grzegorz Herman; Article No. 44; pp. 44:1–44:16
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

should not be confused with the classic hypergraph *edge coloring* problem that also involves assigning colors to the edges. In particular, in the DSC context, it is not required that edges with the same color are disjoint.)

In the *online DSC* problem, the nodes in V are known in advance while the edges in E, assumed hereafter to be totally ordered with edges indexed as $E = \{e_1, \ldots, e_m\}$, arrive sequentially in an online fashion so that edge e_t, $1 \leq t \leq m$, arrives at time t. An online DSC algorithm should decide on the color $C(e_t)$ at (or immediately after) time t, without knowing the future edges $e_{t+1}, \ldots, e_m$, and this decision is irrevocable.

Let $\texttt{Alg}(G)$ be the number of covering colors obtained by (online or offline) DSC algorithm $\texttt{Alg}$ when invoked on hypergraph G. Following the common practice in the realm of online computation (cf. [4]), we measure the quality of online DSC algorithms by means of *competitive analysis*. A deterministic online DSC algorithm $\texttt{Alg}$ is *α-competitive* if for every n, there exists some $\beta = \beta(n) \geq 0$ such that for every n-node hypergraph G, it holds that

$$\texttt{Alg}(G) \geq \frac{\texttt{Opt}(G)}{\alpha} - \beta , \tag{1}$$

where $\texttt{Opt}$ is an optimal offline algorithm. A randomized online DSC algorithm $\texttt{Alg}$ is *α-competitive in expectation* if the bound in (1) holds in expectation; if this bound also holds with high probability (abbreviated *whp*), then $\texttt{Alg}$ is said to be *α-competitive whp.*[2] We emphasize that these probabilistic statements should hold with respect to the coin tosses of $\texttt{Alg}$, making no assumptions on the input edge sequence. Notice that since DSC is a maximization problem, it follows that if $\texttt{Alg}$ is α-competitive whp, then it is $O(\alpha)$-competitive in expectation. We refer to α as the online algorithm's *competitive ratio* and say that this competitive ratio is *pure* if the bound in (1) holds with $\beta = 0$ and *impure* otherwise.

By definition, the *minimum degree* $\delta = \min_{v \in V} |\{e \in E : v \in e\}|$ of hypergraph $G = (V, E)$ serves as an obvious upper bound on $\texttt{Opt}(G)$. Recalling that E may exhibit edge multiplicities, we emphasize that δ (and $\texttt{Opt}(G)$) may become arbitrarily large with respect to n as the length m of the input edge sequence increases. To a large extent, this fact is what makes the online DSC problem interesting: if δ would have been bounded as a function of n, then one could have included it in the additive term β and trivially obtain an (impure) competitive ratio of 1.

1.2 Background and Related Work

The DSC problem is a fundamental combinatorial optimization problem with many applications in both the offline and online domains. These applications include scheduling the operation of sensors in sensor networks, allocating servers to users in file systems, and assigning users to tasks in crowd-sourcing platforms; refer to [9] for more details. The offline version of the problem is known to be NP-hard and it can be approximated to within an asymptotically tight $O(\log n)$ approximation ratio [10].

The rigorous study of the online DSC problem was initiated by Pananjady et al. [9].[3] They first prove that a deterministic online DSC algorithm that does not hold a prior

[2] Throughout this paper, we say that event A holds whp if $\mathbb{P}(A) \geq 1 - n^{-z}$ for an arbitrarily large constant z.

[3] The authors of [9] also define the DSC problem in terms of a hypergraph, however, in that paper, the role of the nodes and edges is reversed so that the nodes in V arrive in an online fashion, each reporting the edges in E to which it belongs. To avoid confusion, we discuss the results of [9] using the current paper's model that follows the common convention in the literature on online and streaming hypergraph algorithms (see, e.g., [12, 7, 8, 6]), where the hypergraph objects that arrive in an online fashion are the edges in E.

knowledge of the minimum degree δ cannot admit a pure competitive ratio better than $\Omega(n)$.[4] Following that, Pananjady et al. focus their attention on online algorithms that know δ (or an approximation thereof) in advance and develop a deterministic purely $O(\log n)$-competitive online DSC algorithm. They also establish an $\Omega(\sqrt{\log n})$ lower bound on the (impure) competitive ratio of any such algorithm.

The DSC problem can be viewed as a (maximization) extension of the classic (minimization) *set cover* problem, where (using our hypergraph terminology) the goal is to construct a covering edge subset of minimum size. In its online version, the hypergraph $G = (V, E)$ is known in advance, but only a subset $V' \subseteq V$ of the nodes should be covered. Those are revealed one-by-one in an online fashion and must be covered immediately upon arrival. Using the *online primal-dual* technique (see [5]), Alon et al. [1] developed a (deterministic) online algorithm for this problem with competitive ratio $O(\log n \log m)$. They also proved that this is optimal up to an $O(\log n + \log m)$ factor.

1.3 Our Contribution

Our goal in this paper is to lift the assumption that the minimum degree δ is known in advance, aiming for online DSC algorithms that do not hold any initial knowledge of that hypergraph parameter, referred to hereafter as *δ-oblivious* online algorithms. As a warm up, we develop a simple deterministic δ-oblivious online algorithm with linear (in n) pure competitive ratio, thus matching the $\Omega(n)$ lower bound of [9]. Nevertheless, we wish to obtain a sublinear competitiveness which means that our online algorithms must be either randomized or admit an impure competitive ratio (or both). We advocate for this compromise: randomization as well as impure competitiveness are omnipresent in the online computation literature and seem like a small price to pay for lifting the often unrealistic assumption that the online algorithm knows the parameter δ in advance, recalling that this parameter would typically increase with the length of the input sequence.

The main technical contribution of the current paper is twofold: On the positive side, we develop a randomized δ-oblivious online DSC algorithm and prove that it is purely $O(\log^2 n)$-competitive in expectation and impurely $O(\log^2 n)$-competitive whp. On the negative side, we prove that no randomized online DSC algorithm can have impure competitive ratio better than $\Omega(\log n / \log \log n)$ in expectation or whp. Interestingly, this result holds even for online algorithms that know δ in advance, thus improving upon the $\Omega(\sqrt{\log n})$ lower bound of [9]. A comparison between the results of [9] and those established in the current paper is presented in Table 1.

1.4 Paper's Organization

The remainder of this paper is organized as follows. Following some preliminaries in Section 2, we present our simple deterministic δ-oblivious online DSC algorithm in Section 3, where we also prove that it is $O(n)$-competitive. Section 4 is then dedicated to our main positive result: a randomized δ-oblivious online DSC algorithm with competitive ratio $O(\log^2 n)$. The $\Omega(\log(n)/ \log \log n)$ lower bound on the competitiveness of randomized online DSC algorithms is established in Section 5. Finally, Section 6 is dedicated to some open questions.

[4] This negative result is obtained on hypergraphs whose δ parameter is proportional to n and the authors of [9] state it as an $\Omega(\delta)$ lower bound. We prefer to view it as an $\Omega(n)$ lower bound since in the current paper, all competitive ratio bounds are expressed as a function of n, and since it does not rule out the existence of a deterministic online DSC algorithm with pure competitive ratio $O(n)$ that works even for instances with $\delta \gg n$ (see Section 1.3).

