The RED-BLUE SEPARATION Problem on Graphs

Subhadeep Ranjan Dev[1], Sanjana Dey[1], Florent Foucaud[2,3](✉), Ralf Klasing[4], and Tuomo Lehtilä[5,6]

[1] ACM Unit, Indian Statistical Institute, Kolkata, India
[2] Université Clermont-Auvergne, CNRS, Mines de Saint-Étienne, Clermont-Auvergne-INP, LIMOS, 63000 Clermont-Ferrand, France
florent.foucaud@uca.fr
[3] Univ. Orléans, INSA Centre Val de Loire, LIFO EA 4022, 45067 Orléans Cedex 2, France
[4] Université de Bordeaux, Bordeaux INP, CNRS, LaBRI, UMR 5800, Talence, France
[5] Univ Lyon, Université Claude Bernard, CNRS, LIRIS - UMR 5205, 69622 Lyon, France
[6] Department of Mathematics and Statistics, University of Turku, Turku, Finland

Abstract. We introduce the RED-BLUE SEPARATION problem on graphs, where we are given a graph $G = (V, E)$ whose vertices are colored either red or blue, and we want to select a (small) subset $\mathcal{S} \subseteq V$, called *red-blue separating set*, such that for every red-blue pair of vertices, there is a vertex $s \in \mathcal{S}$ whose closed neighborhood contains exactly one of the two vertices of the pair. We study the computational complexity of RED-BLUE SEPARATION, in which one asks whether a given red-blue colored graph has a red-blue separating set of size at most a given integer. We prove that the problem is NP-complete even for restricted graph classes. We also show that it is always approximable in polynomial time within a factor of $2 \ln n$, where n is the input graph's order. In contrast, for triangle-free graphs and for graphs of bounded maximum degree, we show that RED-BLUE SEPARATION is solvable in polynomial time when the size of the smaller color class is bounded by a constant. However, on general graphs, we show that the problem is $W[2]$-hard even when parameterized by the solution size plus the size of the smaller color class. We also consider the problem MAX RED-BLUE SEPARATION where the coloring is not part of the input. Here, given an input graph G, we want to determine the smallest integer k such that, *for every possible red-blue-coloring* of G, there is a red-blue separating set of size at most k. We derive tight bounds on the cardinality of an optimal solution of

This study has been carried out in the frame of the "Investments for the future" Programme IdEx Bordeaux - SysNum (ANR-10-IDEX-03-02).

F. Foucaud—Research financed by the French government IDEX-ISITE initiative 16-IDEX-0001 (CAP 20-25) and by the ANR project GRALMECO (ANR-21-CE48-0004-01).

T. Lehtilä—Research supported by the Finnish Cultural Foundation and by the Academy of Finland grant 338797.

© Springer Nature Switzerland AG 2022
C. Bazgan and H. Fernau (Eds.): IWOCA 2022, LNCS 13270, pp. 285–298, 2022.
https://doi.org/10.1007/978-3-031-06678-8_21

Max Red-Blue Separation, showing that it can range from logarithmic in the graph order, up to the order minus one. We also give bounds with respect to related parameters. For trees however we prove an upper bound of two-thirds the order. We then show that Max Red-Blue Separation is NP-hard, even for graphs of bounded maximum degree, but can be approximated in polynomial time within a factor of $O(\ln^2 n)$.

1 Introduction

We introduce and study the Red-Blue Separation problem for graphs. Separation problems for discrete structures have been studied extensively from various perspectives. In the 1960s, Rényi [24] introduced the Separation problem for set systems (a set system is a collection of sets over a set of vertices), which has been rediscovered by various authors in different contexts, see e.g. [2,6,17,23]. In this problem, one aims at selecting a solution subset $\mathcal{S}$ of sets from the input set system to separate every pair of vertices, in the sense that the subset of $\mathcal{S}$ corresponding to those sets to which each vertex belongs to, is unique. The graph version of this problem (where the sets of the input set system are the closed neighborhoods of a graph), called Identifying Code [18], is also extensively studied. These problems have numerous applications in areas such as monitoring and fault-detection in networks [26], biological testing [23], and machine learning [20]. The Red-Blue Separation problem which we study here is a red-blue colored version of Separation, where instead of all pairs we only need to separate red vertices from blue vertices.

In the general version of the Red-Blue Separation problem, one is given a set system $(V, \mathcal{S})$ consisting of a set $\mathcal{S}$ of subsets of a set V of vertices which are either blue or red; one wishes to separate every blue from every red vertex using a solution subset $\mathcal{C}$ of $\mathcal{S}$ (here a set of $\mathcal{C}$ separates two vertices if it contains exactly one of them). Motivated by machine learning applications, a geometric-based special case of Red-Blue Separation has been studied in the literature, where the vertices of V are points in the plane and the sets of $\mathcal{S}$ are half-planes [7]. The classic problem Set Cover over set systems generalizes both Geometric Set Cover problems and graph problem Dominating Set (similarly, the set system problem Separation generalizes both Geometric Discriminating Code and the graph problem Identifying Code). It thus seems natural to study the graph version of Red-Blue Separation.

