Combinatorial Lower Bounds for 3-Query LDCs

Arnab Bhattacharyya
National University of Singapore, Singapore
arnabb@nus.edu.sg

L. Sunil Chandran
Indian Institute of Science, Bangalore, India
sunil@iisc.ac.in

Suprovat Ghoshal
Indian Institute of Science, Bangalore, India
suprovat@iisc.ac.in

Abstract
A code is called a $q$-query locally decodable code (LDC) if there is a randomized decoding algorithm that, given an index $i$ and a received word $w$ close to an encoding of a message $x$, outputs $x_i$ by querying only at most $q$ coordinates of $w$. Understanding the tradeoffs between the dimension, length and query complexity of LDCs is a fascinating and unresolved research challenge. In particular, for 3-query binary LDC’s of dimension $k$ and length $n$, the best known bounds are: $2^{k^{o(1)}} \geq n \geq \tilde{\Omega}(k^2)$.

In this work, we take a second look at binary 3-query LDCs. We investigate a class of 3-uniform hypergraphs that are equivalent to strong binary 3-query LDCs. We prove an upper bound on the number of edges in these hypergraphs, reproducing the known lower bound of $\tilde{\Omega}(k^2)$ for the length of strong 3-query LDCs. In contrast to previous work, our techniques are purely combinatorial and do not rely on a direct reduction to 2-query LDCs, opening up a potentially different approach to analyzing 3-query LDCs.

2012 ACM Subject Classification Theory of computation → Error-correcting codes

Keywords and phrases Coding theory, Graph theory, Hypergraphs

Digital Object Identifier 10.4230/LIPIcs.ITCS.2020.85

Funding Arnab Bhattacharyya: Partially supported by NUS Startup Grant R-252-000-A33-133.

Acknowledgements AB thanks Sivakanth Gopi, Nikhil Srivastava, and Luca Trevisan for many useful discussions about this problem.

1 Introduction

A code $C$ is said to be a $q$-query locally decodable code (LDC) if it is possible to recover any symbol $x_i$ of a message $x$ by querying $C(x)$ on at most $q$ locations, such that even if a constant fraction of $C(x)$ is corrupted, the decoder returns $x_i$ with high probability. LDCs already appeared in the PCP literature (e.g., implicitly in [1]) but they were first explicitly formulated by Katz and Trevisan in [10]. LDCs have attracted attention not only because of their immediate relevance to data transmission and data storage but also because of their surprising connections to complexity theory and cryptography ([3, 13, 8, 11]). In more recent years, the analysis of LDCs has led to a greater understanding of basic problems in incidence geometry, the construction of design matrices and the theory of matrix scaling, e.g. [2, 7, 6].

Although LDCs have been studied now for two decades, some basic questions remain stubbornly open. In particular, we have the following open question for 3-query LDCs:

Open Question 1. What is the length of the shortest 3-query LDC that can encode all $k$-bit binary messages?
A wild variety of techniques have been used to study the problem. For a while, it was believed that the length $n$ should be exponential in $k$ for 3-query LDCs (indeed, for any constant number of queries). This belief was shattered by a breakthrough work of Yekhanin that designed 3-query LDCs of length subexponential in $k$ (conditional on some number-theoretic conjectures). Subsequent work ([9, 5] reformulated the construction in terms of matching vector codes and established an unconditional upper bound of $\exp(O(\sqrt{k \log k \log \log k})) = \exp(k^{o(1)})$ on the length.

As for lower bounds on the length of 3-query LDCs, which is the focus of this work, Katz and Trevisan [10] first gave a super linear lower bound of $\Omega(k^{3/2})$, which was then improved to $\Omega(k^2/\log k^3)$ by Kerenidis and de Wolf [12] using quantum information theoretic techniques. The current state-of-the-art is due to Woodruff [14] from over a decade ago where he showed that $n \geq \Omega(k^2/\log k)$.

Given the state of affairs, it is natural to try to prove lower bounds for stronger variants of LDCs where the task should be easier. In this work, we study a restricted form of LDCs which seem to capture most of the challenges associated with general LDCs.

