RAINBOW PATHS

DOMINGOS DELLAMONICA JR., COLTON MAGNANT, AND DANIEL M. MARTIN

Abstract. A k-rainbow path in a graph with colored edges is a path of length k where each edge has a different color. In this note, we settle the problem of obtaining a constructive k-coloring of the edges of K_n in which one may find, between any pair of vertices, a large number of internally disjoint k-rainbow paths. In fact, our construction obtains the largest possible number of paths. This problem was considered in a less general setting by Chartrand et al. (2007).

1. Introduction

Given an edge-colored simple graph G, a path P in G is called *rainbow* if the edges of P are assigned distinct colors. Let $l \leq k$ be integers. Suppose that the edges of G are k-colored. For $a, b \in V(G)$, denote by $p(a, b)$ the maximum number of internally disjoint rainbow paths of length l having endpoints a and b. The *rainbow* (k, l)*-connectivity* of G is the minimum $p(a, b)$ among all distinct $a, b \in V(G)$.

A related concept has been studied in a sequence of papers by Chartrand et al. [CJMZ07, CJMZ07B, CJMZ08] and also [CLRTY08, JOZ08]. In particular, the following theorem is given.

Theorem 1 ([CJMZ08])**.** *For any r, there exists an explicit 2-coloring of K_r in which the number of bi-chromatic paths of length 2 between any pair of vertices is at least*

$$\lfloor \sqrt{r} - 1 \rfloor.$$

Using our definitions, the theorem above is a statement about the rainbow $(2, 2)$-connectivity of a given 2-coloring of the edges of K_r. In this note, we greatly improve and generalize the above lower bound for graphs of sufficiently large order by providing a different constructive coloring. Our construction attains asymptotically the maximum rainbow connectivity possible.

Theorem 2. *For any $k \geq 2$ and $r \geq r_0 = r_0(k)$ there exists an explicit k-coloring of the edges of K_r having rainbow $(k, 2)$-connectivity*

$$\Big(\frac{k-1}{k} - o(1)\Big)r.$$

More generally, we shall also consider the problem of finding longer rainbow paths.

The first author is supported by a CAPES-Fulbright scholarship.
The third author is supported by *CNPq*/Emory process no. 200729/2005-5.

Theorem 3 (Main result)**.** *For any $3 \leq l \leq k$ there exists $r_0 = r_0(k)$ such that, for every $r \geq r_0$, there is an explicit k-coloring of the edges of K_r having rainbow (k,l)-connectivity*

$$\big(1 - o(1)\big)\frac{r}{l-1}.$$

This result is also asymptotically best possible, since any collection of internally disjoint paths of length l can contain at most $r/(l-1)$ paths.

Our proof employs a very recent breakthrough due to Bourgain [Bou05, Rao07], which consists of a powerful explicit *extractor*. Roughly speaking, an (explicit) extractor is a polynomial time algorithm used to convert some special probability distributions into uniform distributions. (See [Sha02] for a good but somewhat outdated survey on extractors.)

The application of extractors in graph constructions already appears as early as [WZ99], where extractors are used directly to obtain good expander graphs. Similar applications followed, e.g. [CRVW02, TSUZ01]. Although some constructions are obtained by simply looking at extractors from a graph perspective (extractors can be seen as graphs), the analysis of our construction is more delicate and requires additional ingredients.

It is also noteworthy that a random k-coloring of a sufficiently large complete graph has asymptotically optimal rainbow (k,l)-connectivity. Therefore, our problem is to obtain an *explicit* edge coloring. By explicit, we mean that there is a polynomial time algorithm to compute such an edge coloring.[1]

2. Paths of length two

A simple application of the Cauchy-Schwarz Inequality shows that one cannot hope to find a k-coloring of $E(K_r)$ in which every pair of vertices is connected by $(1-1/k)(r-1)$ rainbow paths of length two.

Theorem 4. *For any k-coloring of the edges of K_r there exists a pair of vertices having at least*

$$\frac{r-1}{k} - 1$$

monochromatic paths of length 2 between them.

Proof. Denote by $\chi\colon E(K_r) \to [k]$ the fixed coloring of K_r. Let us count the triples (u,v,w) of different vertices satisfying $\chi(\{u,w\}) = \chi(\{v,w\})$. For fixed $w \in V = V(K_r)$ and $l \in [k]$, denote by $N(w,l) = \big|\big\{v \in V \,:\, \chi(\{v,w\}) = l\big\}\big|$ the number of vertices connected to v by an edge of color l. Then, the number of pairs $u \neq v \in V \setminus \{w\}$ such that uwv is a monochromatic path is given by

$$\sum_{l=1}^{k} N(w,l)\big(N(w,l)-1\big) = \sum_{l=1}^{k} N(w,l)^2 - r + 1.$$

Applying the Cauchy–Schwarz inequality we get that the number of such pairs is at least

$$\frac{1}{k}(r-1)^2 - (r-1).$$

[1]There are stronger notions of explicitness. For instance, one could ask for an algorithm to compute the color of an edge in polynomial time over the size of the input (the pair of vertices), which is $O(\log |V|)$.