Table 1 A comparison between the existing state of the art (top cell in each table entry) and the new results established in the current paper (bottom cell in each table entry). Empty cells indicate the lack of known results or lack of improvement over the existing results.

		known δ		*unknown* δ	
		up. bound	*low. bound*	*up. bound*	*low. bound*
deterministic	*pure*	$O(\log n)$	$\Omega\left(\sqrt{\log n}\right)$		$\Omega(n)$
			$\Omega\left(\frac{\log n}{\log\log n}\right)$	$O(n)$	
	impure	$O(\log n)$	$\Omega\left(\sqrt{\log n}\right)$		$\Omega\left(\sqrt{\log n}\right)$
			$\Omega\left(\frac{\log n}{\log\log n}\right)$	$O(n)$	$\Omega\left(\frac{\log n}{\log\log n}\right)$
rand. whp	*pure*	$O(\log n)$			
			$\Omega\left(\frac{\log n}{\log\log n}\right)$	$O(n)$	$\Omega\left(\frac{\log n}{\log\log n}\right)$
	impure	$O(\log n)$			
			$\Omega\left(\frac{\log n}{\log\log n}\right)$	$O\left(\log^2 n\right)$	$\Omega\left(\frac{\log n}{\log\log n}\right)$
rand. in expect.	*pure*	$O(\log n)$			
			$\Omega\left(\frac{\log n}{\log\log n}\right)$	$O\left(\log^2 n\right)$	$\Omega\left(\frac{\log n}{\log\log n}\right)$
	impure	$O(\log n)$			
			$\Omega\left(\frac{\log n}{\log\log n}\right)$	$O\left(\log^2 n\right)$	$\Omega\left(\frac{\log n}{\log\log n}\right)$

2 Preliminaries

Consider some hypergraph $G = (V, E)$. Given node $v \in V$, let $E(v) = \{e \in E \mid v \in e\}$ be the set of edges that contain v and define the *degree* of v in G to be the size of this set, denoted by $d(v) = |E(v)|$. Let $\delta = \min_{v \in V} d(v)$ denote the minimum degree in G.

For $1 \leq t \leq m$, let $E_t = \{e_1, \dots, e_t\}$ be the set of edges that arrive up to (including) time t. Let $E_t(v) = E_t \cap E(v)$ and let $d_t(v) = |E_t(v)|$ be the degree of node v in the hypergraph (V, E_t). Define $\eta_t = \min_{v \in e_t} d_t(v)$ to be the minimum degree, at time t, among the nodes included in edge e_t.

Recall that the goal in the DSC problem is to assign some color $C(e) \in \mathbb{Z}_{>0}$ to each edge $e \in E$. Color $c \in \mathbb{Z}_{>0}$ is said to *cover* node $v \in V$ if $C(e) = c$ for some edge $e \in E(v)$. Cast in this terminology, the objective of the DSC problem is to maximize the number of covering colors, that is, the colors that cover every $v \in V$ (see Section 1.1).

Given two integers $x \leq x'$, let $[x, x']$ denote the set of integers y satisfying $x \leq y \leq x'$ and let $[x] = [1, x]$. We generalize this notation to $x \in \mathbb{R}_{>1}$ by defining $[x] = [\lceil x \rceil]$. (The notation $[x, x']$ is reserved in the current paper only for integral x and x'.) A $\log(\cdot)$ operator with an unspecified base refers to $\log_2(\cdot)$.

Concentration Bounds

Binary random variables $X_1, \dots, X_k$ are said to be *non-positively correlated* if the following two properties hold for any $I \subseteq [k]$:[5]
(a) $\mathbb{P}\left(\bigwedge_{i \in I} X_i = 0\right) \leq \prod_{i \in I} \mathbb{P}(X_i = 0)$; and
(b) $\mathbb{P}\left(\bigwedge_{i \in I} X_i = 1\right) \leq \prod_{i \in I} \mathbb{P}(X_i = 1)$.
The following theorem, referred to as *Chernoff's bounds for non-positively correlated random variables*, was proved in [11] (see also [3]).

[5] In some literature, the term *negatively correlated* is used instead of non-positively correlated.

▶ **Theorem 2.1.** *Let $X_1, \dots, X_k$ be non-positively correlated binary random variables and let $0 \leq a_1, \dots, a_k \leq 1$. Let $X = \sum_{i \in [k]} a_i X_i$ and let $\mu = \mathbb{E}(X)$. Then,*

- $\mathbb{P}(X \leq (1-\delta)\mu) \leq \exp(-\delta^2 \mu / 2)$ *for any* $0 \leq \delta \leq 1$*; and*
- $\mathbb{P}(X \geq d) \leq 2^{-d}$ *for any* $d \geq 6\mu$*.*

Notice that independent random variables are, in particular, non-positively correlated. Indeed, by replacing the requirement that the random variables $X_1, \dots, X_k$ are non-positively correlated by the requirement that they are independent, one obtains (two of) the classic Chernoff bounds.

3 Warmup: a Deterministic Greedy Algorithm

We begin with a simple δ-oblivious deterministic online DSC algorithm, referred to as `Greedy`, whose competitive ratio is purely $O(n)$, thus matching the $\Omega(n)$ lower bound of [9] for such algorithms. For each color $c \in \mathbb{Z}_{>0}$, the algorithm maintains the variable $U_t(c)$ defined to be the set of all nodes covered by the edges in E_t whose color is c, that is, $U_t(c) = \bigcup_{1 \leq t' \leq t : C(e_{t'}) = c} e_{t'}$.

`Greedy` uses the $U_{t-1}(\cdot)$ variables to decide on the color assignment of edge e_t, $1 \leq t \leq m$, setting $C(e_t)$ to be the smallest color $c \in \mathbb{Z}_{>0}$ such that $e_t \not\subseteq U_{t-1}(c)$. This can be viewed as coloring e_t with the smallest color whose cover "benefits" from this assignment. The analysis of `Greedy`'s competitive ratio relies on the following two observations.

▶ **Observation 3.1.** *If $\delta \geq 1$, then* `Greedy`$(G) \geq 1$*.*

Proof. Follows immediately from the greedy nature of the algorithm that colors edge e_t with color $C(e_t) = 1$ if e_t contains a node that does not belong to any edge $e_1, \dots, e_{t-1}$. ◀

▶ **Observation 3.2.** `Greedy` *colors at most n edges with color c for every $c \in \mathbb{Z}_{>0}$.*

Proof. If edge e_t is assigned with color c, then $|U_t(c)| \geq |U_{t-1}(c)|$. The assertion follows since $U_t(c) \subseteq V$ for every $1 \leq t \leq m$. ◀

We are now ready to prove the following theorem.