**Definition 2.** For a given $\delta \in (0, 1)$, a code $C : \{\pm 1\}^k \rightarrow \{\pm 1\}^n$ is a $(3, \delta)$-strong LDC if for every $i \in [k]$, there exists a set $M_i$ of $\geq \delta n$ disjoint triples in $\binom{[n]}{3}$ such that for every $x \in \{\pm 1\}^k$ and for every triple $(j_1, j_2, j_3) \in M_i$, $x_i = C(x)_{j_1} \cdot C(x)_{j_2} \cdot C(x)_{j_3}$. Moreover, if $i \neq i'$, a triple in $M_i$ intersects a triple in $M_{i'}$ in at most 1 coordinate.

Known constructions of 3-query LDCs are strong. Conceptually, the main restriction that the above definition makes is that each triple in the matching $M_i$ successfully decodes $x_i$ for every $x$. On the other hand, Katz and Trevisan [10] show that general LDCs yield matchings $M_1, \ldots, M_k$ such that each triple in the matching $M_i$ successfully decodes $x_i$ for most (not all) $x$’s.

We show a combinatorial proof of the known $\Omega(k^2/\log k)$ lower bound for the length of code words of 3-query strong LDCs. Here is the main theorem stating the lower bound.

**Theorem 3.** Let $C : \{\pm 1\}^k \rightarrow \{\pm 1\}^n$ be a $(3, \delta)$-strong LDC. Then, $n \geq \Omega_{\delta}(k^2/\log k)$.

### 1.1 Proof Overview

As we already noted, Theorem 3 follows from [14]. Of more interest is our proof technique. Woodruff’s lower bound reduces 3-query LDCs to 2-query LDCs and applies known analytic proofs giving tight bounds for 2-query LDCs [12]. On the other hand, our proof is purely combinatorial and does not seem to be a reduction to 2 queries.

Our starting point is the observation that strong 3-query LDCs are equivalent to **even-colored 3-uniform hypergraphs**. A 3-uniform hypergraph is called **linear** if any two edges intersect in at most one vertex.

**Definition 4.** An $(n, k, \delta)$-even-colored 3-uniform hypergraph is a linear edge-colored hypergraph $H$ on $n$ vertices with each edge having a color in $\{1, \ldots, k\}$ such that:

(i) For each $i \in [k]$, the edges of color $i$ form a matching of size at least $\delta n$, and

(ii) If $H'$ is a subgraph of $H$ such that every vertex has even degree in $H'$, then there are an even number of edges of each color in $H'$.

---

1 For instance, Woodruff in [15] gave an $\Omega(k^2)$ lower bound for the special case of linear 3-query LDCs.

2 The decoding scheme of taking the product (xor) of the codeword bits is without loss of generality (see [14]). The additional condition in Definition 2 about triples in different matchings intersecting only at single coordinates is made for technical convenience and should be avoidable.
Given a \((3, \delta)\)-strong LDCs, define the hypergraph \(H\) which is the union of the matchings \(M_1, \ldots, M_k\) given by Definition 2, and let the color of an edge be the matching it comes from. Then, it is easy to check that both conditions (i) and (ii) are met (see Claim 5). The correspondence naturally goes in the other direction too, although this is not needed in the present work.

We prove an upper bound \(k^2 \leq O(\delta(n \log n))\) for \((n, k, \delta)\)-even-colored 3-uniform hypergraphs, proving our main theorem. To motivate our proof, let us sketch the corresponding argument for 2-query LDCs (which is also new to the best of our knowledge). Suppose we have a (2-uniform) graph which is the union of \(k\) matchings, with edges from the \(i\)’th matching having color \(i\). Analogously to condition (ii) of Definition 4, also suppose that any cycle contains an even number of edges of each color. Then, we prove that the number of vertices \(n\) is at least \(\exp(k)\). For simplicity, suppose the matchings of each color are perfect. Our argument is through coding (ironically!). Fix an arbitrary vertex \(s\). For any vertex \(v \neq s\), let its signature \(S(v)\) be defined as \((n_1, \ldots, n_k)\) where \(n_i\) is the parity of the number of edges of color \(i\) on a path \(P\) from \(s\) to \(v\). We claim that \(S(v)\) does not depend on the path chosen. This is because if two paths from \(s\) to \(v\) gave different signatures, this would yield a cycle in which some color occurred an odd number of times. On the other hand, there are at least \(2^k\) different signatures because for any signature \((s_1, \ldots, s_k) \in \{0, 1\}^k\), there is a path from \(s\) with exactly \(s_i\) edges of color \(i\) (since the matchings are perfect). Hence, the number of vertices is at least \(\exp(k)\).