Summing over every $w \in V$ and averaging over the pairs $u \neq v$, we have $(r-1)/k-1$, which implies that at least one pair has at least this number of monochromatic 2-paths connecting its points. □

Observe that, in view of Theorem 4, the result of Theorem 2 is asymptotically best possible. Theorem 2 follows from the proof of our main result, Theorem 3. We remark that a bipartite version of the same result follows using very similar techniques.

3. Extractors

In order to describe our construction, we need to introduce some background on the machinery of extractors, a subject intensively developed during the past two decades. In particular, we shall use Bourgain's recent breakthrough construction. We follow the presentation of [Rao07] in the brief recollection of Bourgain's results contained in this section.

We start by defining a way to measure randomness in (discrete) probability distributions. In what follows, we shall abuse notation by using the same letter to denote a probability distribution and a random variable following that distribution.

Definition 5. *A* source *is a probability distribution on binary strings of a fixed length. Let X be a source over $\{0,1\}^n$. The* min-entropy *of X is defined as*

$$H^\infty(X) = -\log\left(\max_{a\in\{0,1\}^n} \mathbf{P}[X=a]\right).$$

*Here, and throughout this manuscript, logarithms have base 2. We shall say that X is a δ-*source *if its* min-entropy rate *$r(X) = H^\infty(X)/n$ is at least δ.*

Definition 6. *A source having uniform probability over its support is called a* flat source.

An extractor is a function that converts some distribution into one that is close to uniform. To define precisely what it means for two distributions to be close, we state the notion of statistical difference (also known as total variation distance).

Definition 7. *The* statistical difference *between two sources $X, Y \subseteq \{0,1\}^n$, is defined as*[2]

$$\frac{1}{2}\|X-Y\|_1 = \frac{1}{2}\sum_{a\in\{0,1\}^n} \big|\mathbf{P}\big[X=a\big] - \mathbf{P}\big[Y=a\big]\big|.$$

*We say that X is α-*close *to Y if $\frac{1}{2}\|X-Y\|_1 \leq \alpha$.*

We are now ready to formally describe the extractor that will be used to define our constructive coloring.

Definition 8. *A function $E\colon \{0,1\}^{n\times 2} \to \{0,1\}^m$ is a* two-source-extractor *for min-entropy rate δ and error ε if, for any pair of independent sources X and Y satisfying $r(X), r(Y) \geq \delta$, the distribution of $E(X,Y)$ is ε-close to the uniform distribution over $\{0,1\}^m$.*

We say that the extractor E is strong *if the sets*

$$S = \big\{x \in \{0,1\}^n \,:\, E(x,Y) \text{ is } \varepsilon\text{-close to uniform}\big\}$$

[2]The 1/2 factor is used to keep the statistical distance in the range $[0,1]$.

and

$$T = \{y \in \{0,1\}^n \,:\, E(X,y) \text{ is } \varepsilon\text{-close to uniform}\}$$

are such that $\mathbf{P}[X \in S], \mathbf{P}[Y \in T] \geq 1 - \varepsilon$.

A classical construction of two-source-extractors is due to Vazirani, which generalized Hadamard matrices (corresponding to the case where $m = 1$ in Theorem 9).

Theorem 9 ([Vaz87, DEOR04])**.** *For every constant $\delta > 0$, there exists a polynomial time strong two-source-extractor* $\texttt{Had}\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$*, with $m = \Omega_\delta(n)$ and error $\varepsilon = 2^{-\Omega_\delta(n)}$, that works with any independent sources X and Y having $H^\infty(X) + H^\infty(Y) \geq (1+\delta)n$.*

Bourgain's idea was to first encode the input bits (and produce a larger, redundant string) and then apply the Hadamard extractor of Theorem 9. To describe this encoding, we shall use some basic Finite Field Theory (the reader is referred to the textbook of Dummit and Foote [DF99]).

Let g be a primitive element of $\mathrm{GF}(2^n)$ (*i.e.*, a generator of the multiplicative group $\mathrm{GF}(2^n)^\times$). Every element of $\mathrm{GF}(2^n)$ can be seen as an integer in $[0, 2^n - 1]$ and, clearly, also can be seen as an element of $\{0,1\}^n$. If $x \in \mathrm{GF}(2^n)$ let $\#x$ denote the integer corresponding to x. We define $g^x = g^{\#x}$. Bourgain's extractor can be taken as

$$\texttt{Bou}(x,y) = \texttt{Had}\big(\bar{x}, \bar{y}\big), \tag{1}$$

where the encoding $\bar{\cdot}\colon \{0,1\}^n \to \{0,1\}^{2n}$ is given by

$$x \mapsto \bar{x} = (x, g^x). \tag{2}$$

The above extractor is able to inherit many useful properties of the Hadamard extractor, such as being symmetric and strong. The redundant encoding of x and y serves a purpose: in a more sophisticated analysis of the Hadamard extractor, a source X works essentially as well as tX, the distribution obtained by adding t independent samples from X, with only a polynomial loss in the error parameter ε (*i.e.*, instead of getting error ε we get error ε^{-a} for some constant $a > 0$). Hence, if tX has higher min-entropy—in the case X *grows with addition*—one can extract randomness even if X has min-entropy smaller than what is needed by $\texttt{Had}$. The encoding (2) is proven to grow with addition, as asserted by Lemma 10.