▶ **Theorem 3.3.** `Greedy` *is purely $O(n)$-competitive.*

Proof. If $d(v) = 0$ for some node $v \in V$, then clearly `Greedy`$(G) =$ `Opt`$(G) = 0$, so assume hereafter that $\delta \geq 1$. We argue that `Greedy`$(G) \geq \lfloor \delta / n \rfloor$. Combined with Observation 3.1, this implies that

$$\texttt{Greedy}(G) \geq \max\{1, \delta/n - 1\} \geq \max\{1, \texttt{Opt}(G)/n - 1\} \geq \texttt{Opt}(G)/(2n),$$

thus establishing the assertion. To that end, consider some node $v \in V$ and recall that Observation 3.2 ensures that each color $c \in \mathbb{Z}_{>0}$ is assigned to at most n edges in $E(v)$. Therefore, there must exist at least $\lfloor d(v)/n \rfloor \geq \lfloor \delta/n \rfloor$ colors $c \in \mathbb{Z}_{>0}$ that cover v. Due to the greedy nature of the algorithm, we deduce that the colors $1, \dots, \lfloor \delta/n \rfloor$ cover v which establishes the assertion since this is true for every $v \in V$. ◀

4 The Main Algorithm

In this section, we present our main positive contribution: a randomized δ-oblivious online DSC algorithm, referred to as `Oblv`. We start by providing an intuitive overview for this algorithm in Section 4.1. The algorithm itself is then presented in Section 4.2. In Section 4.3, we establish a combinatorial lemma regarding the online DSC problem in general. This lemma plays a key role in Section 4.4, where we prove that `Oblv` is $O(\log^2 n)$-competitive in expectation. Finally, the proof presented in Section 4.4 is extended in Section 4.5 to show that the same (asymptotic) competitive ratio bound holds also whp.

4.1 Technical Challenges and Intuition

Pananjady et al. [10] developed a randomized *offline* DSC algorithm that on hypergraph $G = (V, E)$, simply colors each edge $e \in E$ by a color $C(e)$ picked uniformly at random (abbreviated hereafter by *uar*) from the *palette* $P = [\Theta(\delta / \log n)]$. Since the degree $d(v)$ of every node $v \in V$ is at least δ, it is easy to see that each color $c \in P$ covers v whp, hence, by the union bound, c covers all V whp. Using standard arguments, one can conclude that the expected number of covering colors is at least $\Omega(|P|) = \Omega(\delta / \log n)$, which is an $O(\log n)$-approximation as $\delta \geq \texttt{Opt}(G)$.

In [9], Pananjady et al. observed that the offline algorithm of [10] can be implemented as an online algorithm assuming that δ is known in advance. Their main technical contribution was then to derandomize this randomized algorithm by employing the method of conditional expectation (see, e.g., [2]), carefully adjusted to work in an online fashion.

In contrast, in the current paper we aim for a δ-oblivious online algorithm and hence, cannot use $P = [\Theta(\delta / \log n)]$ as the palette from which a color is picked for each edge $e_t \in E$. Instead, we estimate δ by the parameter $\eta_t = \min_{v \in e_t} d_t(v)$ that can be calculated at time t as it depends only on information that was already exposed to the algorithm. The combinatorial key to our algorithm is that (at least) a constant fraction of the edges e_t that contain node $v \in V$ satisfy $\eta_t \geq \Omega(d(v)/n)$. This means that we can identify (in hindsight) a sufficiently large subset of the edges $e_t \in E(v)$ for which $\Omega(\delta/n) \leq \eta_t \leq \delta$, or equivalently, $\log \eta_t \leq \log \delta \leq \log \eta_t + O(\log n)$.

We rely on this combinatorial insight for the design of `Oblv`: Upon arrival of edge e_t, the algorithm assigns the variable r_t to be an integer picked uar from the integers in the range $[\log \eta_t, \log \eta_t + O(\log n)]$, thus ensuring that 2^{r_t} is a constant approximation of δ with probability $\Omega(1/\log n)$. The algorithm then uses 2^{r_t} to construct the palette $P_t = [\Omega(2^{r_t} / \log^2 n)]$ from which the color $C(e_t)$ of edge e_t is picked (uar), where the role of the extra $\log n$ factor in the denominator is to account for the probability that 2^{r_t} is a good estimate for δ. The rest of the analysis follows the aforementioned line of arguments, concluding that `Oblv` is $O(\log^2 n)$-competitive in expectation.

For whp competitiveness we have to work a little bit harder though. While we identify (in hindsight) a palette P of size $\Theta(\delta / \log^2 n)$ such that each color $c \in P$ covers V whp, the number of such colors may be too large to apply the union bound over all of them, thus we cannot simply argue that all colors in P cover V (simultaneously) whp. Instead, we show that for each node $v \in V$, the random variables that indicate the events that color $c \in P$ does not cover v are non-positively correlated. By applying the Chernoff bound for non-positively correlated random variables, we conclude that at most a $(1/(2n))$-fraction of the colors in P do not cover v whp, hence the total number of colors in P that do not cover the whole of V is at most $|P|/2$ whp.

4.2 Algorithm's Description

Algorithm Oblv works as follows. Upon arrival of edge e_t, $1 \leq t \leq m$, the algorithm calculates $\eta_t = \min_{v \in e_t} d_t(v)$ and $\ell_t = \lceil \log \eta_t \rceil$ and then assigns the variable r_t to an integer picked uar from the set $[\ell_t, \ell_t + \lceil \log(n-1) \rceil + 2]$. Following that, the color $C(e_t)$ of edge e_t is picked uar from the palette $P_t = \left[\xi \cdot 2^{r_t} / \log^2 n \right]$, where $\xi > 0$ is a constant whose value is determined (implicitly) later on. A pseudocode description of Oblv is provided in Algorithm 1.

Algorithm 1 The operation of Oblv upon arrival of edge e_t, $1 \leq t \leq m$.

```
1: $\eta_t \leftarrow \min_{v \in e_t} d_t(v)$
2: $\ell_t \leftarrow \lceil \log \eta_t \rceil$
3: pick $r_t$ uar from $[\ell_t, \ell_t + \lceil \log(n-1) \rceil + 2]$
4: $P_t \leftarrow \left[ \xi \cdot 2^{r_t} / \log^2 n \right]$                ▷ $\xi > 0$ is a constant
5: color edge $e_t$ with color $C(e_t)$ picked uar from $P_t$
```

4.3 A Combinatorial Lemma

Fix some node $v \in V$. Edge $e_t \in E(v)$ is said to be *heavy* (for v) if

$$d_t(v) \leq 2(n-1)\eta_t \, ;$$

otherwise, we say that it is *light* (for v).

▶ **Lemma 4.1.** *For every time $1 \leq T \leq m$, more than $d_T(v)/2$ of the edges in $E_T(v)$ are heavy.*

Proof. Consider the edge sequence $\sigma = (e_1, \dots, e_T)$. By the definition of η_t, edge $e_t \in E_T(v)$ is light if and only if there exists some node $u \in e_t - \{v\}$ whose degree at time t satisfies $d_t(u) < d_t(v)/(2(n-1))$. On an intuitive level, this means that the challenge in constructing an edge sequence σ that contradicts the assertion, is to increase the degree of v while keeping the degrees of the other nodes small, thus enabling the generation of many light edges with few heavy edges. We employ this intuition to make the following simplifying assumptions.