Problem Definition. In the graph setting, we are given a graph G and a red-blue coloring $c : V(G) \rightarrow \{\text{red}, \text{blue}\}$ of its vertices, and we want to select a (small) subset S of vertices, called *red-blue separating set*, such that for every red-blue pair r, b of vertices, there is a vertex from S whose closed neighborhood contains exactly one of r and b. Equivalently, $N[r] \cap S \neq N[b] \cap S$, where $N[x]$ denotes the closed neighborhood of vertex x; the set $N[x] \cap S$ is called the *code* of x (with respect to S), and thus all codes of blue vertices are different from all codes of red vertices. The smallest size of a red-blue separating set of (G, c) is denoted by $\text{sep}_{\text{RB}}(G, c)$. Note that if a red and a blue vertex have the same closed neighborhood, they cannot be separated. Thus, for simplicity, we will consider only *twin-free* graphs, that is, graphs where no two vertices have the same closed neighborhood. Also, for a twin-free graph, the vertex set $V(G)$ is always a red-blue separating set as all the vertices have a unique subset of neighbors. We have the following associated computational problem.

RED-BLUE SEPARATION
Input: A red-blue colored twin-free graph (G, c) and an integer k.
Question: Do we have $\text{sep}_{\text{RB}}(G, c) \leq k$?

It is also interesting to study the problem when the red-blue coloring is not part of the input. For a given graph G, we thus define the parameter $\text{max-sep}_{\text{RB}}(G)$ which denotes the largest size, over each possible red-blue coloring c of G, of a smallest red-blue separating set of (G, c). The associated decision problem is stated as follows.

MAX RED-BLUE SEPARATION
Input: A twin-free graph G and an integer k.
Question: Do we have $\text{max-sep}_{\text{RB}}(G) \leq k$?

In Fig. 1, to note the difference between sep_{RB} and $\text{max-sep}_{\text{RB}}$, a path of 6 vertices P_6 is shown, where the vertices are colored red or blue.

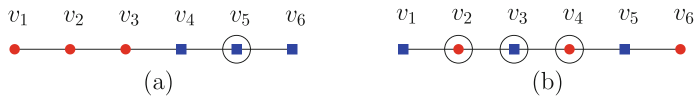

Fig. 1. A path of 6 vertices where (a) $\text{sep}_{\text{RB}}(P_6, c) = 1$ and (b) $\text{max-sep}_{\text{RB}}(P_6) = 3$; the members of the red-blue separating set are circled. Square vertices are blue, round vertices are red. (Color figure online)

Our Results. We show that RED-BLUE SEPARATION is NP-complete even for restricted graph classes such as planar bipartite sub-cubic graphs, in the setting where the two color classes[1] have equal size. We also show that the problem is NP-hard to approximate within a factor of $(1 - \epsilon) \ln n$ for every $\epsilon > 0$, even for split graphs[2] of order n, and when one color class has size 1. On the other hand, we show that RED-BLUE SEPARATION is always approximable in polynomial time within a factor of $2 \ln n$. In contrast, for triangle-free graphs and for graphs of bounded maximum degree, we prove that RED-BLUE SEPARATION is solvable in polynomial time when the smaller color class is bounded by a constant (using algorithms that are in the parameterized class XP, with the size of the smaller color class as parameter). However, on general graphs, the problem is shown to be $W[2]$-hard even when parameterized by the solution size plus the size of the smaller color class. (This is in contrast with the geometric version of separating points by half-planes, for which both parameterizations are known to be fixed-parameter tractable [3, 19]).

[1] One class consists of vertices colored *red* and the other class consists of vertices colored *blue*.

[2] A graph $G = (V, E)$ is called a *split graph* when the vertices in V can be partitioned into an independent set and a clique.

As the coloring is not specified, max-sep$_{\mathrm{RB}}(G)$ is a parameter that is worth studying from a structural viewpoint. In particular, we study the possible values for max-sep$_{\mathrm{RB}}(G)$. We show the existence of tight bounds on max-sep$_{\mathrm{RB}}(G)$ in terms of the order n of the graph G, proving that it can range from $\lfloor \log_2 n \rfloor$ up to $n-1$ (both bounds are tight). For trees however we prove bounds involving the number of support vertices (i.e. which have a leaf neighbor), which imply that max-sep$_{\mathrm{RB}}(G) \leq \frac{2n}{3}$. We also give bounds in terms of the (non-colored) separation number. We then show that the associated decision problem Max Red-Blue Separation is NP-hard, even for graphs of bounded maximum degree, but can be approximated in polynomial time within a factor of $O(\ln^2 n)$.

Related Work. Red-Blue Separation has been studied in the geometric setting of red and blue points in the Euclidean plane [3,5,22]. In this problem, one wishes to select a small set of (axis-parallel) lines such that any two red and blue points lie on the two sides of one of the solution lines. The motivation stems from the Discretization problem for two classes and two features in machine learning, where each point represents a data point whose coordinates correspond to the values of the two features, and each color is a data class. The problem is useful in a preprocessing step to transform the continuous features into discrete ones, with the aim of classifying the data points [7,19,20]. This problem was shown to be NP-hard [7] but 2-approximable [5] and fixed-parameter tractable when parameterized by the size of a smallest color class [3] and by the solution size [19]. A polynomial time algorithm for a special case was recently given in [22].

The Separation problem for set systems (also known as Test Cover and Discriminating Code) was introduced in the 1960s [24] and widely studied from a combinatorial point of view [1,2,6,17] as well as from the algorithmic perspective for the settings of classical, approximation and parameterized algorithms [8,10,23]. The associated graph problem is called Identifying Code [18] and is also extensively studied (see [21] for an online bibliography with almost 500 references as of January 2022); geometric versions of Separation have been studied as well [9,15,16]. The Separation problem is also closely related to the VC Dimension problem [27] which is very important in the context of machine learning. In VC Dimension, for a given set system $(V, \mathcal{S})$, one is looking for a (large) set X of vertices that is *shattered*, that is, for every possible subset of X, there is a set of $\mathcal{S}$ whose trace on X is the subset. This can be seen as "perfectly separating" a subset of $\mathcal{S}$ using X; see [4] for more details on this connection.