Our proof for 3-uniform hypergraphs is in a similar spirit. Instead of a path to define a signature, we use a sequence of cherries, borrowing an idea from [4]. A cherry is a pair of hyperedges which uniquely intersect at a hyperedge; see Figure 1. We observe that if the number of edges is sufficiently large, then there are many cherries. We then use this structure to show that there are even subgraphs (i.e., subgraphs in which all vertices have even degrees) which have an odd number of edges of some color. Namely, we construct a “cycle of cherries” in which we know there is a color that appears on a unique edge, yielding the contradiction. More details follow.

Formally, given an \((n, k, \delta)\)-even colored hypergraph \(H\) which is a union of matchings \(M_1, \ldots, M_k\), define the signature graph \(G = (V', E')\) as follows. The vertex set \(V' = V \times V\), and there is an edge in \(G\) between the vertices \((u_1, v_1)\) and \((u_2, v_2)\) whenever there exists a \(w \in V\) such that \(\{u_1, u_2, w\}\) and \(\{v_1, v_2, w\}\) are hyperedges forming a cherry in \(H\). Moreover, such an edge is labeled by the pair of colors \(\{i_1, i_2\}\) if \(\{u_1, u_2, w\}\) \(\in M_{i_1}\) and \(\{v_1, v_2, w\}\) \(\in M_{i_2}\). The signature graph enjoys the following useful structural property:

For any vertex \(x\) in \(G\) and for any color \(i \in [k]\), there are at most 4 edges incident to \(x\) that have \(i\) in their label. \((\ast)\)

The proof of \((\ast)\) follows from the definition of \(G\) in terms of cherries (see Claim 8).

---

**Figure 1** A cherry formed from the edges \(\{u_1, w, u_2\}\) and \(\{v_1, w, v_2\}\) intersecting at \(w\).
For the sake of contradiction, assume that \( k \geq \sqrt{Cn \log n} \) for some large constant \( C \). This, along with a standard averaging argument, implies that there exists a large subgraph \( G' \) of the signature graph with minimum degree at least \( k^2 n / 4n^2 \geq (C/4) \log n \). Now fixing an arbitrary vertex \( r \in V(G') \) as root, we iteratively grow a sequence of trees \( T_1, T_2, \ldots, T_{i+1} \) using edges in \( G' \), while maintaining the following rainbow condition: For any vertex \( x \in V(T_i) \), no color appears more than once among the colors labeling the unique path from \( r \) to \( x \) in \( T_i \).

We explain how to construct \( T_{i+1} \) from \( T_i \) so that the above rainbow condition is met. Let \( L_i \) denote the leaves of the tree \( T_i \), and let \( N_i \) denote the neighbors of \( y \) in \( T_i \) so that the colors labeling the edge \((x, y)\) do not occur on the path from the root \( r \) to \( x \) in \( T_i \). A short argument (Claim 9) allows us to deduce that \( N_i \) must be disjoint from \( V(T_i) \), because otherwise, condition (ii) of Definition 4 is violated. Hence, the next tree \( T_{i+1} \) can be built by letting \( L_{i+1} = N_i \) and adding one edge from each vertex in \( L_{i+1} \) to a vertex in \( L_i \).

We continue this process until \( |L_{\ell}| < 2|L_{\ell-1}| \) for some iteration \( \ell \). We now sketch how to arrive at a contradiction. From the stopping criteria, we know that for every \( i < \ell - 1 \) we have \( |L_{i+1}| \geq 2|L_i| \geq 2^{i+1} \), and therefore the depth of the tree is at most \( \ell = O(\log n) \). Therefore, for any \( x \in L_{\ell-1} \), the number of colors labeling the path from \( r \) to \( x \) is at most \( O(\log n) \). From property (\ast), we get that there are at least \((C/4) \log n - O(\log n) = C' \log n \) neighbors of \( x \) that are in \( L_\ell \) (for some other constant \( C' \)). Since \( |L_\ell| < 2|L_{\ell-1}| \), there exists a vertex \( w \in L_\ell \) with at least \((C'/2) \log n \) neighbors in \( L_{\ell-1} \). Again, invoking property (\ast), for \( C' \) large enough, there will be a neighbor \( w' \in L_{\ell-1} \) such that the colors labeling \((w, w')\) do not appear among the \( O(\log n) \) labels of the path from \( r \) to \( w \). From here, we can conclude that the unique path between \( w \) and \( w' \) in \( T_\ell \) along with the edge \((w, w')\) forms a cycle in \( G \) in which some color appears exactly once. This structure corresponds to a subgraph in \( H \) that violates condition (ii) of Definition 4.