The first assumption we make for the sake of simplifying the proof is that v is contained in all edges of the sequence σ, that is, $E_T(v) = E_T$. This assumption is clearly without loss of generality since the existence of an edge e_t that does not contain v (and hence is neither heavy nor light) increases the degrees of the nodes in e_t without increasing the degree of v.

Next, notice that all singleton edges of the form $e_t = \{v\}$ are heavy. The second assumption we make for the sake of simplifying the proof is that every heavy edge e_t in σ is a singleton, i.e., $e_t = \{v\}$. To see that this assumption is without lose of generality, suppose that e_t includes additional nodes $u \neq v$ and consider the edge sequence σ' obtained from σ by removing these nodes u from e_t. Comparing σ' to σ, one observes that $d_{t'}(u)$ decreases and $d_{t'}(v)$ remains unchanged for every $t' \geq t$, thus if σ contradicts the assertion, then so does σ'.

The third assumption we make for the sake of simplifying the proof is that every light edge e_t in σ is of size $|e_t| = 2$. To see that this assumption is without lose of generality, suppose that e_t is light with $|e_t| \geq 3$ and let u be a node of minimum degree $d_t(u)$ in e_t. Consider the edge sequence σ' obtained from σ by removing from e_t any node $u' \in e_t - \{v, u\}$. Comparing σ' to σ, one observes that edge e_t remains light (due to the existence of u) while $d_{t'}(u')$ decreases and $d_{t'}(v)$ remains unchanged for every $t' \geq t$, thus if σ contradicts the assertion, then so does σ'.

The fourth assumption we make for the sake of simplifying the proof is that σ is composed of a prefix of heavy edges followed by a suffix of light edges; i.e., there exists some $1 \leq \hat{t} \leq T$ such that edge e_t is heavy if $1 \leq t \leq \hat{t}$ and light if $\hat{t}+1 \leq t \leq T$. To see that this assumption is without loss of generality, suppose that there exists some $1 \leq t \leq T-1$ such that edge $e_t = \{v, u\}$ is light and edge $e_{t+1} = \{v\}$ is heavy and consider the edge sequence σ' obtained from σ by swapping between e_t and e_{t+1}. By construction, this swap does not turn e_t (now arriving at time $t+1$) into a heavy edge as $d_u(t+1) = d_u(t)$ and $d_v(t+1) > d_v(t)$, thus if σ contradicts the assertion, then so does σ'. Our assumption is now justified by repeating these swap operations.

So, based on the aforementioned four assumptions, the edge sequence $\sigma = (e_1, \dots, e_T)$ consists of a prefix $(e_1, \dots, e_{\hat{t}})$ of heavy edges of the form $e_t = \{v\}$ and a suffix $(e_{\hat{t}+1}, \dots, e_T)$ of light edges of the form $e_t = \{v, u\}$ for some node $u \neq v$ referred to hereafter as the *extra* node of edge e_t. The fifth and last assumption we make for the sake of simplifying the proof is that the degrees of the extra nodes are monotonically non-decreasing; that is, if u is the extra node of edge e_t, $\hat{t}+1 \leq t \leq T-1$, and u' is the extra node of edge e_{t+1}, then $d_t(u) \leq d_{t+1}(u')$. To see that this assumption is without loss of generality, suppose that $d_t(u) > d_{t+1}(u')$ and consider the edge sequence σ' obtained from σ by swapping between e_t and e_{t+1}. By construction, since e_t and e_{t+1} are light in σ, they are also light in σ', thus if σ contradicts the assertion, then so does σ'. Our assumption is now justified by repeating these swap operations.

Observe that the last simplifying assumption implies that if u is the extra node of edge e_t, $\hat{t}+1 \leq t \leq T$, and there exists some node $u' \notin \{v, u\}$ with $d_{t-1}(u') < d_{t-1}(u)$, then u' does not appear as the extra node of any edge $e_{t'}$, $t \leq t' \leq T$. This observation allows us to conclude that if u is the extra node of edge e_t, $\hat{t}+1 \leq t \leq T$, then $d_t(u) \geq (t - \hat{t})/(n-1)$.

We are now ready to establish the assertion by proving that $\hat{t} > T/2$. To that end, recall that by the definition of a light edge, if u is the extra node of edge e_t, $\hat{t}+1 \leq t \leq T$, then

$$d_t(u) < \frac{d_t(v)}{2(n-1)} = \frac{t}{2(n-1)} .$$

Put together with the bound $d_t(u) \geq (t - \hat{t})/(n-1)$, we conclude that $t/2 > t - \hat{t}$ which holds if and only if $\hat{t} > t/2$, thus completing the proof by taking $t = T$. ◀

▶ **Corollary 4.2.** *For every time $1 \leq T \leq m$, if $d_T(v) \geq z$, then*

$$\left| \left\{ e_t \in E_T(v) \mid \eta_t > \frac{z}{8(n-1)} \right\} \right| > z/4 .$$

Proof. Let $e_{t(1)}, \dots, e_{t(z)}$ be the first z edges in the sequence $(e_1, \dots, e_T)$ that contain node v, ordered so that $t(1) < \cdots < t(z)$ (we know that these z edges exist as $d_T(v) \geq z$). Lemma 4.1 ensures that more than $z/2$ of the edges $e_{t(j)}$, $1 \leq j \leq z$, are heavy (for v), hence even if all edges in $\{e_{t(j)} \mid 1 \leq j \leq z/4\}$ are heavy, we still have more than $z/4$ heavy edges in $H = \{e_{t(j)} \mid z/4 < j \leq z\}$. Since $d_{t(j)}(v) = j > z/4$ for every edge $e_{t(j)} \in H$, it follows that more than $z/4$ of the edges $e_{t(j)} \in H$ satisfy $\eta_{t(j)} > z/(8(n-1))$, thus establishing the assertion. ◀

4.4 Competitiveness in Expectation

We now turn to bound the competitive ratio of `Oblv` in expectation, based on Corollary 4.2. Let $w = \lfloor \log \delta \rfloor$ and let

$$P = \left\lceil \xi \cdot 2^w / \log^2 n \right\rceil$$

be the palette from which Oblv picks a color $C(e_t)$ (uar) when the random variable r_t is assigned to $r_t = w$. Given node $v \in V$, let

$$T(v) \;=\; \min\{1 \leq t \leq m \mid d_t(v) = 2^w\}$$

be the first time t at which the degree of v reaches $2^w \leq \delta$. Let

$$F(v) \;=\; \left\{ e_t \in E_{T(v)}(v) \mid \eta_t > \frac{2^w}{8(n-1)} \right\},$$

recalling that $E_{T(v)}(v)$ is the set of edges $e_t \in E(v)$ with $1 \leq t \leq T(v)$, and let

$$F^w(v) \;=\; \{e_t \in F(v) \mid r_t = w\}$$

be the (random) set of edges e_t in $F(v)$ for which Oblv picks a color (uar) from the palette P.

