HARDNESS AND ALGORITHMS FOR RAINBOW CONNECTIVITY

SOURAV CHAKRABORTY [1] AND ELDAR FISCHER [1] AND ARIE MATSLIAH [2] AND RAPHAEL YUSTER [3]

[1] Department of Computer Science, Technion, Haifa 32000, Israel.
E-mail address, S. Chakraborty: `sourav@cs.technion.ac.il`
E-mail address, E. Fischer: `eldar@cs.technion.ac.il`

[2] Centrum Wiskunde & Informatica (CWI), Amsterdam, Netherlands.
E-mail address, A. Matsliah: `ariem@cwi.nl`

[3] Department of Mathematics, University of Haifa, Haifa 31905, Israel.
E-mail address, R. Yuster: `raphy@math.haifa.ac.il`

ABSTRACT. An edge-colored graph G is *rainbow connected* if any two vertices are connected by a path whose edges have distinct colors. The *rainbow connectivity* of a connected graph G, denoted $rc(G)$, is the smallest number of colors that are needed in order to make G rainbow connected. In addition to being a natural combinatorial problem, the rainbow connectivity problem is motivated by applications in cellular networks. In this paper we give the first proof that computing $rc(G)$ is NP-Hard. In fact, we prove that it is already NP-Complete to decide if $rc(G) = 2$, and also that it is NP-Complete to decide whether a given edge-colored (with an unbounded number of colors) graph is rainbow connected. On the positive side, we prove that for every $\epsilon > 0$, a connected graph with minimum degree at least ϵn has bounded rainbow connectivity, where the bound depends only on ϵ, and the corresponding coloring can be constructed in polynomial time. Additional non-trivial upper bounds, as well as open problems and conjectures are also presented.

1. Introduction

Connectivity is perhaps the most fundamental graph-theoretic property, both in the combinatorial sense and the algorithmic sense. There are many ways to strengthen the connectivity property, such as requiring hamiltonicity, k-connectivity, imposing bounds on the diameter, requiring the existence of edge-disjoint spanning trees, and so on.

An interesting way to quantitatively strengthen the connectivity requirement was recently introduced by Chartrand et al. in [5]. An edge-colored graph G is *rainbow connected* if any two vertices are connected by a path whose edges have distinct colors. Clearly, if a graph is rainbow connected, then it is also connected. Conversely, any connected graph has a trivial edge coloring that makes it rainbow connected; just color each edge with a distinct color. Thus, one can properly define the

The work was done when Sourav Chakraborty was a Phd student at University of Chicago and Arie Matsliah was a Phd student at Technion, Israel.

SYMPOSIUM ON THEORETICAL ASPECTS OF COMPUTER SCIENCE

rainbow connectivity of a connected graph G, denoted $rc(G)$, as the smallest number of colors that are needed in order to make G rainbow connected. An easy observation is that if G is connected and has n vertices then $rc(G) \leq n-1$, since one may color the edges of a given spanning tree with distinct colors. We note also the trivial fact that $rc(G)=1$ if and only if G is a clique, the (almost) trivial fact that $rc(G)=n-1$ if and only if G is a tree, and the easy observation that a cycle with $k>3$ vertices has rainbow connectivity $\lceil k/2 \rceil$. Also notice that, clearly, $rc(G) \geq \mathrm{diam}(G)$ where $\mathrm{diam}(G)$ denotes the diameter of G.

Chartrand et al. computed the rainbow connectivity of several graph classes including complete multipartite graphs [5]. Caro et al. [6] considered the extremal graph-theoretic aspects of rainbow connectivity. They proved that if G is a connected graph with n vertices and with minimum degree 3 then $rc(G) < 5n/6$, and if the minimum degree is δ then $rc(G) \leq \frac{\ln \delta}{\delta} n(1+f(\delta))$ where $f(\delta)$ tends to zero as δ increases. They also determine the threshold function for a random graph $G(n,p(n))$ to have $rc(G)=2$. In their paper, they conjecture that computing $rc(G)$ is an NP-Hard problem, as well as conjecture that even deciding whether a graph has $rc(G)=2$ in NP-Complete.

In this paper we address the computational aspects of rainbow connectivity. Our first set of results solve, and extend, the complexity conjectures from [6]. Indeed, it turns out that deciding whether $rc(G)=2$ is an NP-Complete problem. Our proof is by a series of reductions, where on the way it is shown that 2-rainbow-colorability is computationally equivalent to the seemingly harder question of deciding the existence of a 2-edge-coloring that is required to rainbow-connect only vertex pairs from a prescribed set.

Theorem 1.1. *Given a graph G, deciding if $rc(G)=2$ is NP-Complete. In particular, computing $rc(G)$ is NP-Hard.*

Suppose we are given an edge coloring of the graph. Is it then easier to verify whether the colored graph is rainbow connected? Clearly, if the number of colors in constant then this problem becomes easy. However, if the coloring is arbitrary, the problem becomes NP-Complete:

Theorem 1.2. *The following problem is NP-Complete: Given an edge-colored graph G, check whether the given coloring makes G rainbow connected.*

For the proof of Theorem 1.2, we first show that the $s-t$ version of the problem is NP-Complete. That is, given two vertices s and t of an edge-colored graph, decide whether there is a rainbow path connecting them.