In the rest of the paper, we present the argument formally with all the details. It is unclear currently how to extend the analysis to \( q \)-query LDCs. But we remain hopeful that by looking at more intricate combinatorial structures than cherries, we can make some progress.

## 2 Preliminaries

In this section and later, we do not invoke the notion of even-colored subgraphs, and we define objects directly in reference to strong 3-query LDCs.

Given a \((3, \gamma)\)-strong LDC, we define the recovery hypergraph \( H \), where \( V(H) = [n] \) and \( E(H) := \bigcup_{i \in [k]} M_i \) to be the 3-uniform hypergraph which is the union of matchings \( M_1, M_2, \ldots, M_k \). For any edge \( e \in E(H) \), we say that the color of the hyperedge \( e \) is \( i \) if \( e \) belongs to matching \( M_i \). We use the notation \( \text{col}(e) \) to denote the color of the hyperedge \( e \). We additionally assume that \( H \) is linear, i.e. no two hyperedges of \( H \) intersect in more than one element.

Let \( L \) be a hypergraph. Then we define an augmentation \( L' \) of \( L \) as follows: \( V(L') = V(L) \) and \( E(L') \) is a multiset where each member \( e \in E(L) \) also belongs to \( E(L) \) but can possibly have a higher multiplicity than the multiplicity of \( e \) in \( E(L) \). With respect to a hypergraph \( L \), where a hyperedge \( e \) is allowed to have multiplicity greater than 1, we denote by \( \text{mult}_L(e) \) the multiplicity of \( e \) in \( E(L) \). We may drop the subscript \( L \), if the hypergraph under consideration is clear from the context. Also for \( v \in V(L) \), \( \deg_L(v) := \sum_{e \in E(L): v \in e} \text{mult}_L(e) \). If for all \( v \in V(L) \), \( \deg_L(v) \) is even, then \( L \) is called an even hypergraph. We use \( \text{col}(L) \) to denote the multiset of colors associated with edges in \( E(L) \).

If \( L \) is an augmentation of the recovery hypergraph \( H \), for \( 1 \leq i \leq k \), we define

\[
\lambda_L(i) := \sum_{e \in E(L): \text{col}(e) = i} \text{mult}_L(e).
\]
Claim 5. Let $L$ be an augmentation of the recovery hypergraph $H$. If $L$ is even, then for $1 \leq i \leq k$, $\lambda_L(i)$ is even.

Proof. Suppose for contradiction that there exists $i$, $1 \leq i \leq k$ such that $\lambda_L(i)$ is odd. Recall that the indices of the code word bits correspond to the vertices of the recovery hypergraph. Let us assume that $y = (y_1, y_2, \ldots, y_n)$ is the code word of a message $(x_1, x_2, \ldots, x_k)$, where $x_i = -1$ and $x_j = 1$ for $j \neq i$. For an edge $e = \{a, b, c\}$ of the recovery hypergraph, $Y_e = y_a y_b y_c$. Now it is clear that $\prod_{e \in E(L)} Y_e = 1$, so $L$ is an even augmentation of $H$.

On the other hand, by definition of the recovery hypergraph, if $\text{col}(e) = t$, then $Y_e = x_t$ for $1 \leq t \leq k$. Therefore $\prod_{e \in E(L)} Y_e = \prod_{1 \leq t \leq k} x_t^{\lambda_L(t)}$. Clearly since for the selected message $x_j = 1$ for $j \neq i$, we infer that $\prod_{e \in E(L)} Y_e = x_i^{\lambda_L(i)} = -1$, if $\lambda_L(i)$ is odd. This is a contradiction. We conclude that for $1 \leq i \leq k$, $\lambda_L(i)$ is even.