▶ **Lemma 4.3.** *If $\delta \geq \Omega(\log^2 n)$, then $|F^w(v)| \geq \Omega(\delta / \log n)$ whp for every node $v \in V$.*

Proof. Consider some edge $e_t \in F(v)$ and let A_t denote the event $e_t \in F^w(v)$. By definition,

$$\frac{2^w}{8(n-1)} \;<\; \eta_t \;\leq\; d_t(v) \;\leq\; 2^w,$$

hence

$$w - (\log(n-1) + 3) \;<\; \log \eta_t \;\leq\; w.$$

Since w is an integer, it follows that

$$w - (\lceil \log(n-1) \rceil + 2) \;\leq\; \ell_t \;\leq\; w,$$

where recall that $\ell_t = \lceil \log \eta_t \rceil$. Therefore, $w \in [\ell_t, \ell_t + \lceil \log(n-1) \rceil + 2]$ and by the design of the algorithm, we conclude that r_t is assigned to w with probability $1/(\lceil \log(n-1) \rceil + 3)$ implying that $\mathbb{P}(A_t) \geq \Omega(1/\log n)$.

Corollary 4.2 guarantees that $|F(v)| > 2^w/4 \geq \Omega(\delta)$, hence

$$\mathbb{E}\left(|F^w(v)|\right) \;=\; \sum_{e_t \in F(v)} \mathbb{P}(A_t) \;\geq\; \Omega(\delta / \log n).$$

If $\delta \geq \Omega(\log^2 n)$, then $\mathbb{E}(|F^w(v)|) \geq \Omega(\log n)$, therefore, as the events A_t are independent, we can apply Theorem 2.1 to conclude that $|F^w(v)| \geq \Omega(\delta / \log n)$ whp. ◀

▶ **Corollary 4.4.** *Fix some color $c \in P$. If $\delta \geq \Omega(\log^2 n)$, then c covers v whp for every node $v \in V$.*

Proof. Lemma 4.3 ensures that $|F^w(v)| \geq \Omega(\delta / \log n)$ whp; condition hereafter on this event. The algorithm is designed so that each edge $e_t \in F^w(v)$ is colored $C(e_t) \leftarrow c$ with probability $1/|P| = \Omega(\log^2(n)/\delta)$. Therefore, the probability that none of the edges in $F^w(v)$ is colored c is at most

$$\left(1 - \Omega\left(\log^2(n)/\delta\right)\right)^{\Omega(\delta/\log n)} \;\leq\; \exp(-\Omega(\log n)),$$

thus establishing the assertion. ◀

We are now ready to establish the desired competitive ratio bound.

▶ **Theorem 4.5.** `Oblv` *is (impurely)* $O(\log^2 n)$*-competitive in expectation.*

Proof. We prove the assertion by showing that

$$\mathbb{E}\left(\texttt{Oblv}(G)\right) \;\geq\; \Omega\left(\texttt{Opt}(G)/\log^2 n\right) - 1\,.$$

This bound holds trivially if $\texttt{Opt}(G) \leq \delta \leq O(\log^2 n)$, so assume that $\delta \geq \Omega(\log^2 n)$. Applying the union bound to the promise of Corollary 4.4, we conclude that color $c \in P$ covers all nodes in V whp and, in particular, with probability at least (say) 1/2. Therefore, by the linearity of expectation,

$$\mathbb{E}(\texttt{Oblv}(G)) \;\geq\; |P|/2 \;\geq\; \Omega\left(\delta/\log^2 n\right) \;\geq\; \Omega\left(\texttt{Opt}(G)/\log^2 n\right)\;, \tag{2}$$

thus completing the proof. ◀

Recall that in Section 3, we presented the deterministic online DSC algorithm `Greedy` whose competitive ratio is purely $O(n)$. By combining it with `Oblv`, we can turn the competitive ratio bound promised by Theorem 4.5 into a pure one. To that end, consider the online algorithm $\texttt{Oblv}_p$ that runs `Oblv` with probability 1/2 and `Greedy` with probability 1/2. If $\delta = 0$, then clearly $\texttt{Opt}(G) = \texttt{Oblv}_p(G) = 0$. If $\delta \geq \Omega(\log^2 n)$, then

$$\mathbb{E}(\texttt{Oblv}_p(G)) \;\geq\; \mathbb{E}(\texttt{Oblv}(G))/2 \;\geq\; \Omega(\texttt{Opt}(G)/\log^2 n)\,,$$

where the last transition holds due to (2). If $1 \leq \delta \leq O(\log^2 n)$, then

$$\mathbb{E}(\texttt{Oblv}_p(G)) \;\geq\; \texttt{Greedy}(G)/2 \;\geq\; 1/2 \geq \Omega(\texttt{Opt}(G)/\log^2 n)\,,$$

where the second transition holds due to Observation 3.1. Put together, we obtain the following corollary.

▶ **Corollary 4.6.** $\texttt{Oblv}_p$ *is purely* $O(\log^2 n)$*-competitive in expectation.*

4.5 Competitiveness with High Probability

We now turn to show that the competitive ratio bound established in Section 4.4 holds also whp (though not purely). For node $v \in V$ and color $c \in P$, define the random variable $X^v(c)$ to be an indicator for the event that color c does *not* cover v, namely, $X^v(c) = 1$ if and only if $C(e_t) \neq c$ for all edges $e_t \in E(v)$. Recall that Corollary 4.4 guarantees that if $\delta \geq \Omega(\log^2 n)$, then

$$\mathbb{E}\left(X^v(c)\right) \leq n^{-z} \tag{3}$$

for an arbitrarily large constant z.

The analysis in this section relies on proving that the random variables $X^v(\cdot)$ are non-positively correlated (Lemma 4.8), based on the following observation.

▶ **Observation 4.7.** *For every node* $v \in V$*, color subset* $Q \subset P$*, and color* $c' \in P - Q$*, we have*

(a) $\mathbb{P}\left(\bigwedge_{c \in Q} X^v(c) = 0 \mid X^v(c') = 0\right) \leq \mathbb{P}\left(\bigwedge_{c \in Q} X^v(c) = 0\right)$*; and*

(b) $\mathbb{P}\left(\bigwedge_{c \in Q} X^v(c) = 1 \mid X^v(c') = 1\right) \leq \mathbb{P}\left(\bigwedge_{c \in Q} X^v(c) = 1\right)$.