We now turn to positive algorithmic results. Our main positive result is that connected n-vertex graphs with minimum degree $\Theta(n)$ have *bounded* rainbow connectivity. More formally, we prove:

Theorem 1.3. *For every $\epsilon > 0$ there is a constant $C = C(\epsilon)$ such that if G is a connected graph with n vertices and minimum degree at least ϵn, then $rc(G) \leq C$. Furthermore, there is a polynomial time algorithm that constructs a corresponding coloring for a fixed ϵ.*

The proof of Theorem 1.3 is based upon a modified degree-form version of Szemerédi's Regularity Lemma that we prove and that may be useful in other applications. From our algorithm it is also not hard to find a probabilistic polynomial time algorithm for finding this coloring with high probability (using on the way the algorithmic version of the Regularity Lemma from [1] or [7]).

We note that connected graphs with minimum degree ϵn have bounded diameter, but the latter property by itself does *not* guarantee bounded rainbow connectivity. As an extreme example, a star with n vertices has diameter 2 but its rainbow connectivity is $n-1$. The following theorem asserts however that having diameter 2 and only logarithmic minimum degree suffices to guarantee rainbow connectivity 3.

Theorem 1.4. *If G is an n-vertex graph with diameter 2 and minimum degree at least $8\log n$ then $rc(G) \le 3$. Furthermore, such a coloring is given with high probability by a uniformly random 3-edge-coloring of the graph G, and can also be found by a polynomial time deterministic algorithm.*

Since a graph with minimum degree $n/2$ is connected and has diameter 2, we have as an immediate corollary:

Corollary 1.5. *If G is an n-vertex graph with minimum degree at least $n/2$ then $rc(G) \le 3$.*

The rest of this paper is organized as follows. The next section contains the hardness results, including the proofs of Theorem 1.1 and Theorem 1.2. Section 3 contains the proof of Theorem 1.3 and the proof of Theorem 1.4. At the end of the proof of each of the above theorems we explain how the algorithm can be derived – this mostly consists of using the conditional expectation method to derandomize the probabilistic parts of the proofs. The final Section 4 contains some open problems and conjectures. Due to space limitations, several proofs have been omitted from this write-up.

2. Hardness results

We first give an outline of our proof of Theorem 1.1. We begin by showing the computational equivalence of the problem of rainbow connectivity 2, that asks for a red-blue edge coloring in which *all* vertex pairs have a rainbow path connecting them, to the problem of *subset rainbow connectivity* 2, asking for a red-blue coloring in which every pair of vertices in a *given subset* of pairs has a rainbow path connecting them. This is proved in Lemma 2.1 below.

In the second step, we reduce the problem of *extending to rainbow connectivity* 2, asking whether a given partial red-blue coloring can be completed to a obtain a rainbow connected graph, to the subset rainbow connectivity 2 problem. This is proved in Lemma 2.2 below.

Finally, the proof of Theorem 1.1 is completed by reducing 3-SAT to the problem of *extending to rainbow connectivity* 2.

Lemma 2.1. *The following problems are polynomially equivalent:*

1. *Given a graph G decide whether $rc(G) = 2$.*
2. *Given a graph G and a set of pairs $P \subseteq V(G) \times V(G)$, decide whether there is an edge coloring of G with 2 colors such that all pairs $(u, v) \in P$ are rainbow connected.*

Lemma 2.2. *The first problem defined below is polynomially reducible to the second one:*

1. *Given a graph $G = (V, E)$ and a partial 2-edge-coloring $\hat{\chi} : \hat{E} \to \{0, 1\}$ for $\hat{E} \subset E$, decide whether $\hat{\chi}$ can be extended to a complete 2 edge-coloring $\chi : E \to \{0, 1\}$ that makes G rainbow connected.*
2. *Given a graph G and a set of pairs $P \subseteq V(G) \times V(G)$ decide whether there is an edge coloring of G with 2 colors such that all pairs $(u, v) \in P$ are rainbow connected.*

We are unable to present the proofs of Lemma 2.1 and Lemma 2.2 due to space limitations.

Proof of Theorem 1.1 We show that Problem 1 of Lemma 2.2 is NP-hard, and then deduce that 2-rainbow-colorability is NP-Complete by applying Lemma 2.1 and Lemma 2.2 while observing that it clearly belongs to NP.

We reduce 3-SAT to Problem 1 of Lemma 2.2. Given a 3CNF formula $\phi = \bigwedge_{i=1}^{m} c_i$ over variables $x_1, x_2, \dots, x_n$, we construct a graph G_ϕ and a partial 2-edge coloring $\chi' : E(G_\phi) \to \{0,1\}$ such that there is an extension χ of χ' that makes G_ϕ rainbow connected if and only if ϕ is satisfiable.

We define G_ϕ as follows:

$$V(G_\phi) = \{c_i : i \in [m]\} \cup \{x_i : i \in [n]\} \cup \{a\}$$

$$E(G_\phi) = \Big\{\{c_i, x_j\} : x_j \in c_i \text{ in } \phi\Big\} \cup \Big\{\{x_i, a\} : i \in [n]\Big\} \cup \Big\{\{c_i, c_j\} : i, j \in [m]\Big\} \cup \Big\{\{x_i, x_j\} : i, j \in [n]\Big\}$$

and we define the partial coloring χ' as follows:

$$\forall_{i,j\in[m]} \chi'(\{c_i, c_j\}) = 0$$

$$\forall_{i,j\in[n]} \chi'(\{x_i, x_j\}) = 0$$

$$\forall_{\{x_i,c_j\}\in E(G_\phi)} \chi'(\{x_i, c_j\}) = 0 \text{ if } x_i \text{ is positive in } c_j,\ 1 \text{ otherwise}$$

while all the edges in $\Big\{\{x_i, a\} : i \in [n]\Big\}$ (and only they) are left uncolored.