Lemma 10. *Let $3\bar{X} = \bar{X} + \bar{X} + \bar{X}$ denote the distribution induced by sampling three independent elements x_1, x_2, x_3 from X and outputing $(x_1, g^{x_1}) + (x_2, g^{x_2}) + (x_3, g^{x_3})$. Then there exists an absolute constant $\alpha > 0$ such that $H^\infty(3\bar{X}) \geq (1+\alpha) \cdot 2 \cdot H^\infty(X)$. In terms of min-entropy rates, $r(3\bar{X}) \geq (1+\alpha)\, r(X)$.*

The price one has to pay in order to use $3\bar{X}$ instead of $\bar{X}$ in the analysis of the extractor is an increase of the output error as described by Lemma 11.

Lemma 11. *Let X and Y be independent sources over $\{0,1\}^n$ and suppose that $H^\infty(t_1 X) \geq k_1$, $H^\infty(t_2 Y) \geq k_2$. Let* $\texttt{Had}\colon \{0,1\}^{n\times 2} \to \{0,1\}^m$ *be the Hadamard extractor and*

$$\varepsilon = \exp\Big\{\frac{n - k_1 - k_2}{c} + m\Big\},$$

for $c = c(t_1, t_2)$. Then $\texttt{Had}(X,Y)$ *is ε-close to the uniform distribution over $\{0,1\}^m$.*

In order to have ε small in the above lemma, one needs to have sufficient min-entropy from t_1X_1 and t_2Y and m, the length of the output, must be sufficiently small. Given Lemma 10, after encoding (2), sources with min-entropy rate slightly smaller than $1/2$ grow with addition into sources with min-entropy rate slightly greater than $1/2$. Applying Lemma 11 to those sources allows us to make m linear in n, while keeping the error exponentially small.

Theorem 12 ([Bou05, Rao07])**.** *There exists an absolute constant $\nu > 0$ such that the (explicit) function* $\texttt{Bou}\colon \{0,1\}^{n\times 2} \to \{0,1\}^m$ *is a strong extractor for sources of min-entropy $(1/2-\nu)n$ with $m = \Omega(n)$ and, furthermore, has error $\varepsilon = 2^{-\Omega(n)}$.*

Another important observation regarding the above encoding is the following.

Claim 13. *Given any X, we have $H^\infty(2\bar{X}) \geq 2H^\infty(X) - 1$.*

Proof. First, let us change variables by setting $y = g^x$ and then $(x, g^x) = (\log_g y, y)$. (We may assume that $x \neq 0$ by paying a small price in terms of min-entropy, for instance, one can "move" the probability from 0 to an arbitrary fixed string). Since we have $\log_g y_1 + \log_g y_2 = \log_g y_1y_2$, an element of $2\bar{X} = \bar{X} + \bar{X}$ would be of the form $\mathbf{x} = (\log_g y_1y_2, y_1 + y_2)$. If we have $\mathbf{x} = (\log_g y_3y_4, y_3 + y_4)$ then the following system of equations must be satisfied

$$\begin{cases} y_1y_2 = y_3y_4 \\ y_1 + y_2 = y_3 + y_4. \end{cases}$$

Since this field has characteristic 2, we have $y_1 + y_3 = y_2 + y_4$ and it follows that

$$\begin{aligned} y_1^2 + y_3^2 = (y_1 + y_3)^2 = (y_1 + y_3)(y_2 + y_4) \\ = (y_1y_2 + y_3y_4) + y_1y_4 + y_2y_3 = y_1y_4 + y_2y_3. \end{aligned} \tag{3}$$

Hence, $y_1(y_1 + y_4) = y_3(y_2 + y_3)$ and we also have $y_1 + y_4 = y_2 + y_3$. Therefore either $y_1 = y_4$ (and $y_2 = y_3$) or $y_1 = y_3$ (and $y_2 = y_4$). Hence, an element of $2\bar{X}$ is the image of at most 2 pairs in $\bar{X} \times \bar{X}$. The probability of any such element is thus at most 2 times the maximum probability of a pair in $\bar{X}^2$. By the independence assumption, a pair in $\bar{X}^2$ has probability bounded by $2^{-2H^\infty(X)}$. It follows that an element in $2\bar{X}$ has probability bounded by $2 \times 2^{-2H^\infty(X)} = 2^{-2H^\infty(X)+1}$ and the claim follows. □

From this we get that another interesting property of the Hadamard extractor is inherited by Bourgain's extractor.