The Signature Graph

We define a graph called the signature graph $G$ as follows: $V(G) = \{(u, v) : u, v \in V(H); u \neq v\}$ and an edge exists between two vertices $(u_1, v_1)$ and $(u_2, v_2)$ of $G$ if and only if $(u_1, v_1) \cap (u_2, v_2) = \emptyset$ and there exists a vertex $w \in V(H)$ such that $(u_1, u_2, w), (v_1, v_2, w) \in E(H)$. Note that since the recovery hypergraph $H$ is linear, if there exists an edge between two vertices $(u_1, v_1)$ and $(u_2, v_2)$, there is a unique vertex $w$ such that $(u_1, u_2, w), (v_1, v_2, w) \in E(H)$. We may say that the vertex $w$ causes the edge $((u_1, v_1), (u_2, v_2))$. Given an edge $e \in E(G)$, we define $T(e) = \{(u_1, u_2, w), (v_1, v_2, w)\}$ if $w$ causes the edge $e$. We may abuse the notation and use $T(e)$ to denote the corresponding unordered set. We define col$(e) = \{i_1, i_2\}$ if $(u_1, u_2, w) \in M_{i_1}$ and $(v_1, v_2, w) \in M_{i_2}$. Note that $i_1 \neq i_2$ since $w$ cannot be in two different edges of the same matching.

Claim 6. The number of edges in the signature graph $G$ is at least $12\gamma^2 nk^2$.

Proof. Since each matching is of size at least $\gamma n$, the number of hyperedges $m$ in $H$ is at least $\gamma nk$. It follows that $\sum_{v \in V(H)} \deg(v) = 3m \geq 3\gamma nk$. For any vertex $w \in V(H)$ consider a pair of incident edges, say $(u_1, u_2, w)$ and $(v_1, v_2, w)$. Since $H$ is linear, $(u_1, u_2) \cap (v_1, v_2) = \emptyset$. It is easy to see that based on this pair of incident edges, $w$ can cause 4 distinct edges of the signature graph $G$. Therefore the vertex $w$ causes $4(d(v))^2$ distinct edges of $G$, where $d(v) = \deg(v)$. As we have mentioned earlier, two different vertices $w$ and $w'$ cannot cause the same edge in $G$. Therefore $|E(G)| \geq 4 \sum_{v \in V(H)} (d(v))^2 = 2 \sum_{v \in V(H)} (d(v))^2 - d(v))$. Recall that $\sum_{v \in V(H)} d(v) = 3m$. Using Cauchy-Schwarz inequality, we get $\sum_{v \in V(H)} (d(v))^2 \geq \frac{9m^2}{n}$. It follows that $|E(G)| \geq \frac{6m^2}{n}(3m - n) \geq \frac{12m^2}{n} \geq 12\gamma^2 nk^2$. Here we have used $m \geq \gamma nk$ and $m \geq n$, since $\gamma k \geq 1$.

For a subgraph $J$ of the signature graph $G$, we define $H_J$ to be the augmentation of $H$, with $V(H_J) = V(H)$ and $E(H_J) = \cup_{e \in E(J)} T(e)$. Note that when we take the union here, we retain multiple copies of a hyperedge if that hyperedge appears in multiple sets $T(e)$ taking part in the union operation. Thus $E(H_J)$ is by definition a multi-set. We extend some of the notation used for hypergraphs to subgraphs of signature graphs also in the following way: We use col$(J)$ to denote the multiset col$(H_J)$. A hypergraph $H'$ is rainbow colored with respect to an edge coloring if there exist no two hyperedges having the same color. (In

\[^3\text{For vectors } X, Y, \text{ the Cauchy-Schwarz inequality states that } \|X\| \cdot \|Y\| \geq X^T Y. \text{ Now take } X = (d(v_1), d(v_2), \ldots, d(v_n))^T, \text{ where } V(H) = \{v_1, v_2, \ldots, v_n\} \text{ and } Y = (1, 1, \ldots, 1)^T \text{ to get the required lower bound.}\]
Combinatorial Lower Bounds for 3-Query LDCs

In particular, there will not be any hyperedge with multiplicity greater than 1. A subgraph $J$ of the signature graph is rainbow colored, if $H_J$ is rainbow colored. We may also say $J$ (or $H$) is rainbow, shortening the phrase rainbow colored.

\textbf{Claim 7.} Let $J$ be an even subgraph of the signature graph $G$, i.e. $\forall v \in V(J)$, $\deg(v)$ is even. Then $H_J$ is an even augmentation of $H$.