Assuming without loss of generality that all variables in ϕ appear both as positive and as negative, one can verify that a 2-rainbow-coloring of the uncolored edges corresponds to a satisfying assignment of ϕ and vice versa. ∎

The proof of Theorem 1.2 is based upon the proof of the following theorem.

Theorem 2.3. *The following problem is NP-complete: Given an edge colored graph G and two vertices s, t of G, decide whether there is a rainbow path connecting s and t.*

Proof. Clearly the problem is in NP. We prove that it is NP-Complete by reducing 3-SAT to it. Given a 3CNF formula $\phi = \bigwedge_{i=1}^{m} c_i$ over variables $x_1, x_2, \dots, x_n$, we construct a graph G_ϕ with two special vertices s, t and a coloring $\chi : E(G_\phi) \to [|E(G_\phi)|]$ such that there is a rainbow path connecting s and t in G_ϕ if and only if ϕ is satisfiable.

We start by constructing an auxiliary graph G' from ϕ. The graph G' has $3m + 2$ vertices, that are partitioned into $m + 2$ layers $V_0, V_1, \dots, V_m, V_{m+1}$, where $V_0 = \{s\}$, $V_{m+1} = \{t\}$ and for each $i \in [m]$, the layer V_i contains the three vertices corresponding to the literals of c_i (a clause in ϕ). The edges of G' connect between all pairs of vertices residing in consecutive layers. Formally,

$$E(G') = \Big\{\{u, v\} : \exists i \in [m+1] \text{ s.t. } u \in V_{i-1} \text{ and } v \in V_i\Big\}.$$

Intuitively, in our final colored graph G_ϕ, every rainbow path from s to t will define a satisfying assignment of ϕ in a way that for every $i \in [m]$, if the rainbow path contains a vertex $v \in V_i$ then the literal of c_i that corresponds to v is satisfied, and hence c_i is satisfied. Since any path from s to t must contain at least one vertex from every layer V_i, this will yield a satisfying assignment for the whole formula ϕ. But we need to make sure that there are no contradictions in this assignment, that is, no opposite literals are satisfied together. For this we modify G' by replacing each literal-vertex with a gadget, and we define an edge coloring for which rainbow paths yield only consistent assignments.

For every variable x_j, $j \in [n]$, let $v_{j_1}, v_{j_2}, \dots, v_{j_k}$ be the vertices of G' corresponding to the positive literal x_j, and let $\overline{v}_{j_1}, \overline{v}_{j_2}, \dots, \overline{v}_{j_\ell}$ be the vertices corresponding to the negative literal $\overline{x}_j$. We can assume without loss of generality that both $k \geq 1$ and $\ell \geq 1$, since otherwise the formula ϕ can be simplified. For every such variable x_j we also introduce $k \times \ell$ distinct colors $\alpha^j_{1,1}, \dots, \alpha^j_{k,\ell}$. Next, we transform the auxiliary graph G' into the final graph G_ϕ.

For every $a \in [k]$ we replace the vertex v_{j_a} that resides in layer (say) V_i with $\ell + 1$ new vertices $v_1, v_2, \dots, v_{\ell+1}$ that form a path in that order. We also connect all vertices in V_{i-1} to v_1 and connect all vertices in V_{i+1} to $v_{\ell+1}$. For every $b \in [\ell]$, we color the edge $\{v_b, v_{b+1}\}$ in the new path with the color $\alpha^j_{a,b}$. Similarly, for every $b \in [\ell]$ we replace the vertex $\overline{v}_{j_b}$ from layer (say) $V_{i'}$ with $k + 1$ new vertices $\overline{v}_1, \overline{v}_2, \dots, \overline{v}_{k+1}$ that form a path, and connect all vertices in $V_{i'-1}$ to $\overline{v}_1$ and all vertices in $V_{i'+1}$ to $\overline{v}_{k+1}$. For every $a \in [k]$, we color the edge $\{v_a, v_{a+1}\}$ with $\alpha^j_{a,b}$. All other edges of G_ϕ (which were the original edges of G') are colored with fresh distinct colors.

Clearly, any path from s to t in G_ϕ must contain at least one of the newly built paths in each layer. On the other hand, it is not hard to verify that any two paths of opposite literals of the same variable have edges sharing the same color. ∎

Proof of Theorem 1.2. We reduce from the problem in Theorem 2.3. Given an edge colored graph $G = (V, E)$ with two special vertices s and t, we construct a graph $G' = (V', E')$ and define a coloring $\chi' : E' \to [|E'|]$ of its edges such that s and t are rainbow connected in G if and only if the coloring of G' makes G' rainbow connected.