Structure of the Paper. We start with the algorithmic results on Red-Blue Separation in Sect. 2. We then present the bounds on max-sep$_{\mathrm{RB}}$ in Sect. 3 and the hardness result for Max Red-Blue Separation in Sect. 4. Due to space constraints, we have omitted some proofs or parts of proofs.

2 Complexity and Algorithms for Red-Blue Separation

We will prove some algorithmic results for Red-Blue Separation by reducing to or from the following problems.

SET COVER
Input: A set of elements U, a family $\mathcal{S}$ of subsets of U and an integer k.
Question: Does their exist a cover $\mathcal{C} \subseteq \mathcal{S}$, with $|\mathcal{C}| \leq k$ such that $\bigcup_{C \in \mathcal{C}} C = U$?

DOMINATING SET
Input: A graph $G = (V, E)$ and an integer k.
Question: Does there exist a set $D \subseteq V$ of size k with $\forall v \in V, N[v] \cap D \neq \emptyset$?

2.1 Hardness

Theorem 1. RED-BLUE SEPARATION *cannot be approximated within a factor of* $(1 - \epsilon) \cdot \ln n$ *for any* $\epsilon > 0$ *even when the smallest color class has size* 1 *and the input is a split graph of order* n*, unless P = NP. Moreover,* RED-BLUE SEPARATION *is W[2]-hard when parameterized by the solution size together with the size of the smallest color class, even on split graphs.*

Proof. For an instance $((U, \mathcal{S}), k)$ of SET-COVER, we construct in polynomial time an instance $((G, c), k)$ of RED-BLUE SEPARATION where G is a split graph and one color class has size 1. The statement will follow from the hardness of approximating MIN SET COVER proved in [11], and from the fact that SET COVER is W[2]-hard when parameterised by the solution size [12].

We create the graph (G, c) by first creating vertices corresponding to all the sets and the elements. We connect a vertex u_i corresponding to an element $i \in U$ to a vertex v_j corresponding to a set $S_j \in \mathcal{S}$ if $u_i \in S_j$. We color all these vertices blue. We add two isolated blue vertices b and b'. We connect all the vertices of type $u_i \in U$ to each other. Also, we add a red vertex r and connect all vertices $u_i \in U$ to r. Now, note that the vertices $U \cup \{r\}$ form a clique whereas the vertices v_j along with b and b' form an independent set. Thus, our constructed graph (G, c) with the coloring c is a split graph. See Fig. 2.

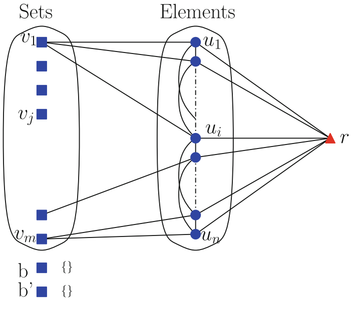

Fig. 2. Reduction from SET COVER to RED-BLUE SEPARATION of Theorem 1. All vertices are blue, except vertex r, which is red. (Color figure online)

Claim 1. $\mathcal{S}$ has a set cover of size k if and only if G has a red-blue separating set of size at most $k + 1$. $\square$

Theorem 2. Red-Blue Separation *is NP-hard for bipartite planar sub-cubic graphs of girth at least 12 when the color classes have almost the same size.*

Proof. We reduce from Dominating Set, which is NP-hard for bipartite planar sub-cubic graphs with girth at least 12 that contain some degree-2 vertices [28]. We reduce any instance (G, k) of Dominating Set to an instance $((H, c), k')$ of Red-Blue Separation, where $k' = k + 1$ and the number of red and blue vertices in c differ by at most 2.

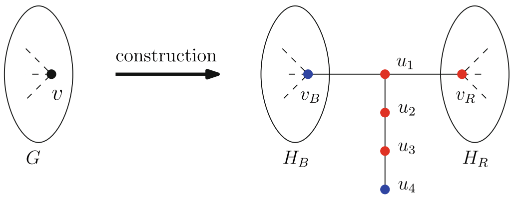

Fig. 3. Reduction from Dominating Set to Red-Blue Separation of Theorem 2. Vertices v_B and u_4 are blue, the others are red. (Color figure online)

Construction. We create two disjoint copies of G namely H_B and H_R and color all vertices of H_B blue and all vertices of H_R red. Select an arbitrary vertex v of degree-2 in G (we may assume such a vertex exists in G by the reduction of [28]) and look at its corresponding vertices $v_R \in V(H_R)$ and $v_B \in V(H_B)$. We connect v_R and v_B with the head of the path u_1, u_2, u_3, u_4 as shown in Fig. 3. The tail of the path, i.e. the vertex u_4, is colored blue and the remaining three vertices u_1, u_2 and u_3 are colored red. Our final graph H is the union of H_R, H_B and the path u_1, u_2, u_3, u_4 and the coloring c as described. Note that if G is a connected bipartite planar sub-cubic graph of girth at least g, then so is H (since v was selected as a vertex of degree-2). We make the following claim.

Claim 2. The instance (G, k) is a YES-instance of Dominating Set if and only if $\text{sep}_{\text{RB}}(H, c) \leq k' = k + 1$. □

In the previous reduction, we could choose any class of instances for which Dominating Set is known to be NP-hard. We could also simply take two copies of the original graph and obtain a coloring with two equal color class sizes (but then we obtain a disconnected instance). In contrast, in the geometric setting, the problem is fixed-parameter-tractable when parameterised by the size of the smallest color class [3], and by the solution size [19]. It is also 2-approximable [5].