Proof. To see that property (a) holds, notice that if $X^v(c') = 0$, then at least one edge in $E(v)$ is colored c' which means that there is one less edge available for the colors in Q, hence $\mathbb{P}\left(\bigwedge_{c \in Q} X^v(c) = 0\right)$ decreases. To see that property (b) holds, notice that if $X^v(c') = 1$, then none of the edges in P is colored c' which means that there is one less color in P to compete with the colors in Q, hence $\mathbb{P}\left(\bigwedge_{c \in Q} X^v(c) = 1\right)$ decreases. ◀

▶ **Lemma 4.8.** *For every node $v \in V$, the random variables $X^v(c)$, $c \in P$, are non-positively correlated.*

Proof. Fix some $Q \subseteq P$. We prove that $\mathbb{P}\left(\bigwedge_{c \in Q} X^v(c) = 0\right) \leq \prod_{c \in Q} \mathbb{P}(X^v(c) = 0)$; the proof that $\mathbb{P}\left(\bigwedge_{c \in Q} X^v(c) = 1\right) \leq \prod_{c \in Q} \mathbb{P}(X^v(c) = 1)$ is analogous. To that end, we let $Q = \{c_1, \dots, c_k\}$ and prove by induction on k that

$$\mathbb{P}\left(\bigwedge_{i=1}^{k} X^v(c_i) = 0\right) \leq \prod_{i=1}^{k} \mathbb{P}\left(X^v(c_i) = 0\right) .$$

The case $k = 1$ holds trivially, so assume that the inequality holds for $k - 1$ and develop

$$\begin{aligned}\mathbb{P}\left(\bigwedge_{i=1}^{k} X^v(c_i) = 0\right) &= \mathbb{P}\left(\bigwedge_{i=1}^{k-1} X^v(c_i) = 0 \mid X^v(c_k) = 0\right) \cdot \mathbb{P}\left(X^v(c_k) = 0\right) \\ &\leq \mathbb{P}\left(\bigwedge_{i=1}^{k-1} X^v(c_i) = 0\right) \cdot \mathbb{P}\left(X^v(c_k) = 0\right) \\ &\leq \prod_{i=1}^{k-1} \mathbb{P}\left(X^v(c_i) = 0\right) \cdot \mathbb{P}\left(X^v(c_k) = 0\right) = \prod_{i=1}^{k} \mathbb{P}\left(X^v(c_i) = 0\right) ,\end{aligned}$$

where the second transition follows from Observation 4.7 and the third transition holds due to the inductive hypothesis. ◀

Assume hereafter that $\delta \geq z' n \log^3 n$ for a sufficiently large constant z' which means that $|P| \geq 2zn \log n$ for a constant z that can be made arbitrarily large. Consider some node $v \in V$ and let $X^v = \sum_{c \in P} X^v(c)$. Applying the linearity of expectation to (3), we deduce that $\mathbb{E}(X^v) \leq |P| \cdot n^{-z}$ for an arbitrarily large constant z. Lemma 4.8 allows us to apply Theorem 2.1, thus obtaining the bound

$$\mathbb{P}\left(X^v \geq \frac{|P|}{2n}\right) \leq 2^{-|P|/(2n)} \leq 2^{-2zn \log n/(2n)} = n^{-z} \tag{4}$$

for an arbitrarily large constant z. We are now ready to establish the desired competitive ratio bound.

▶ **Theorem 4.9.** `Oblv` *is (impurely) $O(\log^2 n)$-competitive whp.*

Proof. We prove the assertion by showing that

$$\texttt{Oblv}(G) \geq \Omega\left(\texttt{Opt}(G)/\log^2 n\right) - O(n \log n)$$

whp. This bound holds trivially if $\texttt{Opt}(G) \leq \delta < z' n \log^3 n$ (recall that z' is a constant), so assume that $\delta \geq z' n \log^2 n$ which means that the bound in (4) holds for every node $v \in V$.

Let X be the random variable that takes on the number of colors in P that do *not* cover V. Notice that by the definition of X^v, we know that $X \leq \sum_{v \in V} X^v$. Applying the union bound over all nodes to (4), we conclude that $X^v < |P|/(2n)$ for all nodes $v \in V$ simultaneously whp; condition hereafter on this event. We can now develop

$$X \leq \sum_{v \in V} X^v < |P|/2$$

which means that $\mathtt{Oblv}(G) > |P|/2$. The assertion follows since $|P| \geq \Omega(\delta / \log^2 n)$. ◀

5 Lower Bound

This section is dedicated to our main negative result: there does not exist a randomized online DSC algorithm with (impure) competitive ratio better than $\Omega(\log(n)/\log\log n)$ in expectation (and thus also whp). This lower bound is derived from the following theorem by Yao's min-max principle.

▶ **Theorem 5.1.** *For every n_0 and δ_0, there exist $n \geq n_0$, $\delta \geq \delta_0$, and a distribution $\mathcal{D}$ over n-node hypergraphs with minimum degree δ such that (1) $\mathtt{Opt}(G) = \delta$ for every hypergraph G in the support of $\mathcal{D}$; and (2) $\mathbb{E}_{G \sim \mathcal{D}}(\mathtt{Alg}(G)) \leq O\left(\delta \frac{\log\log n}{\log n}\right)$ for any deterministic online DSC algorithm $\mathtt{Alg}$.*

Theorem 5.1 is established in two stages. First, in Section 5.1, we construct the promised distribution $\mathcal{D}$ for the special case that $\delta = \Theta(\log(n)/\log\log n)$ (and $\mathtt{Alg}(G) \leq O(1)$). Then,in Section 5.2, we show how this construction is extended for arbitrarily large values of the parameter δ

5.1 The Basic Construction

Let $q = 2^{2^z}$ for an arbitrarily large integer z and let $r = q/(2\log q)$ (an integer by the choice of q). Each hypergraph in the support of $\mathcal{D}$ has 2^q nodes, $q + r$ edges, and minimum degree $\delta = r$. We present the construction of a random hypergraph $G = (V, E)$ in $\mathcal{D}$ and then show that $\mathtt{Opt}(G) = r$, whereas

$$\mathbb{E}_G\left(\mathtt{Alg}(G)\right) < 3 \tag{5}$$

for any deterministic online DSC algorithm $\mathtt{Alg}$, thus establishing Theorem 5.1 under the restriction that $\delta = \Theta(\log(n)/\log\log n)$.

The nodes in V are identified with the vectors in $\{0,1\}^q$. The edges in E arrive in the form of a deterministic prefix $e_1^p, \dots, e_q^p$ followed by a random suffix $e_1^s, \dots, e_r^s$. The prefix is defined by setting

$$e_i^p = \{v \in \{0,1\}^q \mid v(i) = 1\}$$

for every $i \in [q]$. For the suffix, we pick a partition $\mathcal{S} = \{S_1, \dots, S_r\}$ of $[q]$ into r equally sized clusters (each of size $|S_\ell| = q/r = 2\log q$) uar among the collection of all such partitions. The suffix is then defined by setting

$$e_\ell^s = \{v \in \{0,1\}^q \mid v(i) = 0 \text{ for all } i \in S_\ell\}$$

for every $\ell \in [r]$. Refer to Figure 1 for an illustration of the suffix edges.

▶ **Lemma 5.2.** *The hypergraph G satisfies $\mathtt{Opt}(G) = \delta = r$.*

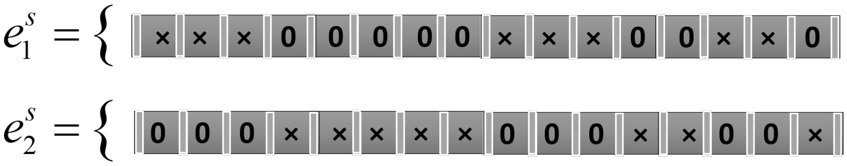


Figure 1 The construction of the suffix edges for $q = 16$ assuming that the random partition $\mathcal{S} = \{S_1, S_2\}$ consists of the clusters $S_1 = \{4, 5, 6, 7, 8, 12, 13, 16\}$ and $S_2 = \{1, 2, 3, 9, 10, 11, 14, 15\}$. The × symbol represents a 'dont-care' vector entry, i.e., it can be a 0 or a 1.