Let $V = \{v_1 = s, v_2, \dots, v_n = t\}$ be the vertices of the original graph G. We set

$$V' = V \cup \{s', t', b\} \cup \{s^1, v_2^1, v_2^2, \dots, v_{n-1}^1, v_{n-1}^2, t^2\}$$

and

$$E' = E \cup \Big\{\{s', s\}, \{t', t\}, \{s, s^1\}, \{t, t^2\}\Big\} \cup \Big\{\{b, v_i\} : i \in [n]\Big\} \cup$$
$$\cup \Big\{\{v_i, v_i^j\} : i \in [n],\ j \in \{1, 2\}\Big\} \cup \Big\{\{v_i^a, v_j^b\} : i, j \in [n],\ a, b \in \{1, 2\}\Big\}.$$

The coloring χ' is defined as follows:

- all edges $e \in E$ retain the original color, that is $\chi'(e) = \chi(e)$;
- the edges $\{t, t'\}, \{s, b\}$ and $\Big\{\{v_i, v_i^1\} : i \in [n-1]\Big\}$ are colored with a special color c_1;
- the edges $\{s, s'\}, \{t, b\}$ and $\Big\{\{v_i, v_i^2\} : i \in [2, n]\Big\}$ are colored with a special color c_2;
- the edges in $\Big\{\{v_i, b\} : i \in [2, n-1]\Big\}$ are colored with a special color c_3;
- the edges in $\Big\{\{v_i^a, v_j^b\} : i, j \in [n],\ a, b \in \{1, 2\}\Big\}$ are colored with a special color c_4.

One can verify that χ' makes G' rainbow connected if and only if there was a rainbow path from s to t in G. ∎

3. Upper bounds and algorithms

The proof of our main Theorem 1.3 is based upon a modified degree-form version of Szemerédi's Regularity Lemma, that we prove here and that may be useful in other applications. We begin by introducing the Regularity Lemma and the already known degree-form version of it.

3.1. Regularity Lemma

The Regularity Lemma of Szemerédi [9] is one of the most important results in graph theory and combinatorics, as it guarantees that every graph has an ϵ-approximation of constant descriptive size, namely a size that depends only on ϵ and not on the size of the graph. This approximation "breaks" the graph into a constant number of pseudo-random bipartite graphs. This is very useful in many applications since dealing with random-like graphs is much easier than dealing with arbitrary graphs. In particular, as we shall see, the Regularity Lemma allows us to prove that graphs with linear minimum degree have bounded rainbow connectivity.

We first state the lemma. For two nonempty disjoint vertex sets A and B of a graph G, we define $E(A,B)$ to be the set of edges of G between A and B. The *edge density* of the pair is defined by $d(A,B) = |E(A,B)|/(|A||B|)$.

Definition 3.1 (ϵ-regular pair)**.** A pair (A,B) is *ϵ-regular* if for every $A' \subseteq A$ and $B' \subseteq B$ satisfying $|A'| \geq \epsilon|A|$ and $|B'| \geq \epsilon|B|$, we have $|d(A',B') - d(A,B)| \leq \epsilon$.

An ϵ-regular pair can be thought of as a pseudo-random bipartite graph in the sense that it behaves almost as we would expect from a random bipartite graph of the same density. Intuitively, in a random bipartite graph with edge density d, all large enough sub-pairs should have similar densities.

A partition $V_1,\dots,V_k$ of the vertex set of a graph is called an *equipartition* if $|V_i|$ and $|V_j|$ differ by no more than 1 for all $1 \leq i < j \leq k$ (so in particular every V_i has one of two possible sizes). The *order* of an equipartition denotes the number of partition classes (k above). An equipartition $V_1,\dots,V_k$ of the vertex set of a graph is called *ϵ-regular* if all but at most $\epsilon\binom{k}{2}$ of the pairs (V_i,V_j) are ϵ-regular. Szemerédi's Regularity Lemma can be formulated as follows.

Lemma 3.2 (Regularity Lemma [9])**.** *For every $\epsilon > 0$ and positive integer K, there exists $N = N_{3.2}(\epsilon,K)$, such that any graph with $n \geq N$ vertices has an ϵ-regular equipartition of order k, where $K \leq k \leq N$.*

As mentioned earlier, the following variation of the lemma comes useful in our context.

Lemma 3.3 (Regularity Lemma - degree form [8])**.** *For every $\epsilon > 0$ and positive integer K there is $N = N_{3.3}(\epsilon,K)$ such that given any graph $G = (V,E)$ with $n > N$ vertices, there is a partition of the vertex-set V into $k+1$ sets $V'_0, V'_1, \dots, V'_k$, and there is a subgraph G' of G with the following properties:*

(1) $K \leq k \leq N$,
(2) $s \triangleq |V'_0| \leq \epsilon^5 n$ *and all other components* V'_i, $i \in [k]$ *are of size* $\ell \triangleq \frac{n-s}{k}$,
(3) *for all* $i \in [k]$, V'_i *induces an independent set in* G',
(4) *for all* $i,j \in [k]$, *the pair* (V'_i, V'_j) *is* ϵ^5*-regular in* G', *with density either* 0 *or at least* $\frac{\epsilon}{4}$,

(5) *for all* $v \in V$, $\deg_{G'}(v) > \deg_G(v) - \frac{\epsilon}{3}n$.

This form of the lemma (see e.g. [8]) can be obtained by applying the original Regularity Lemma (with a smaller value of ϵ), and then "cleaning" the resulting partition. Namely, adding to the exceptional set V_0' all components V_i incident to many irregular pairs, deleting all edges between any other pairs of clusters that either do not form an ϵ-regular pair or they do but with density less than ϵ, and finally adding to V_0 also vertices whose degree decreased too much by this deletion of edges.