2.2 Positive Algorithmic Results

We start with a reduction to Set Cover implying an approximation algorithm.

Proposition 3. Red-Blue Separation *has a polynomial time* $(2 \ln n)$*-factor approximation algorithm.*

Proposition 4. *Let* (G, c) *be a red-blue colored triangle-free and twin-free graph with* R, B *the two color classes. Then,* $sep_{RB}(G, c) \leq 3 \min\{|R|, |B|\}$*.*

Proof. Without loss of generality, we assume $|R| \leq |B|$. We construct a red-blue separating set S of (G, c). First, we add all red vertices to S. It remains to separate every red vertex from its blue neighbors. If a red vertex v has at least two neighbors, we add (any) two such neighbors to S. Since G is triangle-free, no blue neighbor of v is in the closed neighborhood of both these neighbors of v, and thus v is separated from all its neighbors. If v had only one neighbor w, and it was blue, then we separate w from v by adding one arbitrary neighbor of w (other than v) to S. Since G is triangle-free, v and w are separated. Thus, we have built a red-blue separating set S of size at most $3|R|$. □

Proposition 5. *Let* (G, c) *be a red-blue colored twin-free graph with maximum degree* $\Delta \geq 3$*. Then,* $sep_{RB}(G, c) \leq \Delta \min\{|R|, |B|\}$*.*

Proof. Without loss of generality, let us assume $|R| \leq |B|$. We construct a red-blue separating set S of (G, c). Let v be any red vertex. If there is a blue vertex w whose closed neighborhood contains all neighbors of v (w could be a neighbor of v), we add both v and w to S. If v is adjacent to w, since they cannot be twins, there must be a vertex z that can separate v and w; we add z to S. Now, v is separated from every blue vertex in G.

If such a vertex w does not exist, then we add all neighbors of v to S. Now again, v is separated from every vertex of G. Thus, we have built a red-blue separating set S of size at most $\Delta|R|$. □

The previous propositions imply that RED-BLUE SEPARATION can be solved in XP time for the parameter "size of a smallest color class" on triangle-free graphs and on graphs of bounded degree (by a brute-force search algorithm). This is in contrast with the fact that in general graphs, it remains hard even when the smallest color class has size 1 by Theorem 1.

Theorem 6. RED-BLUE SEPARATION *on graphs whose vertices belong to the color classes* R *and* B *can be solved in time* $O(n^{3\min\{|R|,|B|\}})$ *on triangle-free graphs and in time* $O(n^{\Delta\min\{|R|,|B|\}})$ *on graphs of maximum degree* Δ*.*

3 Extremal Values and Bounds for max-sep$_{\mathbf{RB}}$

We denote by $\mathrm{sep}(G)$ the smallest size of a (non-colored) separating set of G, that is, a set that separates *all* pairs of vertices. We will use the relation $\text{max-sep}_{\mathrm{RB}}(G) \leq \mathrm{sep}(G)$, which clearly holds for every twin-free graph G.

3.1 Lower Bounds for General Graphs

We can have a large twin-free colored graph with solution size 2 (for example, in a large blue path with a single red vertex, two vertices suffice). We show that in every twin-free graph, there is always a coloring that requires a large solution.

Theorem 7. *For any twin-free graph G of order $n \geq 1$ and $n \notin \{8, 9, 16, 17\}$, we have max-sep$_{RB}(G) \geq \lfloor \log_2(n) \rfloor$.*

Proof. Let G be a twin-free graph of order n with $\text{max-sep}_{\text{RB}}(G) = k$. There are 2^n different red-blue colorings of G. For each such coloring c, we have $\text{sep}_{\text{RB}}(G, c) \leq k$. Consider the set of vertex subsets of G which are separating sets of size k for some red-blue colorings of G. Notice that each red-blue coloring has a separating set of cardinality k. There are at most $\binom{n}{k} \leq n^k$ such sets.

Consider such a separating set S and consider the set $I(S)$ of subsets S' of S for which there exists a vertex v of G with $N[v] \cap S = S'$. Let i_S be the number of these subsets: we have $i_S \leq 2^{|S|} \leq 2^k$. If S is a separating set for (G, c), then all vertices having the same intersection between their closed neighborhood and S must receive the same color by c. Thus, there are at most $2^{i_S} \leq 2^{2^k}$ red-blue colorings of G for which S is a separating set. Overall, we thus have $2^n \leq \binom{n}{k} 2^{2^k} \leq n^k 2^{2^k}$, and thus $n \leq k \log_2(n) + 2^k$.

We now claim that this implies that $k \geq \log_2(n - \log_2(n)\log_2(n))$. Suppose to the contrary that this is not the case, then we would obtain:

$$n < \log_2(n - \log_2(n)\log_2(n))\log_2(n) + n - \log_2(n)\log_2(n)$$
$$n < \log_2(n)\log_2(n) + n - \log_2(n)\log_2(n)$$

And thus $n < n$, a contradiction. Since k is an integer, we actually have $k \geq \lceil \log_2(n - \log_2(n)\log_2(n)) \rceil$. To conclude, one can check that whenever $n \geq 70$, we have $\lceil \log_2(n - \log_2(n)\log_2(n)) \rceil \geq \lfloor \log_2(n) \rfloor$. Moreover, if we compute values for $2^n - \binom{n}{k} 2^{2^k}$ when $1 \leq n \leq 69$ and $k = \lfloor \log_2(n) \rfloor - 1$, then we observe that this is negative only when $n \in \{8, 9, 16, 17\}$. Thus, $\lfloor \log_2(n) \rfloor$ is a lower bound for $\text{max-sep}_{\text{RB}}(G)$ as long as $n \notin \{8, 9, 16, 17\}$. □

The bound of Theorem 7 is tight for infinitely many values of n.

