The RED-BLUE SEPARATION problem on graphs [☆],[☆☆]

Subhadeep Ranjan Dev [a], Sanjana Dey [b],[*],[1], Florent Foucaud [c], Ralf Klasing [d], Tuomo Lehtilä [e],[2]

[a] *ACM Unit, Indian Statistical Institute, Kolkata, India*
[b] *National University of Singapore, Singapore*
[c] *Université Clermont Auvergne, CNRS, Clermont Auvergne INP, Mines Saint Étienne, LIMOS, 63000 Clermont-Ferrand, France*
[d] *Université de Bordeaux, Bordeaux INP, CNRS, LaBRI, UMR 5800, Talence, France*
[e] *University of Turku, Department of Mathematics and Statistics, Turku, Finland*

ARTICLE INFO

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 $S \subseteq V$, called *red-blue separating set*, such that for every red-blue pair of vertices, there is a vertex $s \in S$ 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 order of the input graph. 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 of the graph 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 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 $O(\ln^2 n)$, where n is the order of the input graph. We derive tight bounds on the cardinality of an optimal solution of 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.

© 2023 Elsevier B.V. All rights reserved.

[*] Corresponding author.
E-mail address: info4.sanjana@gmail.com (S. Dey).
[1] Most of the work of this author was done when she was at the Indian Statistical Institute, Kolkata, India.
[2] Most of the work of this author was done when he was in Univ Lyon, Université Claude Bernard, CNRS, LIRIS - UMR 5205, F69622, France.

1. Introduction

We introduce and study the RED-BLUE SEPARATION problem on graphs. Separation problems for discrete structures have been studied extensively from various perspectives. In the 1960s, Rényi [30] 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. [2,6,22,29].

In this problem, one is given a set system, that is, a collection $\mathcal{S}$ of sets over a set V of vertices. One wishes to select a solution subset $\mathcal{C}$ of sets in $\mathcal{S}$ to *separate* every pair of vertices, that is, for every pair of vertices, there is a set of $\mathcal{C}$ that contains one vertex but not the other (in other words, for each vertex, the subset of sets in $\mathcal{C}$ containing it is unique). These problems have numerous applications in areas such as monitoring and fault-detection in networks [32], biological testing [29], and machine learning [25].

Some special cases, which are extensively studied as well, are as follows. The graph version of SEPARATION (where the sets of the input set system are the closed neighborhoods of a graph and the set system vertices are the graph vertices), is called IDENTIFYING CODE [4,16,17,23]. The geometric version of SEPARATION (where the sets of the input set system are geometric objects and the set system vertices are points), called GEOMETRIC DISCRIMINATING CODE or POINT SEPARATION, is also extensively studied, see [5,11,18,20].

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].

A classic topic in algorithms and discrete mathematics is the problem SET COVER over set systems, that generalizes both GEOMETRIC SET COVER problems and the graph problem DOMINATING SET. Similarly, as explained above, the set system problem SEPARATION generalizes both the GEOMETRIC DISCRIMINATING CODE problem and the graph problem IDENTIFYING CODE. It is thus natural to initiate the study of the graph version of RED-BLUE SEPARATION. That is the goal of this paper.

Problem definition   In the graph setting, we are given a graph G and a red-blue coloring $c : V(G) \rightarrow \{\text{red, blue}\}$ 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 of vertices r, b (say), there is a vertex from S whose closed neighborhood contains exactly one of r or 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 max-$\text{sep}_{\text{RB}}(G)$ 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 max-$\text{sep}_{\text{RB}}(G) \leq k$?

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

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[3] 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[4] of order n, and when one color class has

---

[3] One class consists of vertices colored *red* and the other class consists of vertices colored *blue*.
[4] 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.

Fig. 1. A path of 6 vertices where (a) $\text{sep}_{RB}(P_6, c) = 1$ and (b) $\text{max-sep}_{RB}(P_6) = 3$. The square vertices are blue, the round ones are red; the members of the red-blue separating set are circled. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

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 [3,24].)

As the coloring is not specified, $\text{max-sep}_{RB}(G)$ is a parameter that is worth studying from a structural viewpoint. We 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 $O(\ln^2 n)$.