3.2. A modified degree form version of the Regularity Lemma

In order to prove that graphs with linear minimum degree have bounded rainbow connectivity number, we need a special version of the Regularity Lemma, which is stated next.

Lemma 3.4 (Regularity Lemma - new version)**.** *For every* $\epsilon > 0$ *and positive integer* K *there is* $N = N_{3.4}(\epsilon, K)$ *so that the following holds: If* $G = (V, E)$ *is a graph with* $n > N$ *vertices and minimum degree at least* ϵn *then there is a subgraph* G'' *of* G*, and a partition of* V *into* $V_1'', \dots, V_k''$ *with the following properties:*

(1) $K \leq k \leq N$,
(2) *for all* $i \in [k]$, $(1-\epsilon)\frac{n}{k} \leq |V_i''| \leq (1+\epsilon^3)\frac{n}{k}$,
(3) *for all* $i \in [k]$, V_i'' *induces an independent set in* G'',
(4) *for all* $i, j \in [k]$, (V_i'', V_j'') *is an* ϵ^3*-regular pair in* G''*, with density either* 0 *or at least* $\frac{\epsilon}{16}$,
(5) *for all* $i \in [k]$ *and every* $v \in V_i''$ *there is at least one other class* V_j'' *so that the number of neighbors of* v *in* G'' *belonging to* V_j'' *is at least* $\frac{\epsilon}{2}|V_j''|$.

We also note that the above a partition as guaranteed by our modified version of the Regularity Lemma can be found in polynomial time for a fixed ϵ (with somewhat worse constants), by using the exact same methods that were used in [1] for constructing an algorithmic version of the original Regularity Lemma. We are unable to give the complete proof of Lemma 3.4 due to space limitations.

3.3. Proof of Theorem 1.3

In this section we use our version of the Regularity Lemma to prove Theorem 1.3. First we need some definitions. Given a graph $G = (V, E)$ and two subsets $V_1, V_2 \subseteq V$, let $E(V_1, V_2)$ denote the set of edges having one endpoint in V_1 and another endpoint in V_2. Given a vertex v, let $\Gamma(v)$ denote the set of v's neighbors, and for $W \subseteq V$, let $\Gamma_W(v)$ denote the set $W \cap \Gamma(v)$.

For an edge coloring $\chi : E \to \mathcal{C}$, let π_χ denote the corresponding partition of E into (at most) $|\mathcal{C}|$ components. For two edge colorings χ and χ', we say that χ' is a *refinement* of χ if $\pi_{\chi'}$ is a refinement of π_χ, which is equivalent to saying that $\chi'(e_1) = \chi'(e_2)$ always implies $\chi(e_1) = \chi(e_2)$.

Observation 3.5. Let χ and χ' be two edge-colorings of a graph G, such that χ' is a refinement of χ. For any path P in G, if P is a rainbow path under χ, then P is a rainbow path under χ'. In particular, if χ makes G rainbow connected, then so does χ'. ∎

We define a set of eight distinct colors $\mathcal{C} = \{a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4\}$. Given a coloring $\chi : E \to \mathcal{C}$ we say that $u, v \in V$ are *a-rainbow connected* if there is a rainbow path from u to v using only the colors a_1, a_2, a_3, a_4. We similarly define *b-rainbow connected* pairs. The following is a central lemma in the proof of Theorem 1.3. The proof of Lemma 3.6 is given in Section 3.4.

Lemma 3.6. *For any $\epsilon > 0$, there is $N = N_{3.6}(\epsilon)$ such that any connected graph $G = (V, E)$ with $n > N$ vertices and minimum degree at least ϵn satisfies the following. There is a partition Π of V into $k \le N$ components $V_1, V_2, \ldots, V_k$, and a coloring $\chi : E \to \mathcal{C}$ such that for every $i \in [k]$ and every $u, v \in V_i$, the pair u, v is both a-rainbow connected and b-rainbow connected under χ.*

Using Lemma 3.6 we derive the proof of Theorem 1.3. For a given $\epsilon > 0$, set $N = N_{3.6}(\epsilon)$ and set $C = \frac{3}{\epsilon}N + 8$. Clearly, any connected graph $G = (V, E)$ with $n \le C$ vertices satisfies $rc(G) \le C$. So we assume that $n > C \ge N$, and let $\Pi = V_1, \ldots, V_k$ be the partition of V from Lemma 3.6, while we know that $k \le N$.

First observe that since the minimal degree of G is ϵn, the diameter of G is bounded by $3/\epsilon$. This can be verified by e.g. by taking an arbitrary vertex $r \in V$ and executing a BFS algorithm from it. Let $L_1, \ldots, L_t$ be the layers of vertices in this execution, where L_i are all vertices at distance i from r. Observe that since the minimal degree is at least ϵn, the total number of vertices in every three consecutive layers must be at least ϵn, thus $t \le 3/\epsilon$. Since the same claim holds for any $r \in V$, this implies that $\mathrm{diam}(G) \le t \le 3/\epsilon$.