Proposition 8. *For any integers $k \geq 1$ and $n = 2^k$, there exists a graph G of order n with max-sep$_{RB}(G) = k$.*

We next relate parameter $\text{max-sep}_{\text{RB}}$ to other graph parameters.

Theorem 9. *Let G be a graph on n vertices. Then, sep$(G) \leq \min\{\lceil \log_2(n) \rceil \cdot$ max-sep$_{RB}(G), \lceil \log_2(\Delta(G) + 1) \rceil \cdot$ max-sep$_{RB}(G) + \gamma(G)\}$, where $\gamma(G)$ is the domination number of G and $\Delta(G)$ its maximum degree.*

Proof. Let G be a graph on $2^{k-1} + 1 \leq n \leq 2^k$ vertices for some integer k. We denote each vertex by a different k-length binary word $x_1 x_2 \cdots x_k$ where each $x_i \in \{0, 1\}$. Moreover, we give k different red-blue colorings $c_1, \ldots, c_k$ such that vertex $x_1 x_2 \cdots x_k$ is red in coloring c_i if and only if $x_i = 0$ and blue otherwise. For each i, let S_i be an optimal red-blue separating set of (G, c_i). We have $|S_i| \leq \text{max-sep}_{\text{RB}}(G)$ for each i. Let $S = \bigcup_{i=1}^{k} S_i$. Now, $|S| \leq k \cdot \text{max-sep}_{\text{RB}}(G) = \lceil \log_2(n) \rceil \cdot \text{max-sep}_{\text{RB}}(G)$. We claim that S is a separating set of G. Assume to the contrary that for two vertices $x = x_1 x_2 \cdots x_k$ and $y = y_1 y_2 \cdots y_k$, $N[x] \cap S =$

$N[y] \cap S$. For some i, we have $y_i \neq x_i$. Thus, in coloring c_i, vertices x and y have different colors and hence, there is a vertex $s \in c_i$ such that $s \in N[y] \triangle N[x]$, a contradiction which proves the first bound.

Let S be an optimal red-blue separating set for such a coloring c and let D be a minimum-size dominating set in G; $S \cup D$ is also a red-blue separating set for coloring c. At most $\Delta(G) + 1$ vertices of G may have the same closed neighborhood in D. Thus, we may again choose $\lceil \log_2(\Delta(G) + 1) \rceil$ colorings and optimal separating sets for these colorings, each coloring (roughly) halving the number of vertices having the same vertices in the intersection of separating set and their closed neighborhoods. Since each of these sets has size at most max-sep$_{\mathrm{RB}}(G)$, we get the second bound. □

We do not know whether the previous bound is reached, but as seen next, there are graphs G such that sep$(G) = 2$max-sep$_{\mathrm{RB}}(G)$.

Proposition 10. *Let $G = K_{k_1,\ldots,k_t}$ be a complete t-partite graph for $t \geq 2$, $k_i \geq 5$ odd for each i. Then sep$(G) = n - t$ and max-sep$_{RB}(G) = (n - t)/2$.*

3.2 Upper Bound for General Graphs

We will use the following classic theorem in combinatorics to show that we can always spare one vertex in the solution of Max Red-Blue Separation.

Theorem 11 (Bondy's Theorem [2]). *Let V be an n-set with a family $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n\}$ of n distinct subsets of V. There is an $(n-1)$-subset X of V such that the sets $\mathcal{A}_1 \cap X, \mathcal{A}_2 \cap X, \mathcal{A}_3 \cap X, \ldots, \mathcal{A}_n \cap X$ are still distinct.*

Corollary 12. *For any twin-free graph G on n vertices, we have max-sep$_{RB}(G)$ $\leq$ sep$(G) \leq n - 1$.*

This bound is tight for every even n for complements of *half-graphs* (studied in the context of identifying codes in [14]).

Definition 13 (Half-graph [13]). *For any integer $k \geq 1$, the* half-graph H_k *is the bipartite graph on vertex sets $\{v_1, \ldots, v_k\}$ and $\{w_1, \ldots, w_k\}$, with an edge between v_i and w_j if and only if $i \leq j$.*

The complement $\overline{H_k}$ of H_k thus consists of two cliques $\{v_1, \ldots, v_k\}$ and $\{w_1, \ldots, w_k\}$ and with an edge between v_i and w_j if and only if $i > j$.

Proposition 14. *For every $k \geq 1$, we have max-sep$_{RB}(\overline{H_k}) = 2k - 1$.*

3.3 Upper Bound for Trees

We will now show that a much better upper bound holds for trees.

Degree-1 vertices are called *leaves* and the set of leaves of the tree T is $L(T)$. Vertices adjacent to leaves are called *support vertices*, and the set of support vertices of T is denoted $S(T)$. We denote $\ell(T) = |L(T)|$ and $s(T) = |S(T)|$. The set of support vertices with exactly i adjacent leaves is denoted $S_i(T)$ and the set of leaves adjacent to support vertices in $S_i(T)$ is denoted $L_i(T)$. Observe that $|L_1(T)| = |S_1(T)|$. Moreover, let $L_+(T) = L(T) \setminus L_1(T)$ and $S_+(T) = S(T) \setminus S_1(T)$. We denote the sizes of these four types of sets $s_i(T), \ell_i(T), s_+(T)$ and $\ell_+(T)$.

To prove our upper bound for trees, we need Theorems 15 and 16.