Lemma 14. *For any positive constant γ, there is $m = \Omega(\gamma n)$ for which*

$$\texttt{Bou}\colon \{0,1\}^{n\times 2} \to \{0,1\}^m,$$

as defined in (1) *is a strong extractor for any X and Y such that either*
A) $H^\infty(X) + H^\infty(Y) \geq (1+\gamma)n$ or
B) $H^\infty(X), H^\infty(Y) \geq (1/2-\nu)n$, where ν is the constant of Theorem 12.
Moreover, Bou *has error $\varepsilon = 2^{-\Omega(\gamma n)}$.*

Proof. Let $m_1 = -\frac{\gamma n}{2c(2,2)}$, where $c(\cdot,\cdot)$ is the implicit function in Lemma 11. Let $m_2 = \Omega(n)$ be the output length of the Bourgain extractor of Theorem 12. Set $m = \min\{m_1, m_2\}$ and let the output length of both Had and Bou be m in what follows.

A: Given independent sources X and Y such that $H^\infty(X)+H^\infty(Y) = (1+\gamma)n$, by Claim 13, $H^\infty(2\bar{X}) + H^\infty(2\bar{Y}) \geq (1+\gamma)2n - 2$. By Lemma 11, $\texttt{Bou}(X,Y) = \texttt{Had}(\bar{X}, \texttt{Had}Y)$ is $2^{-\Omega(\gamma n)}$-close to the uniform distribution.

B: If X and Y satisfy $H^\infty(X), H^\infty(Y) \geq (1/2-\nu)n$ then the conclusion is immediate from Theorem 12. □

4. Coloring the edges of the complete graph

Let us define a coloring of the complete graph on 2^n vertices. Fix $\texttt{Bou}$ as the extractor of Lemma 14 with $\gamma = \nu/2$ and suppose that m is the output length and ε is the error of the extractor. Let the vertex-set of K_{2^n} be $V = \{0,1\}^n$. Let $S_1, \ldots, S_k$ be a balanced partition of $\{0,1\}^m$ (that means $\big||S_i| - |S_j|\big| \leq 1$ for every i, j). Let $f\colon \{0,1\}^m \to [k]$ be defined as $f(x) = i$ iff $x \in S_i$. An edge $\{u,v\} \in \{0,1\}^{n\times 2}$ gets color $\chi(\{u,v\}) = f\big(\texttt{Bou}(u,v)\big)$ (this is well defined since $\texttt{Bou}$ is symmetric).

Note that we are coloring a graph on 2^n vertices. This coloring could have several local flaws (pairs of vertices which have low rainbow connectivity). We shall later take an edge-coloring induced by an appropriate subset of these vertices as the final coloring. More concretely, we have $n = O(\log r)$, where r is the order of the final graph.

Define $\Gamma(v,l) = \big\{w \in V \,:\, \texttt{Bou}(v,w) \in S_l\big\}$ and set $N(v,l) = |\Gamma(v,l)|$.

Lemma 15. *The above coloring satisfies the following for all but at most $k\,2^{\nu n/2}$ vertices v: for any color $l \in [k]$, we have $N(v,l) \geq (1-2k\varepsilon)2^n/k$.*

Proof. Let T_l be a set of vertices v having $N(v,l) < (1-2k\varepsilon)2^n/k$. Assume that $|T_l| \geq 2^{\nu n/2}$. Let X_l be a flat source over T_l (thus having $H^\infty(X_l) \geq \nu n/2$). Let U_n be uniform over $V = \{0,1\}^n$. By Lemma 14, the distribution $\texttt{Bou}(X_l, U_n)$ should be ε-close to uniform. But this leads to the following contradiction:

$$2^{-m}|S_l| - \varepsilon \leq \mathbf{P}\big[\texttt{Bou}(X_l,U_n) \in S_l\big] = \frac{1}{|T_l|}\sum_{v\in T_l}\frac{N(v,l)}{|V|} < (1-2k\varepsilon)\frac{1}{k}.$$

Hence $|T_l| < 2^{\nu n/2}$ and $\big|\bigcup_{l\in[k]} T_l\big| \leq k\,2^{\nu n/2}$. □

From this we can also conclude the following about pairs of vertices.