Now let $T = (V_T, E_T)$ be a connected subtree of G on at most $k \cdot \mathrm{diam}(G) \le \frac{3}{\epsilon}N$ vertices such that for every $i \in [k]$, $V_T \cap V_i \ne \emptyset$. Such a subtree must exist in G since as observed earlier, $\mathrm{diam}(G) \le 3/\epsilon$. Let $\chi : E \to \mathcal{C}$ be the coloring from Lemma 3.6, and let $\mathcal{H} = \{h_1, h_2, \ldots, h_{|E_T|}\}$ be a set of $|E_T| \le \frac{3}{\epsilon}N$ fresh colors. We refine χ by recoloring every $e_i \in E(T)$ with color $h_i \in \mathcal{H}$. Let $\chi' : E \to \left(\mathcal{C} \cup \mathcal{H}\right)$ be the resulting coloring of G. The following lemma completes the proof of Theorem 1.3.

Lemma 3.7. *The coloring χ' makes G rainbow connected. Consequently, $rc(G) \le |E_T| + 8 \le C$.*

Proof. Let $u, v \in V$ be any pair of G's vertices. If u and v reside in the same component V_i of the partition Π, then (by Lemma 3.6) they are connected by a path P of length at most four, which is a rainbow path under the the original coloring χ. Since χ' is a refinement of χ, the path P remains a rainbow path under χ' as well (see Observation 3.5).

Otherwise, let $u \in V_i$ and $v \in V_j$ for $i \ne j$. Let t_i and t_j be vertices of the subtree T, residing in V_i and V_j respectively. By definition of χ', there is a rainbow path from t_i to t_j using colors from $\mathcal{H}$. Let P_t denote this path. In addition, by Lemma 3.6 we know that for the original coloring χ, there is a rainbow path P_a from u to t_i using colors $a_1, \ldots, a_4$ and there is a rainbow path P_b from v to t_j using colors $b_1, \ldots, b_4$. Based on the fact that χ' is a refinement of χ, it is now easy to verify that P_t, P_a and P_b can be combined to form a rainbow path from u to v under χ'. ∎

This concludes the proof of Theorem 1.3, apart from the existence of a polynomial time algorithm for finding this coloring. We note that all arguments above apart from Lemma 3.6 admit polynomial algorithms for finding the corresponding structures. The algorithm for Lemma 3.6 will be given with its proof.

3.4. Proof of Lemma 3.6

First we state another auxiliary lemma, which is proved in the next section.

Lemma 3.8. *For every $\epsilon > 0$ there exists $N = N_{3.8}(\epsilon)$ such that any graph $G = (V, E)$ with $n > N$ vertices and minimum degree at least ϵn satisfies the following: There exists a partition $\Pi = V_1, \ldots, V_k$ of V such that for every $i \in [k]$ and every $u, v \in V_i$, the number of edge disjoint paths of length at most four from u to v is larger than $8^5 \log n$. Moreover, these sets can be found using a polynomial time algorithm for a fixed ϵ.*

Proof. **(of Lemma 3.6)** First we apply Lemma 3.8 to get the partition Π. Now the proof follows by a simple probabilistic argument. Namely, we color every edge $e \in E$ by choosing one of the colors in $\mathcal{C} = \{a_1, \ldots, a_4, b_1, \ldots, b_4\}$ uniformly and independently at random. Observe that a fixed path P of length at most four is an a-rainbow path with probability at least 8^{-4}. Similarly, P is a b-rainbow path with probability at least 8^{-4}. So any fixed pair $u, v \in V_i$ is not both a-rainbow-connected and b-rainbow-connected with probability at most $2(1-8^{-4})^{8^5 \log n} < n^{-2}$, and therefore the probability that all such pairs are both a-rainbow connected and b-rainbow connected is strictly positive. Hence the desired coloring must exist.

To find the coloring algorithmically, we note that for every *partial* coloring of the edges of the graph it is easy to calculate the *conditional* probability that the fixed pair of vertices u, v is not both a-rainbow-connected and b-rainbow-connected. Therefore we can calculate the conditional expectation of the number of pairs that are not so connected for any partial coloring. Now we can derandomize the random selection of the coloring above by using the conditional expectation method (cf. [2]): In every stage we color one of the remaining edges in a way that does not increase the conditional expectation of the number of unconnected pairs. Since this expectation is smaller than 1 in the beginning, in the end we will have less than 1 unconnected pair, and so all pairs will be connected. ∎

3.5. Proof of Lemma 3.8

Given $\epsilon > 0$ let $L = N_{3.4}(\epsilon, 1)$ and set N to be the smallest number that satisfies $\epsilon^4 \frac{N}{L} > 8^5 \log N$. Now, given any graph $G = (V, E)$ with $n > N$ vertices and minimum degree at least ϵn, we apply Lemma 3.4 with parameters ϵ and 1. Let $\Pi = V_1, V_2, \ldots, V_k$ be the partition of V obtained from Lemma 3.4, while as promised, $k \le L = N_{3.4}(\epsilon)$.

Fix $i \in [k]$ and $u, v \in V_i$. From Lemma 3.4 we know that there is a component V_a such that u has at least $\frac{\epsilon}{3k} n$ neighbors in V_a. Similarly, there is a component V_b such that v has at least $\frac{\epsilon}{3k} n$ neighbors in V_b. Let $\Gamma_{u,a}$ denote the set of u's neighbors in V_a, and similarly, let $\Gamma_{v,b}$ denote v's neighbors in V_b. We assume in this proof that $V_a \neq V_b$, and at the end it will be clear that the case $V_a = V_b$ can only benefit.