Theorem 15. *For any tree T of order $n \geq 5$, we have max-sep$_{RB}(T) \leq \frac{n+s(T)}{2}$.*

Proof. Observe that the claim holds for stars (select the vertices of the smallest color class among the leaves, and at least two leaves). Thus, we assume that $s(T) \geq 2$. Let c be a coloring of T such that $\text{max-sep}_{\text{RB}}(T) = \text{sep}_{\text{RB}}(T, c)$.

We build two separating sets C_1 and C_2; the idea is that one of them is small. We choose a non-leaf vertex x and add to the first set C_1' every vertex at odd distance from x and every leaf. If there is a support vertex $u \in S_1(T) \cap C_1'$ and an adjacent leaf $v \in L_1(T) \cap N(u)$, we create a separating set C_1 from C_1' by shifting the vertex away from leaf v to some vertex $w \in N(u) \setminus L(T)$. We construct in a similar manner sets C_2' and C_2, except that we add the vertices at even distance from x to C_2' and do the shifting when $u \in S_1(T)$ has even distance to x.

Claim 3. Both C_1 and C_2 are separating sets.

Let us denote by $NS_3(T)$ a smallest set of vertices in T such that for each vertex $v \in S_3(T)$ which has $N(v) \cap S_+(T) = \emptyset$, we have at least one vertex $u \in N(v) \setminus L(T)$ in $NS_3(T)$ (such a set exists since T is not a star).

We assume that out of the two sets C_1' and C_2', C_a' ($a \in \{1, 2\}$) has less vertices among the vertices in $V(T) \setminus (L(T) \cup S_+(T) \cup NS_3(T))$. In particular, it contains at most half of those vertices and we have $|C_a' \setminus (L(T) \cup S_+(T) \cup NS_3(T))| \leq (n - \ell(T) - s_+(T) - |NS_3(T)|)/2$. Now, we will construct set C from C_a'. Let us start by having each vertex in C_a' be in C. Let us then, for each support vertex $u \in S_+(T)$, remove from C every adjacent leaf $w \in L_+(T) \cap N(u)$ such that w is in the more common color class within the vertices in $N(u) \cap L_+(T)$ in coloring c. We then add some vertices to C as follows. For $u \in S_i(T)$, $i \geq 4$, we add u to C and some leaves so that there are at least two vertices in $N(u) \cap C$. We have at most $|L(T) \cap N[u]|/2 + 1$ vertices in $C \cap (N[u] \cap L(T) \cup \{u\})$.

For $i = 3$, we add u and any $v \in NS_3(T) \cap N(u) \setminus C$, depending on which one already belongs to C. Then, if all leaves in $N(u)$ have the same color, we add one of them to C. Hence, we have $|C \cap (L_2(T) \cup NS_3(T))|/s_3(T) \leq 2$.

Finally, for $i = 2$, if the two leaves have same color and $u \notin C_a'$, we add u and one of the two leaves to C. If the two leaves have the same color and $u \in C_a'$, we add a non-leaf neighbor of u to C. If the leaves have different colors, one of them, say v, has the same color as u. We add u to C and shift the vertex in C in the leaves so that v is in C. We added at most two vertices to C in this case. Notice that now we have $S_+(T) \subseteq C$.

Each time, we added to C at most half the considered vertices in $N(u)$, and at most one additional vertex. After these changes, we shift some vertices in C

away from $L_1(T)$ the same way we built C_a from C'_a. As $|C'_a \setminus (L(T) \cup S_+(T) \cup NS_3(T))| \leq (n - \ell(T) - s_+(T) - |NS_3(T)|)/2$, we get:

$$|C| \leq \frac{n - \ell(T) - s_+(T) - |NS_3(T)|}{2} + \ell_1(T) + \frac{\ell_+(T) + |NS_3(T)|}{2} + s_+(T)$$
$$= \frac{n + \ell_1(T) + s_+(T)}{2} = \frac{n + s(T)}{2}.$$

Claim 4. C is a red-blue separating set for coloring c. □

The upper bound of Theorem 15 is tight. Consider, for example, a path on eight vertices. Also, the trees presented in Proposition 18 are within 1/2 from this upper bound. In the following theorem, we offer another upper bound for trees which is useful when the number of support vertices is large.

Theorem 16. *For any tree T of order $n \geq 5$, sep$(T) \leq n - s(T)$.*

The following corollary is a direct consequence of Theorems 15 and 16. Indeed, we have max-sep$_{\mathrm{RB}}(T) \leq \min\{n - s(T), (n + s(T))/2\}$.

Corollary 17. *For any tree T of order $n \geq 5$, we have max-sep$_{RB}(T) \leq \frac{2n}{3}$.*

We next show that Corollary 17 (and Theorem 15) is not far from tight.

Proposition 18. *For any $k \geq 1$, there is a tree T of order $n = 5k + 1$ with max-sep$_{RB}(T) = \frac{3(n-1)}{5} = \frac{n+s(T)-1}{2}$.*

4 Algorithmic Results for Max Red-Blue Separation

The problem Max Red-Blue Separation does not seem to be naturally in the class NP (it is in the second level of the polynomial hierarchy). Nevertheless, we show that it is NP-hard by reduction from a special version of 3-SAT [25].

> 3-SAT-2L
> **Input**: A set of m clauses $C = \{c_1, \ldots, c_m\}$ each with at most three literals, over n Boolean variables $X = \{x_1, \ldots, x_n\}$, and each literal appears at most twice.
> **Question**: Is there an assignment of X where each clause has a true literal?