Lemma 16. *For any vertex $v \in V$ such that $N(v,l) \geq (1-2k\varepsilon)2^n/k$ holds for all $l \in [k]$ there are at most $k^2 2^{\nu n/2+2k}$ vertices w such that, for some $(j,l) \in [k]$, we have*

$$\big|\Gamma(w,j) \cap \Gamma(v,l)\big| < (1-2k\varepsilon)^2\frac{2^n}{k^2}. \tag{4}$$

Proof. The proof is similar to that of Lemma 15. For some fixed pair (j,l), let $T_{j,l}$ be the set of vertices w for which (4) holds and assume $|T_{j,l}| \geq 2^{\nu n/2+2k}$. Let $X_{j,l}$ be the flat source over $T_{j,l}$ and Y be the flat source over $\Gamma(v,l)$. Observe that $H^\infty(Y) \geq n + \log\{1/k - 2\varepsilon\} \geq n - \log 2k$ and hence, $H^\infty(X_{j,l}) + H^\infty(Y) \geq (1+\nu/2)n$. It follows that $\texttt{Bou}(X_{j,l}, Y)$ is ε-close to uniform. On the other hand,

$$\frac{|S_j|}{2^m} - \varepsilon \leq \mathbf{P}\big[\texttt{Bou}(X_{j,l},Y) \in S_j\big] = \frac{1}{|T_{j,l}|}\sum_{w\in T_{j,l}}\frac{|\Gamma(w,j)\cap\Gamma(v,l)|}{N(v,l)} < (1-2k\varepsilon)\frac{1}{k},$$

which is a contradiction. Therefore, $\big|\bigcup_{(j,l)\in[k]^2} T_{j,l}\big| \leq k^2\,2^{\nu n/2+2k}$. □

Along the same lines we have the following.

Lemma 17. *Given two sets $X, Y \subseteq V$ with $|X|, |Y| \geq k\, 2^{(1/2-\nu)n}$ all but at most $k\, 2^{(1/2-\nu)n}$ vertices $x \in X$ have edges of all k colors going to Y.*

4.1. **Randomly selecting a subgraph.** Although the coloring provided by the Bourgain extractor is such that for *most* pairs (a, b), the number of k-rainbow paths between a and b is very close to the best possible, we do not have any guarantee that this holds for *all* pairs. In order to deal with this technicality, we first show that randomly selecting a small subset of vertices of K_{2^n} (much less than $\sqrt{2^n}$) the coloring induced on the edges spanned by those vertices is one satisfying the requirements. The use of the Bourgain extractor (instead of the classical Hadamard extractor) is justified if one wishes to analyse the coloring induced by such a small set of vertices. Since our final goal is a constructive coloring, we shall derandomize the vertex selection in Subsection 4.2.

Let $p = 2^{-(1+\nu)n/2}$. Assume that we pick elements from $\{0,1\}^n$ uniformly and independently with probability p forming some set V'. Consider the coloring formed in the induced graph V'. Observe that the expected cardinality of $|V'|$ is $2^{(1-\nu)n/2}$. Let us say that a pair $(v, w) \in V$ is *bad* (with respect to V) if either v or w fails Lemma 15 for some color or if (4) holds for some $(j, l) \in [k]^2$. A pair is called *good* otherwise.

Let us estimate the expected number of bad pairs that are contained in V'. The expected number of pairs containing (at least) one vertex failing Lemma 15 is at most $p^2 \cdot k2^{\nu n/2} \cdot 2^n = k2^{-\nu n/2}$. The expected number of pairs $v \neq w \in V$ such that both do not fail Lemma 15 but (4) holds for some $(j, l) \in [k]^2$ is, by Lemma 16, $p^2 \cdot k^2 2^{\nu n/2+2k} \cdot 2^n = k^2 2^{-\nu n/2+2k}$.

We also estimate the expected number of good pairs of V that are bad with respect to V': that is, $u \neq v \in V$ is not a bad pair but, for some fixed $\gamma > 0$ and some $(j, l) \in [k]^2$,

$$\big|\Gamma(w, j) \cap \Gamma(v, l) \cap V'\big| < (1-\gamma)(1-2k\varepsilon)^2 \frac{2^{(1-\nu)n/2}}{k^2}. \tag{5}$$

It follows by Chernoff's inequality (see [Hoe63]) and the union bound that the expected number of quadruples (w, v, j, l) satisfying (5) is at most

$$k^2 \cdot 2^{2n} \cdot \exp\Big\{-cp(1-2k\varepsilon)^2 \frac{2^n}{k^2}\Big\},$$

where $c = c(\gamma) > 0$.

The number of bad pairs with respect to V' has expectation $o(1)$. We also have that $|V'|$ is strongly concentrated around its expectation. By choosing an appropriate value of p, we have $|V'| \in [r, r + r^{2/3}]$ with probability $1 - o(1)$. By removing at most $r^{2/3}$ arbitrary vertices, we get the final graph with the prescribed number r of vertices. Observe that since we are removing a very small number of vertices in the end, the effect on equation (5) is negligible. We shall derandomize the selection of V' in order to obtain a graph in which every pair is good (with respect to V').