We say that a set $W_u = \{w_1, \ldots, w_t\} \subseteq V_i$ is *distinctly reachable from* u if there are distinct vertices $w'_1, \ldots, w'_t \in \Gamma_{u,a}$ such that for every $j \in [t]$, $\{w_j, w'_j\} \in E$. Notice that the collection of pairs $\{w_j, w'_j\}$ corresponds to a matching in the graph G, where all edges of the matching have one endpoint in V_i and the other endpoint in $\Gamma_{u,a}$. Similarly, we say that $W_v \subseteq V_i$ is distinctly reachable from v if there are distinct vertices $w'_1, \ldots, w'_t \in \Gamma_{v,b}$ such that for every $j \in [t]$, $\{w_j, w'_j\} \in E$.

Observe that it is enough to prove that there exists a set $W \subseteq V_i$ of size $\epsilon^4 \frac{N}{L} > 8^5 \log N$ which is distinctly reachable from both u and v. This will imply the existence of $8^5 \log N$ edge disjoint paths of length four from u to v.

Our first goal is to bound from below the size of the maximal set W_u as above. Since (by Lemma 3.4) V_a and V_i are ϵ^3-regular pairs with density $\geq \frac{\epsilon}{16}$ and since $\epsilon^3 < \epsilon/3$, the number of edges between $\Gamma_{u,a}$ and V_i is at least $\left(\frac{\epsilon}{16} - \epsilon^3\right) |\Gamma_{u,a}| \cdot |V_i|$. Before proceeding, we make the following useful observation.

Observation 3.9. Let $H = (A, B)$ be a bipartite graph with $\gamma|A||B|$ edges. Then H contains a matching M of size $\gamma \frac{|A||B|}{|A|+|B|}$.

Proof. Consider the following process that creates M. Initially $M_0 = \emptyset$. Then in step i, we pick an arbitrary edge $\{a, b\} \in E(H)$, set $M_{i+1} = M_i \cup \{a, b\}$ and remove from $E(H)$ all the edges incident with either a or b. Clearly, in each step the number of removed edges is bounded by $|A| + |B|$, so the process continues for at least $\frac{E(H)}{|A|+|B|} = \gamma \frac{|A||B|}{|A|+|B|}$ steps. Hence $|M| = |\bigcup_i M_i| \geq \gamma \frac{|A||B|}{|A|+|B|}$. ■

Returning to the proof of Lemma 3.8, by Observation 3.9 the size of a maximal set W_u as above is at least

$$\left(\frac{\epsilon}{16} - \epsilon^3\right) \frac{|\Gamma_{u,a}||V_i|}{|\Gamma_{u,a}| + |V_i|} \geq \left(\frac{\epsilon}{16} - \epsilon^3\right) \frac{\Big(\epsilon n/(3k)\Big)(n/k)}{\epsilon n/(3k) + n/k} \geq \frac{\epsilon^2}{64k} n.$$

To prove that $W = W_u \cap W_a$ is large, we similarly use the regularity condition, but now on the pair $(\Gamma_{v,b}, W_u)$. We get,

$$|E(\Gamma_{v,b}, W_u)| \geq \left(\frac{\epsilon}{16} - \epsilon^3\right) |\Gamma_{v,b}||W_u|.$$

Here too, by Observation 3.9 we can bound from below the size of a maximal matching in the pair $(\Gamma_{v,b}, W_u)$ with

$$\left(\frac{\epsilon}{16} - \epsilon^3\right) \frac{|\Gamma_{v,b}||W_u|}{|\Gamma_{v,b}| + |W_u|} \geq \left(\frac{\epsilon}{16} - \epsilon^3\right) \frac{\left(\frac{\epsilon}{3k} n\right)\left(\frac{\epsilon^2}{64k} n\right)}{\frac{\epsilon}{3k} n + \frac{\epsilon^2}{64k} n} \geq \epsilon^4 \frac{n}{k} \geq \epsilon^4 \frac{n}{L} > 8^5 \log N,$$

where the last inequality follows from our choice of N. Recall that the matching that we found defines the desired set W, concluding the proof. An algorithmic version of this lemma can be derived by simply using an algorithmic version of Lemma 3.4 in the selection of $V_1, \ldots, V_k$ above. ■

3.6. Graphs with diameter 2

Proof of Theorem 1.4. Consider a random 3-coloring of E, where every edge is colored with one of three possible colors uniformly and independently at random. It is enough to prove that for all pairs $u, v \in V$ the probability that they are not rainbow connected is at most $1/n^2$. Then the proof follows by the union bound (cf. [2]).

Let us fix a pair $u, v \in V$, and bound from above the probability that this pair is not rainbow connected. We know that both $\Gamma(u)$ and $\Gamma(v)$ (the neighborhoods of u and v) contain at least $8 \log n$ vertices.