Proof. Since the zero vector $v_0 = (0, \ldots, 0)$ is included in all suffix edges e^s_ℓ, $\ell \in [r]$ and is not included in any prefix edge v^p_i, $i \in [q]$, it follows that $\delta \leq d(v_0) = r$. The proof is completed by showing that $\texttt{Opt}(G) \geq r$, recalling that $\texttt{Opt}(G) \leq \delta$.

Consider the color assignment $C^* : E \to [r]$ that colors each suffix edge e^s_ℓ, $\ell \in [r]$, by color $C^*(e^s_\ell) = \ell$ and each prefix edge e^p_i, $i \in [q]$, by color $C^*(e^p_i) = \ell(i)$ defined to be the unique $\ell \in [r]$ that satisfies $i \in S_\ell$ (this is well defined since $\mathcal{S} = \{S_1, \ldots, S_r\}$ is a partition of $[q]$). We argue that under color assignment C^*, color ℓ covers V for every $\ell \in [r]$. Indeed, if vector $v \in \{0,1\}^q$ is not included in e^s_ℓ, then $v(i) = 1$ for some $i \in S_\ell$, hence v is included in edge e^p_i. This means that $\ell(i) = \ell$, thus $C^*(e^p_i) = \ell$. The assertion follows. ◀

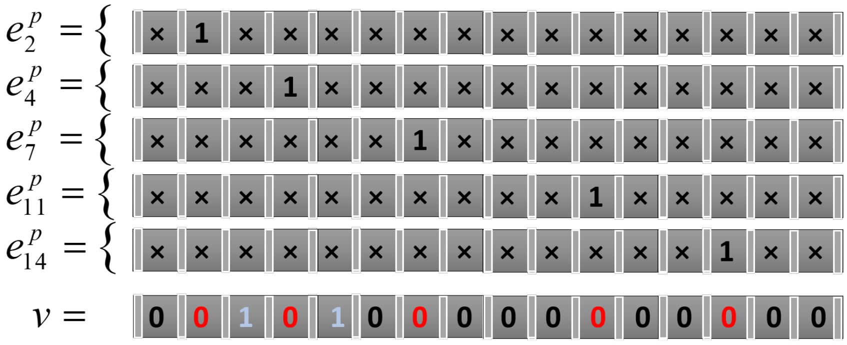
The rest of this section is dedicated to proving that (5) holds. Fix some deterministic DSC algorithm $\texttt{Alg}$. We assume that $\texttt{Alg}$ uses only (a subset of) the colors in $[q]$. To see that this assumption is without loss of generality, notice that if color $c \in \mathbb{Z}_{>0}$ is not assigned to any prefix edge e^p_i, $i \in [q]$, then it cannot cover V since the vector $(1, \ldots, 1)$ is not included in any suffix edge.

So, let $C : E \to [q]$ be the color assignment returned by $\texttt{Alg}$. Color $c \in [q]$ is said to be *heavy*, if it is assigned to at least $q/2$ prefix edges, i.e., $|\{i \in [q] \mid C(e^p_i) = c\}| \geq q/2$; otherwise, it is said to be *light*. By definition, there exists at most 2 heavy colors, so $\texttt{Alg}(G)$ is bounded from above by 2 plus the number of covering light colors. The proof that (5) holds is completed by the following lemma due to the linearity of expectation as clearly, there are at most q light colors in $[q]$.

▶ **Lemma 5.3.** *If color $c \in [q]$ is light, then c covers V with probability smaller than $1/q$.*

Proof. Consider some light color c and let $I = \{i \in [q] \mid C(e^p_i) = c\}$. Color c is said to be *ℓ-free*, $\ell \in [r]$, if $S_\ell \nsubseteq I$, that is, if there exists some index $j \in S_\ell$ such that the prefix edge e^p_j is not colored c. It is said to be *free* if it is ℓ-free for all $\ell \in [r]$.

We argue that if c is free, then it does not cover V even if all suffix edges are colored c. To that end, let b_ℓ, $\ell \in [r]$, be some index in $S_\ell - I$ (this is well defined since c is ℓ-free) and let $B = \{b_\ell \mid \ell \in [r]\}$. Consider the vector v defined by setting $v(i) = 1$ if $i \in B$; and $v(i) = 0$ otherwise. The vector v is not included in any prefix edge e^p_i, $i \in I$, because $B \cap I = \emptyset$, hence $v(i) = 0$ for all $i \in I$. It is also not included in any suffix edge e^s_ℓ, $\ell \in [q]$, because $v(b_\ell) = 1$. Therefore, if c is free, then there exists at least one vector in $\{0,1\}^q$ that it does not cover. Refer to Figure 2 for an illustration. To complete the proof, we show that c is free with probability greater than $1 - 1/q$. Fix some $\ell \in [q]$ and recall that the cluster S_ℓ is a random subset of $[q]$ of size $q/r = 2 \log q$. For the sake of this proof, we think of S_ℓ as being formed by randomly choosing $2 \log q$ indices from $[q]$ without repetitions; denote these indices by $i_1, \ldots, i_{2 \log q}$. By definition, color c is not ℓ-free if and only if $i_j \in I$ for all $1 \leq j \leq 2 \log q$.



$e_2^p = \{$ × 1 × × × × × × × × × × × × × ×

$e_4^p = \{$ × × × 1 × × × × × × × × × × × ×

$e_7^p = \{$ × × × × × × 1 × × × × × × × × ×

$e_{11}^p = \{$ × × × × × × × × × × 1 × × × × ×

$e_{14}^p = \{$ × × × × × × × × × × × × × 1 × ×

$v =$ 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0

Figure 2 Building on the example depicted in Figure 1, the color c with $I = \{2, 4, 7, 11, 14\}$ is free: it is 1-free because $5 \in S_1 - I$; it is 2-free because $3 \in S_2 - I$. Therefore, the vector v is not covered by any edge in $\{e_2^p, e_4^p, e_7^p, e_{11}^p, e_{14}^p\} \cup \{e_1^s, e_2^s\}$.

We can now develop

$$\mathbb{P}\left(\bigwedge_{j=1}^{2\log q} i_j \in I\right) = \prod_{j=1}^{2\log q} \mathbb{P}\left(i_j \in I \mid \bigwedge_{j'=1}^{j-1} i_{j'} \in I\right) = \prod_{j=0}^{2\log q-1} \frac{|I| - j}{q - j} \leq (|I|/q)^{2\log q}\,.$$

As c is assumed to be light, we have $|I| < q/2$, hence the probability that c is not ℓ-free is smaller than $(1/2)^{2\log q} = 1/q^2$. By the union bound over all $\ell \in [r]$, we conclude that the probability that c is not ℓ-free for any (at least one) $\ell \in [r]$ is smaller than $r/q^2 < 1/q$, thus establishing the assertion. ◀

5.2 The Multiplied Construction

In this section, we extend the distribution $\mathcal{D}$ presented in Section 5.1 to a distribution $\mathcal{D}_k$, where k is an arbitrarily large (positive) integer. Each hypergraph in the support of $\mathcal{D}_k$ has 2^q nodes, $k(q + r)$ edges, and minimum degree $\delta_k = kr$. We present the construction of a random hypergraph $G_k = (V_k, E_k)$ in $\mathcal{D}_k$ and then show that $\texttt{Opt}(G_k) = kr$, whereas

$$\mathbb{E}_{G_k}\left(\texttt{Alg}(G_k)\right) < 3k \tag{6}$$

for any deterministic online DSC algorithm $\texttt{Alg}$, thus completing the proof of Theorem 5.1.