4.2. **Derandomization.** In order to obtain a constructive coloring of the edges of the complete graph, we have to derandomize the random choices made above. So far, the following procedure has been defined: some large enough n is set and a coloring of the edges K_{2^n} is given by projecting the output of Bourgain's extractor

onto the set of colors. A random vertex subset of this graph is taken and, with probability $1 - o(1)$, the induced edge-coloring has the desired properties.

We would like to stress that the reason we take such a strong construction as the Bourgain extractor (instead of relying on the well-known Hadamard extractor) lies in the fact that we must select very few vertices of the initial complete graph in order to ensure that no two vertices form a bad pair. On the other hand, we must be able to say something about the distribution of the colors in subsets of the induced subgraph, and those subsets are very small relative to the original graph.

The derandomization technique that we shall use is the *Method of Conditional Expectations* [AS00]. The random induced subgraph is determined by picking vertices independently with probability p. Suppose that the random decisions have been made for all vertices in a subset S of the vertex set V. Namely, for each $v \in S$ a random decision has been taken and a subset $S' = S \cap V'$ has been selected. The remaining choices (for vertices in $V \setminus S$) are independent from the choices made for S and it is simple to compute expectations in the conditional space where $S' \subseteq S$ is fixed.

Let $N = N_0 + N_1 + N_2$, N_0, N_1 and N_2 be the random variables such that N_0 counts the number of bad pairs; N_1 counts the number of good pairs in V that are not good with respect to V', namely, for some $(j, l) \in [k]^2$, they satisfy (5); and $N_2 = 2\big||V'| - (r + r^{2/3}/2)\big|/r^{2/3}$. We showed that $\mathbf{E}[N_i] = o(1)$ for $i = 1, 2, 3$ (if we set $p = (r + r^{2/3}/2)2^{-n}$).

Note that, given $S' \subseteq S$, we can compute the conditional expectation $\mathbf{E}[N \mid V' \cap S = S']$ in polynomial-time (over $2^n = \text{poly}(r)$). Indeed, computing N_0 is just a matter of enumerating all bad pairs in K_{2^n} and adding the conditional probability of each one being selected. Computing N_1 requires a somewhat similar computation: for each pair of vertices and each pair of colors, the conditional probability of satisfying (5) is readily evaluated. Clearly, N_2 is computed in constant time. The method works as follows: initially, $S = \emptyset$; given an arbitrary $v \in V \setminus S$, we decide whether vertex v will be selected by computing two conditional expectations, $E_1 = \mathbf{E}[N \mid V' \cap (S + v) = S' + v]$ and $E_2 = \mathbf{E}[N \mid V' \cap (S + v) = S']$. Note that

$$E_0 = \mathbf{E}[N \mid V' \cap S = S'] = pE_1 + (1 - p)E_2.$$

Clearly, $\min\{E_1, E_2\} \leq E_0$. Take $i \in \{1, 2\}$ such that $E_i \leq E_0$ and put v in S' if and only if $i = 1$. Update $S \leftarrow S + v$ and repeat.

Eventually the set S exhausts all elements of V. When that happens, there is a deterministically chosen set $S' = V' \subseteq V$ for which $N = 0$ (since $N = o(1)$ is an integer). It follows that $|V'| \in [r, r + r^{2/3}]$ and every pair of vertices in V' does not satisfy (5) for any $(j, l) \in [k]^2$. Remove an arbitrary set of vertices from V' so that we get $|V'| = r$ and let the induced colored graph define the coloring of K_r.

In practice, to speed up the process, once a vertex v is chosen to be part of S', every vertex which forms a bad pair with v is known not to belong to S' and we can update S to contain all these vertices while S' only gets v.

5. Internally disjoint rainbow paths

The powerful construction of Bourgain's extractor resembles a random structure in such a way that a greedy algorithm to find disjoint k-rainbow paths, which can easily be seen to work (almost surely) in a random coloring, also works with this constructive coloring.

Proof of Theorem 3. After the derandomization of Subsection 4.2, we obtain a coloring of K_r (with $V' = V(K_r)$) in which, for every pair $v \neq w$ and every pair of colors $(j,l) \in [k]^2$,

$$\left|\Gamma(w,j) \cap \Gamma(v,l) \cap V'\right| \geq (1-o(1))\frac{r}{k^2}. \tag{6}$$

Observe that this also implies an upper bound $\left|\Gamma(w,j)\cap\Gamma(v,l)\cap V'\right| \leq \big(1+o(1)\big)r/k^2$ for all $(j,l) \in [k]^2$.

For simplicity, in this proof we shall consider the case where the length l of the rainbow paths is equal to k. The same proof yields the result for all $3 \leq l \leq k$ with obvious modifications.