Theorem 19. Max Red-Blue Separation *is NP-hard even for graphs of maximum degree 12.*

Proof. To show hardness we reduce from the 3-SAT-2L problem. Given an instance σ of 3-SAT-2L with m clauses and n variables, we create an instance (G, k) of Max Red-Blue Separation as follows.

First let us explain the construction of a domination gadget and its properties. A *domination gadget* on vertices v_1 and v_2 is represented in Fig. 4. The vertices v_1 and v_2 may be connected to each other or to some other vertices which is represented by the dashed edges. Both v_1 and v_2 are also connected to

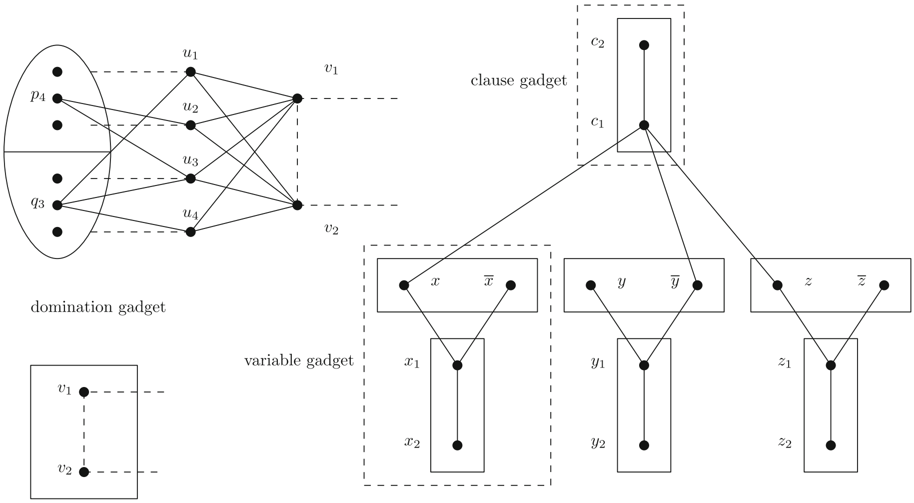

Fig. 4. Reduction from 3-SAT-2L to MAX RED-BLUE SEPARATION.

the vertices u_1, u_2, u_3 and u_4 as shown in the figure. Next we have a clique K_{10} consisting of the vertices $\{p_1, \ldots, p_6, q_1, \ldots, q_4\}$. Every vertex p_i is connected to a unique pair of vertices from $\{u_1, u_2, u_3, u_4\}$ and every vertex q_j is connected to a unique triple of vertices from $\{u_1, u_2, u_3, u_4\}$. For example in the figure we have p_4 connected with the pair of vertices u_2 and u_3 and q_3 connected with the triplet of vertices u_1, u_3 and u_4.

Let $H(v_1, v_2)$ be a subgraph of some graph G such that H is connected to the rest of G only by the vertices v_1 and v_2. We define a worst-coloring of G as any red-blue coloring of G where $\text{sep}_{\text{RB}}(G, c) = \text{max-sep}_{\text{RB}}(G)$. We make the following claim.

Claim 5. For any worst-coloring c of G the optimal red-blue separating code of (G, c) will always contain the vertices u_1, u_2, u_3 and u_4.

The *variable gadget* for a variable x consists of the graph $H(x_1, x_2)$ and $H(x, \overline{x})$ with additional edges $(x_1, x_2), (x_1, x)$ and $(x_1, \overline{x})$. If x_1 and x_2 are colored differently, then either x or $\overline{x}$ needs to be in the red-blue separating set. Selecting at least one of x or $\overline{x}$ also separates x and $\overline{x}$ themselves. The *clause gadget* for a clause $c = (x \vee \overline{y} \vee z)$ is $H(c_1, c_2)$, where c_1 is connected to the vertices $x, \overline{y}$ and z. If c_1 and c_2 are colored differently, then the red-blue separating set should contain at least one of $x, \overline{y}$ or z in order to separate them. This is used to show the following, and complete the proof.

Claim 6. σ is satisfiable if and only if $\text{max-sep}_{\text{RB}}(G) \leq k = 4m + 9n$. □

We can use Theorem 9 and a reduction to SET COVER to show the following.

Theorem 20. MAX RED-BLUE SEPARATION *can be approximated within a factor of $O((\ln n)^2)$ on graphs of order n in polynomial time.*

5 Conclusion

We have initiated the study of Red-Blue Separation and Max Red-Blue Separation on graphs, problems which seem natural given the interest that their geometric version has gathered, and the popularity of its "non-colored" variants Identifying Code on graphs or Test Cover on set systems.

When the coloring is part of the input, the solution size of Red-Blue Separation can be as small as 2, even for large instances; however, we have seen that this is not possible for Max Red-Blue Separation since $\text{max-sep}_{\text{RB}}(G) \geq \lfloor \log_2(n) \rfloor$ for twin-free graphs of order n. $\text{max-sep}_{\text{RB}}(G)$ can be as large as $n-1$ in general graphs, yet, on trees, it is at most $2n/3$ (we do not know if this is tight, or if the upper bound of $3n/5$, which would be best possible, holds). It would also be interesting to see if other interesting upper or lower bounds can be shown for other graph classes.

We have shown that $\text{sep}(G) \leq \lceil \log_2(n) \rceil \cdot \text{max-sep}_{\text{RB}}(G)$. Is it true that $\text{sep}(G) \leq 2\text{max-sep}_{\text{RB}}(G)$? As we have seen, this would be tight.

We have also shown that Max Red-Blue Separation is NP-hard, yet it does not naturally belong to NP. Is the problem actually hard for the second level of the polynomial hierarchy?