\textbf{Proof.} Recall that each edge $e = ((u_1, v_1), (u_2, v_2)) \in E(G)$ corresponds to exactly 2 edges in $E(H_J)$, namely the two edges of $T(e) = \{(u_1, u_2, w), \{v_1, v_2, w\}\}$, where $w$ is the unique vertex which caused the edge $e$. We say that $w$ appears in the role of an intermediate vertex and $u_1, v_1, u_2$ and $v_2$ appear in the role of signature vertices in $T(e)$. It is easy to see that since in $T(e)$ itself the degree $w$ is even, each vertex plays the role of an intermediate vertex an even number of times. Noting that in $T(e)$ each vertex appears in the role of a signature vertex exactly once, it is easy to see that if $x = (u, v)$ is a vertex of $J$, then $u$ (also $v$) plays the role of a signature vertex in $\cup_{e \in E_J} T(e)$ (where $E_J(x)$ denotes the set of edges incident on $x$ in $J$) exactly $\deg_J(x)$ times. Since $\deg_J(x)$ is even, it follows that $\deg_{H_J}(u)$ and $\deg_{H_J}(v)$ are even numbers.

\textbf{Claim 8.} Let $x$ be a vertex of the signature graph $G$ and let $E(x)$ be the set of edges incident on $x$ in $G$. Let $\mathcal{C} \subseteq [k]$ be a subset of colors. Let $E(x, \mathcal{C}) = \{ e \in E(x) : \text{col}(e) \cap \mathcal{C} \neq \emptyset \}$. Then $|E(x, \mathcal{C})| \leq 4|\mathcal{C}|$.

\textbf{Proof.} Let $x = (u, v)$. For an edge $e \in E(x)$, $T(e)$ contains 2 hyperedges, exactly one of which contains $u$ and the other one contains $v$. Let us denote by $T(e)_1$ and $T(e)_2$ the hyperedges in $T(e)$ that contain $u$ and $v$ respectively. For $1 \leq i \leq k$, $E(x, \mathcal{C}) = \cup_{i \in \mathcal{C}} E^i_1 \cup E^i_2$, where $E^i_j = \{ e \in E(x) : \text{col}(T(e)_j) = i \}$ for $j = 1, 2$. First we will show that $|E^i_1| \leq 2$. To see this, note that if $\text{col}(T(e)_1) = i$, then $T(e)_1 \in M_i$ and $T(e)_2$ contains $u$ as mentioned earlier. There is a unique hyperedge with these properties since $M_i$ is a matching. Let $T(e)_1 = (u, a, b)$. Then either $a$ or $b$ could have caused the edge $e$. If $a$ caused the edge $e$, then $T(e)_2$ contains both $v$ and $a$, and there is a unique edge in $E(H)$ that is a superset of $\{v, a\}$ since $H$ is linear. Similarly if $b$ caused $e$, then $T(e)_2$ is uniquely determined, since it should contain both $v$ and $b$. It follows that $|E^i_2| \leq 2$ for all $i$, $1 \leq i \leq k$. A similar argument shows that $|E^i_2| \leq 2$. It follows that $|E(x, \mathcal{C})| \leq 4|\mathcal{C}|$.

Now we are ready to prove Theorem 3.

\section{Proof of Theorem 3}

For contradiction, we shall assume that $k^2 > Cn \log n$ where $C = C(\gamma)$ is a sufficiently large constant. We can then lower bound the average degree of $G$ as follows. From claim 6, we know that $|E(G)| \geq 12\gamma^2 n k^2$. On the other hand, the number of vertices in $G$ is at most $n^2$. Therefore, for $k^2 \geq Cn \log n$, for $C$ large enough, the average degree of $G$ is at least $(2C') \log n$ so that we can find a subgraph $G' \subseteq G$ with minimum degree at least $C' \log n$, where $C'$ is a sufficiently large constant.

Now we fix a vertex $r \in V(G')$, and we grow a rainbow tree rooted at $r$ in $G'$ by level as follows. Let $T_0$ be the tree consisting only of the root $r$ and $T_1$ be the tree consisting of $r$ and all its neighbors. At the $i$th stage, we will have a tree $T_i$ where $V(T_i) = \bigcup_{i=0} L_i$, where $L_i$ is the set of vertices in level $i$. Note that $V_0 = L_0 = \{r\}$ and $T_1$ consists of 2 levels, $L_0 = \{r\}$ and $L_1 = N_{G'}(r)$. For two vertices $x$ and $y$, the unique path in $T_i$ from $x$ to $y$ will be denoted by $P(x, y)$.