Like the construction of $G = (V, E)$ presented in Section 5.1, the nodes in V_k are also identified with the vectors in $\{0, 1\}^q$. The basic idea behind the construction of the edge set E_k is to multiply the edges in E, creating k copies for each one of them. A naive attempt to do so would be to simply introduce k independent instantiations of E one after the other with the hope that the arguments used in Section 5.1 can be applied to each instantiation separately. The problem with this approach is that the prefix edges of the $(j + 1)$-st instantiation arrive after the suffix edges of the j-th instantiation, allowing the online algorithm to "color them together" optimally.

To overcome this obstacle, we design the edge sequence so that (all copies of) the prefix edges arrive before (all copies of) the suffix edges. Specifically, the edges in E_k arrive in the form of a deterministic prefix $e_{1,1}^p, \dots, e_{1,k}^p, e_{2,1}^p, \dots, e_{2,k}^p, \dots, e_{q,1}^p, \dots, e_{q,k}^p$ followed by a random suffix $e_{1,1}^s, \dots, e_{1,k}^s, e_{2,1}^s, \dots, e_{2,k}^s, \dots, e_{r,1}^s, \dots, e_{r,k}^s$, where $e_{i,1}^p, \dots, e_{i,k}^p$ and $e_{\ell,1}^s, \dots, e_{\ell,k}^s$ are k identical copies of the edges e_i^p, $i \in [q]$, and e_ℓ^s, $\ell \in [r]$, respectively, as defined in Section 5.1. We emphasize that the same (random) partition $\mathcal{S} = \{S_1, \dots, S_r\}$ is used to determine all copies of the suffix edges and that this partition is revealed to the online algorithm only after (all copies of) all prefix edges have been colored.

Since G_k is obtained from G by edge multiplicity, it follows that $\delta_k = k\delta = kr$, and by Lemma 5.2, we conclude that $\texttt{Opt}(G_k) = \delta_k = kr$, so it remains to show that (6) holds. To that end, we employ the same proof scheme as in Section 5.1: Fix some deterministic online DSC algorithm `Alg`. Since a color that is not assigned to any prefix edge does not cover the vector $(1, \dots, 1)$, we assume without loss of generality that `Alg` uses only (a subset of) the colors in $[kq]$. As in Section 5.1, we classify the colors in $[kq]$ according to the number of prefix edges they are assigned to, with heavy colors assigned to at least $q/2$ prefix edges and light colors assigned to less than $q/2$ prefix edges. Lemma 5.3 ensures that each light color covers $V = V_k$ with probability smaller than $1/q$, hence, since there are at most kq light colors, we conclude by the linearity of expectation that the expected gain of `Alg` from all light colors is smaller than k. The proof that (6) holds is completed by noticing that there are at most $kq/(q/2) = 2k$ heavy colors.

6 Discussion

Our investigation of the online DSC problem leaves several interesting open questions. The first one concerns the gap between our $O(\log^2 n)$ upper bound and $\Omega(\log(n)/\log\log n)$ lower bound on the competitive ratio of randomized δ-oblivious online DSC algorithms. Since our lower bound holds for online DSC algorithms that know δ in advance as well, one also wonders about the gap it leaves from the $O(\log n)$ upper bound of Pananjady et al. [9] for such algorithms.

The role of randomization in δ-oblivious online DSC algorithms is also not fully understood yet. While the lower bound of [9] states that a deterministic δ-oblivious online DSC algorithm cannot have a pure competitive ratio better than $\Omega(n)$, we still do not know if this is true also for the impure competitiveness of such online algorithms. In particular, it is not clear if the method of conditional expectation applied by Pananjady et al. [9] to derandomize their online algorithm can be applied also to our randomized online algorithm, especially since the derandomization technique of Pananjady et al. relies heavily on the knowledge of δ.

Finally, recall our assumption that the nodes of the hypergraph, and in particular their number n, are known in advance. While the simple deterministic online algorithm presented in Section 3 can be implemented to operate without this assumption, coming up with such an implementation of the randomized online algorithm of Section 4 seems to be a challenging task. More generally, it would be interesting to design online DSC algorithms that are (initially) oblivious to all "global" parameters of the input hypergraph, including both n and δ.

References

1 Noga Alon, Baruch Awerbuch, Yossi Azar, Niv Buchbinder, and Joseph Naor. The Online Set Cover Problem. *SIAM J. Comput.*, 39(2):361–370, 2009.

2 Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley Publishing, 4th edition, 2016.

3 Anne Auger and Benjamin Doerr. *Theory of randomized search heuristics: Foundations and recent developments*, volume 1, chapter 1: Analyzing Randomized Search Heuristics: Tools from Probability Theory, pages 1–20. World Scientific, 2011.

4 Allan Borodin and Ran El-Yaniv. *Online computation and competitive analysis*. Cambridge University Press, 1998.

5 Niv Buchbinder and Joseph Naor. The Design of Competitive Online Algorithms via a Primal-Dual Approach. *Foundations and Trends in Theoretical Computer Science*, 3(2-3):93–263, 2009.

6 Yuval Emek and Adi Rosén. Semi-Streaming Set Cover. *ACM Trans. Algorithms*, 13(1):6:1–6:22, 2016.

7 Sudipto Guha, Andrew McGregor, and David Tench. Vertex and Hyperedge Connectivity in Dynamic Graph Streams. In *Proceedings of the 34th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, PODS '15, pages 241–247, 2015.

8 Dmitry Kogan and Robert Krauthgamer. Sketching Cuts in Graphs and Hypergraphs. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, ITCS '15, pages 367–376, 2015.

9 Ashwin Pananjady, Vivek Kumar Bagaria, and Rahul Vaze. The online disjoint set cover problem and its applications. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, pages 1221–1229. IEEE, 2015.

10 Ashwin Pananjady, Vivek Kumar Bagaria, and Rahul Vaze. Optimally Approximating the Coverage Lifetime of Wireless Sensor Networks. *IEEE/ACM Transactions on Networking*, 25(1):98–111, 2017.

11 Alessandro Panconesi and Aravind Srinivasan. Randomized Distributed Edge Coloring via an Extension of the Chernoff–Hoeffding Bounds. *SIAM J. Comput.*, 26(2):350–368, 1997.

12 Barna Saha and Lise Getoor. On Maximum Coverage in the Streaming Model & Application to Multi-topic Blog-Watch. In *Proceedings of the SIAM International Conference on Data Mining SDM*, pages 697–708, 2009.