References

1. Bollobás, B., Scott, A.D.: On separating systems. Eur. J. Comb. **28**, 1068–1071 (2007)
2. Bondy, J.A.: Induced subsets. J. Comb. Theory Ser. B **12**(2), 201–202 (1972)
3. Bonnet, É., Giannopoulos, P., Lampis, M.: On the parameterized complexity of red-blue points separation. J. Comput. Geom. **10**(1), 181–206 (2019)
4. Bousquet, N., Lagoutte, A., Li, Z., Parreau, A., Thomassé, S.: Identifying codes in hereditary classes of graphs and VC-dimension. SIAM J. Discret. Math. **29**(4), 2047–2064 (2015)
5. Călinescu, G., Dumitrescu, A., Karloff, H.J., Wan, P.: Separating points by axis-parallel lines. Int. J. Comput. Geom. Appl. **15**(6), 575–590 (2005)
6. Charbit, E., Charon, I., Cohen, G., Hudry, O., Lobstein, A.: Discriminating codes in bipartite graphs: bounds, extremal cardinalities, complexity. Adv. Math. Commun. **2**(4), 403–420 (2008)
7. Chlebus, B.S., Nguyen, S.H.: On finding optimal discretizations for two attributes. In: Polkowski, L., Skowron, A. (eds.) RSCTC 1998. LNCS (LNAI), vol. 1424, pp. 537–544. Springer, Heidelberg (1998). https://doi.org/10.1007/3-540-69115-4_74
8. Crowston, R., Gutin, G., Jones, M., Muciaccia, G., Yeo, A.: Parameterizations of test cover with bounded test sizes. Algorithmica **74**(1), 367–384 (2016)
9. Dey, S., Foucaud, F., Nandy, S.C., Sen, A.: Discriminating codes in geometric setups. In: Proceedings of the 31st International Symposium on Algorithms and Computation (ISAAC 2020). Leibniz International Proceedings in Informatics, vol. 181, pp. 24:1–24:16 (2020)
10. De Bontridder, K.M.J., et al.: Approximation algorithms for the test cover problem. Math. Program. Ser. B **98**, 477–491 (2003)
11. Dinur, I., Steurer, D.: Analytical approach to parallel repetition. In: ACM Symposium on Theory of Computing, vol. 46, pp. 624–633 (2014)

12. Downey, R.G., Fellows, M.R.: Parameterized Complexity. Springer, Heidelberg (1999)
13. Erdös, P.: Some combinatorial, geometric and set theoretic problems in measure theory. In: Kölzow, D., Maharam-Stone, D. (eds.) Measure Theory Oberwolfach 1983. LNM, vol. 1089, pp. 321–327. Springer, Heidelberg (1984). https://doi.org/10.1007/BFb0072626
14. Foucaud, F., Guerrini, E., Kovše, M., Naserasr, R., Parreau, A., Valicov, P.: Extremal graphs for the identifying code problem. Eur. J. Comb. **32**(4), 628–638 (2011)
15. Gledel, V., Parreau, A.: Identification of points using disks. Discret. Math. **342**, 256–269 (2019)
16. Har-Peled, S., Jones, M.: On separating points by lines. Discret. Comput. Geom. **63**, 705–730 (2020)
17. Henning, M.A., Yeo, A.: Distinguishing-transversal in hypergraphs and identifying open codes in cubic graphs. Graphs Comb. **30**(4), 909–932 (2014)
18. Karpovsky, M.G., Chakrabarty, K., Levitin, L.B.: On a new class of codes for identifying vertices in graphs. IEEE Trans. Inf. Theory **44**, 599–611 (1998)
19. Kratsch, S., Masařík, T., Muzi, I., Pilipczuk, M., Sorge, M.: Optimal discretization is fixed-parameter tractable. In: Proceedings of the 32nd ACM-SIAM Symposium on Discrete Algorithms (SODA 2021), pp. 1702–1719 (2021)
20. Kujala, J., Elomaa, T.: Improved algorithms for univariate discretization of continuous features. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) PKDD 2007. LNCS (LNAI), vol. 4702, pp. 188–199. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74976-9_20
21. Lobstein, A.: Watching systems, identifying, locating-dominating and discriminating codes in graphs: a bibliography. https://www.lri.fr/~lobstein/debutBIBidetlocdom.pdf
22. Misra, N., Mittal, H., Sethia, A.: Red-blue point separation for points on a circle. In: Proceedings of the 32nd Canadian Conference on Computational Geometry (CCCG 2020), pp. 266–272 (2020)
23. Moret, B.M.E., Shapiro, H.D.: On minimizing a set of tests. SIAM J. Sci. Stat. Comput. **6**(4), 983–1003 (1985)
24. Rényi, A.: On random generating elements of a finite Boolean algebra. Acta Scientiarum Mathematicarum Szeged **22**, 75–81 (1961)
25. Tovey, C.A.: A simplified NP-complete satisfiability problem. Discret. Appl. Math. **8**(1), 85–89 (1984)
26. Ungrangsi, R., Trachtenberg, A., Starobinski, D.: An implementation of indoor location detection systems based on identifying codes. In: Aagesen, F.A., Anutariya, C., Wuwongse, V. (eds.) INTELLCOMM 2004. LNCS, vol. 3283, pp. 175–189. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30179-0_16
27. Vapnik, V.N., Chervonenkis, A.Y.: On the uniform convergence of relative frequencies of events to their probabilities. Theory Probab. Appl. **16**(2), 264–280 (1971)
28. Zvervich, I.E., Zverovich, V.E.: An induced subgraph characterization of domination perfect graphs. J. Graph Theory **20**(3), 375–395 (1995)