Moreover at the $i$th stage, we will make sure that the tree $T_i$ satisfies the following property:

For any vertex $x \in V(T_i)$, $P(r, x)$ is a rainbow path. (1)
\[\text{Proof. Suppose not. Let } z \text{ be the least common ancestor of } x \text{ and } y \text{ in } T_i. \text{ Since } P(z, x) \subseteq P(r, x) \text{ is a rainbow path by property (1), } P(z, x) \cup \{(x, y)\} \text{ is also rainbow. Since } |P(z, y)| \leq |P(z, x)|, \text{ clearly we have } |\text{col}(P(z, x) \cup \{(x, y)\})| > |\text{col}(P(z, y))|. \text{ It follows that there is at least one matching color, say } i \text{ in the cycle } C = P(z, x) \cup \{(x, y)\} \cup P(y, z), \text{ such that } \lambda_{H_C}(i) = 1. \text{ But since } C \text{ is an even subgraph of } G, H_C \text{ is an even augmentation of } H \text{ by Claim 7. Then by Claim 5, } \lambda_{H_C}(i) \text{ should be an even number, a contradiction.} \]

Now we describe how to construct \(T_{i+1}\) from \(T_i\) by adding a new level \(L_{i+1}\). For a vertex \(x \in L_i\) define \(N'(x) = \{y \in N_G(x) : \text{col}(x, y) \cap \text{col}(P(r, x)) = \emptyset\}\). Observe that \(N'(x) \cap V(T_i) = \emptyset\). This follows from Claim 9, since if there is an edge in \(G'\) from a vertex \(x \in L_i\) to a vertex \(y \in L_j, j \leq i\) then \(\text{col}(x, y) \cap \text{col}(P(r, x)) \neq \emptyset\), and therefore \(y \notin N'(x)\). Define \(L_{i+1} = \cup_{x \in L_i} N'(x)\). Clearly \(L_{i+1} \cap V(T_i) = \emptyset\). If \(L_{i+1} \neq \emptyset\), define a bipartite graph \(B_i = (L_i, L_{i+1})\) such that for \(x \in L_i\) and \(y \in L_{i+1}\), \((x, y) \in E(B_i)\) if and only if \(y \in N'(x)\). Now for each \(y \in L_{i+1}\), select one vertex \(y'\) from \(L_i\) such that \((y, y') \in E(B_i)\) to be its parent. Thereby obtaining the new tree \(T_{i+1}\). From the way we defined \(N'(x)\) for \(x \in L_i\), it is clear that property (1) is satisfied by \(T_{i+1}\). If \(|L_{i+1}| \geq 2|L_i|\), we proceed to add the next level. Otherwise we stop the procedure and define the final tree \(T\) to be \(T_{i+1}\).

Let \(L_\ell\) be the last level added to the tree. Clearly \(\ell > 2\). We observe that \(\ell \leq \log n + 1\). Otherwise \(|L_{\ell-1}| \geq 2^{\ell-1} > n\), since \(|L_{i+1}| \geq 2|L_i|\) for \(i \leq \ell - 2\). Now consider the bipartite graph \(B_{\ell-1}\): For each vertex \(x \in L_{\ell-1}\), we know by applying Claim 8 with \(C = \text{col}(P(r, x))\) that \(E(x, C) \leq 4|\text{col}(P(r, x))| \leq 8 \log n\). But \(|E_{G'}(x)| = \deg_G(x) \geq C' \log n\) and therefore \(|N'(x)| \geq (C' - 8) \log n\). Therefore \(B_{\ell-1}\) has at least \(|L_{\ell-1}|(C' - 8) \log n\) edges. Since \(|L_\ell| < 2|L_{\ell-1}|\), there exists a vertex \(w \in L_\ell\) such that its degree in \(B_{\ell-1}\) is at least \((C'/2 - 4) \log n\). Again by applying Claim 8, this time with \(C = \text{col}(P(r, w))\), at most \(8(\log n + 1)\) of these edges can have a common color with any edge in \(P(r, w)\). It follows that if \(C\) is taken large enough, \(w\) has a neighbor \(w' \in L_{\ell-1}\) such that \(\text{col}(w, w') \cap \text{col}(P(r, w)) = \emptyset\). This contradicts Claim 9 applied to the tree \(T = T_\ell\). The situation is depicted in Figure 2. Thus we infer that \(k^2 \leq Cn \log n\), which in turn implies that \(n = \Omega\left(\frac{k^2}{\log k}\right)\).

\[\text{Figure 2 The cycle formed by the concatenation of } P(z, w'), P(z, w), \text{ and } (w, w') \text{ corresponds to an even subgraph in } H \text{ with an odd number of edges having a particular color.}\]
References