Let us fix some pair $v \neq w$ and check that there are many disjoint k-rainbow paths between them. In order to find those paths we use greedy Algorithm 1. We first classify vertices according to the color of the edges connecting them to both v and w: if the edge xv has color j and the edge xw has color l, we put x in X_{jl}. We wish to use vertices from those classes in a uniform manner, so that they always have roughly the same cardinality. This is done by choosing the $k-1$ largest classes $X_{j_1l_1}, \ldots, X_{j_{k-1}l_{k-1}}$ and reordering them[3] so that $j_1 \neq l_{k-1}$. Since the distribution of colors of edges between the classes should be very uniform whenever the classes have at least $k2^{1+(1/2-\nu)}$ elements, if the classes are large, we may find a sequence of vertices $v_m \in X_{j_m,l_m}$ for $m = 1, \ldots, k-1$, $v_0 = v$, $v_k = w$ such that the edges v_iv_{i+1} have all distinct colors.

To simplify the notation, we denote $X \cup \{x\}$ by $X + x$ and, similarly, $X \setminus \{x\}$ is denoted $X - x$.

It is straightforward to check that Algorithm 1 either aborts or obtains a collection of internally disjoint rainbow k-paths. Let us first prove that the algorithm does not abort (when the coloring is given by our construction).

The only way this algorithm aborts is if there are two sets, $Y = Y_{m+1}$ and $X = X_{j_ml_m}$ with $|X| > 2k2^{(1/2-\nu)n}$ and $|Y| \geq k2^{(1/2-\nu)n}$ such that less than $k2^{(1/2-\nu)n}$ vertices $x \in X$ have edges of all colors going to Y. But this is a contradiction with Lemma 17.

To prove that the path system $\mathcal{P}$ is large, we observe that by selecting the largest classes on line 5, the cardinalities of the sets X_{jl} become balanced. This is formalized, when $k \geq 4$, by Claim 18. A similar claim holds for $k = 3$ with some slight modifications on the argument.

Claim 18. *Suppose that there is a collection of k^2 positive numbers such that the maximum difference between them is at most Δ. If a procedure takes the $k-1$ largest elements and decreases them by one at each step, after at most $\Delta(k+3)$ steps, the maximum difference between any pair becomes bounded by 1.*

Proof. Let M_i be the largest element and m_i be the smallest element at the ith step. We denote by $T_{i,j}$ the jth greatest element at the ith step.

At any given step i, we have the following possibilities:

(i) $M_i \leq m_i + 1$: the next step also satisfies case (i);

[3]The case $k = 3$ (or $l = 3$ in the general case) is slightly more complicated and we may have to replace one of the classes by a smaller one to find such ordering.

Algorithm 1: Finding disjoint paths

Input: vertices $v \neq w$.

1 $\mathcal{P} \leftarrow \emptyset$;
2 **foreach** $(j,l) \in [k]^2$ **do**
3 $\quad X_{jl} \leftarrow \Gamma(v,j) \cap \Gamma(w,l) \cap V'$;
4 **while** $\min_{(j,l)\in[k]^2} |X_{jl}| \geq k2^{1+(1/2-\nu)n}$ **do**
5 $\quad$ let $X_{j_1l_1}, \ldots, X_{j_{k-1}l_{k-1}}$ be some collection of $k-1$ sets satisfying $j_1 \neq l_{k-1}$ and having maximum $\sum_{m=1}^{k-1} \left|X_{j_ml_m}\right|$;
6 $\quad Y_{k-1} \leftarrow X_{j_{k-1}l_{k-1}}$;
7 $\quad$ **for** $m \leftarrow k-2$ **downto** 1 **do**
8 $\quad\quad Y_m \leftarrow \left\{x \in X_{j_ml_m} : \Gamma(x,j) \cap Y_{m+1} \neq \emptyset \text{ for all } j = 1, \ldots, k\right\}$;
9 $\quad\quad$ **if** $|Y_m| < k2^{(1/2-\nu)n}$ **then**
10 $\quad\quad\quad$ **abort**;
11 $\quad$ pick $v_1 \in Y_1$;
12 $\quad X_{j_1l_1} \leftarrow X_{j_1l_1} - v_1$;
13 $\quad K \leftarrow \{j_1, l_{k-1}\}$;
14 $\quad$ **for** $m = 2$ **to** $k-1$ **do**
15 $\quad\quad c \leftarrow \min([k] \setminus K)$;
16 $\quad\quad K \leftarrow K + c$;
17 $\quad\quad$ pick $v_m \in \Gamma(v_{m-1}, c) \cap Y_m$;
18 $\quad\quad X_{j_ml_m} \leftarrow X_{l_ml_m} - v_m$;
19 $\quad \mathcal{P} \leftarrow \mathcal{P} + vv_1v_2 \cdots v_{k-1}w$;

(ii) $M_{i+1} = M_i \geq m_i + 2$ and $m_{i+1} = m_i$: this can only happen if $T_{i,k} = M_i$. Note that (ii) can hold for at most $k+1$ steps (since each time it happens, $k-1$ numbers equal to M_i are decreased).