(1) If $\{u, v\} \in E$ then we are done.
(2) If $|\Gamma(u) \cap \Gamma(v)| \geq 2\log n$ then there are at least $2\log n$ edge-disjoint paths of length two from u to v. In this case, the probability that none of these paths is a rainbow path is bounded by $(1/3)^{2\log n} < 1/n^2$, and we are done.
(3) Otherwise, let $A = \Gamma(u) \setminus \Gamma(v)$ and $B = \Gamma(v) \setminus \Gamma(u)$. We know that $|A|, |B| \geq 6\log n$, and in addition, since the first two cases do not hold and the diameter of G is two, all the (length two) shortest paths from A's vertices to v go through the vertices in B. This implies that every vertex $x \in A$ has a neighbor $b(x) \in B$ ($b(x)$ need not be a one-one function). Let us consider the set of at least $6\log n$ edge-disjoint paths $P = \{u, x, b(x) : x \in A\}$. For each $x \in A$, the probability that $u, x, b(x), v$ is a rainbow path (given the color of the edge $(b(x), v)$) is $2/9$. Moreover, this event is independent of the corresponding events for all other members of A, because this proabablity does not change even with full knowledge of the colors of all edges incident with v. Therefore, the probability that none of the paths in P extends to a rainbow path from u to v is at most $(7/9)^{6\log n} \leq 1/n^2$, as required.

The above proof immediately implies a probabilistic polynomial expected time randomized algorithm with zero error probability (since we can also efficiently check if the coloring indeed makes G 3-rainbow connected). The algorithm can be derandomized and converted to a polynomial time probabilistic algorithm using the method of conditional expectations (cf. [2]) similarly to the proof of Lemma 3.6: For every partial coloring of the edges we can efficiently bound the conditional probability that a fixed pair u, v is not rainbow-connected, using the relevant one of the three cases concerning u and v that were analyzed above. Now we can color the edges one by one, at each time taking care not to increase the bound on the conditional expectation of unconnected pairs that results from the above probability bound for every u and v. Since the bound on the expectation was smaller than 1 before the beginning of the process, in the end we would get a valid 3-rainbow-coloring of G. ■

4. Concluding remarks and open problems

- Theorem 1.3 asserts that a connected graph with minimum degree at least ϵn has bounded rainbow connectivity. However, the bound obtained is huge as it follows from the Regularity Lemma. It would be interesting to find the "correct" bound. It is even possible that $rc(G) \leq C/\epsilon$ for some absolute constant C.
- The proof of Theorem 1.1 shows that deciding whether $rc(G) = 2$ is NP-Complete. Although this suffices to deduce that computing $rc(G)$ is NP-Hard, we still do not have a proof that deciding whether $rc(G) \leq k$ is NP-Complete for every fixed k. It is tempting to conjecture that for every k it is NP-Hard even to distinguish between 2-rainbow-colorable graphs and graphs that are not even k-rainbow-colorable.
- A parameter related to rainbow connectivity is the *rainbow diameter*. In this case we ask for an edge coloring so that for any two vertices, there is a rainbow *shortest* path connecting them. The rainbow diameter number, denoted $rd(G)$ is the smallest number of colors used in such a coloring. Clearly, $rd(G) \geq rc(G)$ and obviously every connected graph with n vertices has $rd(G) < \binom{n}{2}$. Unlike rainbow connectivity, which is a monotone graph

property (adding edges never increases the rainbow connectivity number) this is not the case for the rainbow diameter (although we note that constructing an example that proves non-monotonicity is not straightforward). Clearly, computing $rd(G)$ is NP-Hard since $rc(G) = 2$ if and only if $rd(G) = 2$. It would be interesting to prove a version of Theorem 1.3 for rainbow diameter. We conjecture that, indeed, if G is a connected graph with minimum degree at least ϵn then it has a bounded rainbow diameter.

- Suppose that we are given a graph G for which we are *told* that $rc(G) = 2$. Can we rainbow-color it in polynomial time with $o(n)$ colors? For the usual coloring problem, this version has been well studied. It is known that if a graph is 3-colorable (in the usual sense), then there is a polynomial time algorithm that colors it with $\tilde{O}(n^{3/14})$ colors [3].

Acknowledgment We thank Zsolt Tuza for pointing out an error in one of our concluding remarks.

References

[1] N. Alon, R.A. Duke, H. Lefmann, V. Rödl, and R. Yuster, *The algorithmic aspects of the Regularity Lemma*, Journal of Algorithms 16 (1994), 80–109.

[2] N. Alon and J. H. Spencer, **The Probabilistic Method**, *Second Edition*, Wiley, New York, 2000.

[3] A. Blum and D. Karger, *An $\tilde{O}(n^{3/14})$-coloring algorithm for 3-colorable graphs*, Inform. Process. Lett., 61(1) : 49-53, 1997.

[4] B. Bollobás, **Modern Graph Theory**, Graduate Texts in Mathematics 184, Springer-Verlag 1998.

[5] G. Chartrand, G. L. Johns, K. A. McKeon, and P. Zhang, *Rainbow connection in graphs*, Math. Bohem., 133 (2008), no. 1, 85-98.

[6] Y. Caro, A. Lev, Y. Roditty, Z. Tuza, and R. Yuster, *On rainbow connectivity*, The Electronic Journal of Combinatorics 15(2008), Paper R57.

[7] E. Fischer, A. Matsliah, and A. Shapira, *Approximate Hypergraph Partitioning and Applications*, Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS) 579–589 (2007).

[8] J. Komlós and M. Simonovits, *Szemerédis Regularity Lemma and its applications in graph theory*, In: Combinatorics, Paul Erdös is Eighty (D. Miklós, V. T. Sós, and T. Szönyi, Eds.), Bolyai Society Mathematical Studies, Vol. 2, Budapest (1996), 295-352.

[9] E. Szemerédi Regular partitions of graphs, *Proc. Colloque Inter. CNRS 260*, (CNRS, Paris) 399–401 (1978).