(iii) $M_{i+1} = M_i - 1 \geq m_i + 1$ and $m_{i+1} = m_i - 1$: in this case we must have $T_{i,k-1} = T_{i,k} = \cdots = T_{i,k^2} = m_i$. Observe that a step in which (iii) holds cannot be followed by a step in which (ii) or (iii) holds.

(iv) $M_i \geq m_i + 2$, $T_{i,k} < M_i$ and $T_{i,k-1} > m_i$: in this case, we have $M_{i+1} = M_i - 1$ and $m_{i+1} = m_i$.

Before we reach case (i), there can be at most $k+2$ steps consecutive steps in which only either (ii) or (iii) occurs. Every time (iv) occurs, the difference between the largest and smallest element decreases by 1. Hence, in at most $\Delta(k+3)$ steps we must reach case (i). This completes the proof of Claim 18. □

Note that, for the family of sets $\{X_{jl}\}_{(j,l)\in[k]^2}$, the initial difference is

$$\Delta = \max_{(j,l)\in[k]^2} |X_{jl}| - \min_{(j,l)\in[k]^2} |X_{jl}| = o(r/k^2).$$

Hence, after $o(r/k)$ steps, the sets are all balanced. In particular, since the procedure described by Claim 18 never increases the maximum difference, the condition of the **while** loop (line 4) remains true throughout the balancing process. This shows that, when this loop finishes, every X_{jl} has cardinality at most $k2^{1+(1/2-\nu)n} + 1$. Therefore, $o(r)$ vertices remain in $\bigcup X_{jl}$, all of the other vertices are used in (internally disjoint) k-paths of $\mathcal{P}$, thus proving the theorem. □

From the same constructive edge-coloring, we get Theorem 2.

Proof of Theorem 2. It suffices to observe that the construction of Theorem 3 satisfies equation (6) for every pair of vertices and every pair of colors. In particular, summing the left side of the inequality (6) over $j \neq l$, we are counting the number of bi-chromatic paths of length 2 between v and w. The same sum on the right side of the inequality results in $k(k-1)\big(1-o(1)\big)\frac{r}{k^2}$ and therefore we conclude the proof of the theorem. □

References

[AS00] Noga Alon and Joel Spencer. *The Probabilistic Method*. John Wiley and Sons, New York, 2nd edition, 2000.

[Bou05] J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1:1–32, 2005.

[CLRTY08] Yair Caro, Arie Lev, Yehuda Roditty, Zsolt Tuza and Raphael Yuster. On Rainbow Connection. Elec. J. Comb. 15(1) (2008) R57.

[CJMZ08] Gary Chartrand, Garry L. Johns, Kathleen A. McKeon, Ping Zhang. The Rainbow Connectivity of a Graph. Networks - To appear.

[CJMZ07] Gary Chartrand, Garry L. Johns, Kathleen A. McKeon, Ping Zhang. Rainbow Connection in Graphs. Math. Bohem. - 133 (2008) 85-98.

[CJMZ07B] Gary Chartrand, Garry L. Johns, Kathleen A. McKeon, Ping Zhang. The Rainbow Connectivity of Cages. Congr. Numer. 184 (2007) 209-222.

[CRVW02] Michael Capalbo, Omer Reingold, Salil Vadhan, and Avi Wigderson. Randomness conductors and constant-degree lossless expanders. In *STOC '02: Proceedings of the thiryfourth annual ACM symposium on Theory of computing*, pages 659–668, New York, NY, USA, 2002. ACM Press.

[DEOR04] Dodis, Elbaz, Oliveira, and Raz. Improved randomness extraction from two independent sources. In *RANDOM: International Workshop on Randomization and Approximation Techniques in Computer Science*. LNCS, 2004.

[DF99] David S. Dummit and Richard M. Foote. *Abstract algebra*. Prentice Hall Inc., second edition, 1999.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

[JOZ08] G. L. Johns, F. Okamoto, P. Zhang The rainbow connectivities of small cubic graphs. *Ars. Combin.*, To appear.

[Rao07] Anup Rao. An exposition of Bourgain's 2-source extractor. In *ECCCTR: Electronic Colloquium on Computational Complexity, technical reports*, 2007.

[Sha02] R. Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, pages 67–95, 2002.

[TSUZ01] Amnon Ta-Shma, Christopher Umans, and David Zuckerman. Loss-less condensers, unbalanced expanders, and extractors. In *STOC: ACM Symposium on Theory of Computing (STOC)*, 2001.

[Vaz87] Umesh Vazirani. Strong communication complexity or generating quasi-random sequences from two communicating semi-random sources. *Combinatorica*, 7(4):375–392, 1987.

[WZ99] Avi Wigderson and David Zuckerman. Expanders that beat the eigenvalue bound: Explicit construction and applications. *Combinatorica*, 19(1):125–138, 1999.

E-mail address: ddellam@mathcs.emory.edu

Emory University, Department of Mathematics and Computer Science, 400 Dowman Dr, Atlanta, GA 30322 – USA.