We also study the possible values for $\text{max-sep}_{RB}(G)$. We show the existence of tight bounds on $\text{max-sep}_{RB}(G)$ in terms of n, the order 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 $\text{max-sep}_{RB}(G) \leq \frac{2n}{3}$. We also give bounds in terms of the (non-colored) separation number.

An extended abstract of this paper was presented in the conference IWOCA'22 and appeared in the proceedings as [10]. The present paper includes all the proofs that were missing from the conference version.

Related work RED-BLUE SEPARATION has been studied in the geometric setting of red and blue points in the Euclidean plane [3,5,28]. 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, respectively, 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 [7,24,25]. 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 [24]. A polynomial time algorithm for a special case was recently given in [28].

The SEPARATION problem for set systems (also known as TEST COVER and DISCRIMINATING CODE) was introduced in the 1960s [30] and widely studied from a combinatorial point of view [1,2,6,22] as well as from the algorithmic perspective for the settings of classical, approximation and parameterized algorithms [8,12,29]. The associated graph problem is called IDENTIFYING CODE [23] and has also been extensively studied (see [26] for an online bibliography with almost 500 references as of January 2022); geometric versions of SEPARATION have been studied as well [11,18,20]. The SEPARATION problem is also closely related to the VC DIMENSION problem [33] 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*, i.e., 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.

Another way of transferring the problem definition of the geometric RED-BLUE SEPARATION problem to graphs, is the MULTICUT problem [9,27], where separation is achieved by disconnecting the graph rather than by dominating the vertices. Here, one is given a graph G and a set of vertex-pairs, and one wishes to select a minimum-size set S of vertices or edges of the graph so that the two vertices of every input pair belong to two distinct connected components of the graph $G \setminus S$. If one was given a red-blue colored graph and wished to separate every red vertex from every blue vertex in this way, then one may set as the input pairs of the MULTICUT problem, all possible red-blue vertex pairs.

Structure of the paper We start with the algorithmic results on RED-BLUE SEPARATION in Section 2. In Subsection 2.1 we consider the complexity of RED-BLUE SEPARATION and in Subsection 2.2 we give algorithms for the problem. We then present the bounds on max-sep_{RB} in Section 3. In particular, Subsection 3.1 considers general lower bounds, Subsection 3.2 considers general upper bounds and Subsection 3.3 considers bounds for trees. Then, in Section 4, we give an approximation algorithm for MAX RED-BLUE SEPARATION (using a bound from Section 3) and show that it is an NP-hard problem. We conclude in Section 5.

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 there 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 $\forall v \in V, N[v] \cap D \neq \emptyset$?

Subsection 2.1 contains hardness results on RED-BLUE SEPARATION and Subsection 2.2 contains algorithms for the problem.

2.1. Hardness

In this subsection, we show that RED-BLUE SEPARATION is an NP-hard problem on subcubic bipartite planar graphs, that it is hard to approximate, and that the problem is $W[2]$-hard when parameterized by the solution size.

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 [13], and from the fact that SET COVER is W[2]-hard when parameterized by the solution size [14].

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.

Claim 1.1. $\mathcal{S}$ *has a set cover of size at most k if and only if G has a red-blue separating set of size at most $k + 1$.*

Proof of claim. Let $\mathcal{C}$ be a set cover of $(U, \mathcal{S})$ of size k. We construct a red-blue separating set S of (G, c) of size $k + 1$ as follows. For each set $s_j \in \mathcal{S}$ selected in the set cover $\mathcal{C}$, we choose the corresponding vertex v_j in S. We also include the vertex r in S. Observe that some blue vertices may have the empty code and the red vertex has itself as the code. Also, the vertices u_i are dominated by the vertices v_j and have some unique code different from r. Therefore, S is a separating set of (G, c) of size $k + 1$.

Conversely, consider a red-blue separating set S of (G, c) of size $k' = k + 1$. Since the vertices in $U \cup \{r\}$ form a clique, choosing any vertex from this set will not separate any two vertices of the clique. Hence, to separate r from the vertices in U, each vertex in U has to be dominated by a vertex in S. Furthermore, these vertices form a set cover. Moreover, at least one of vertices b and r is dominated by a vertex. Thus, there exists a set cover of size at most $k' - 1 = k$. (□)

This completes the proof of the theorem. □

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 [34]. 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.

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 [34]) 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 u_1 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.
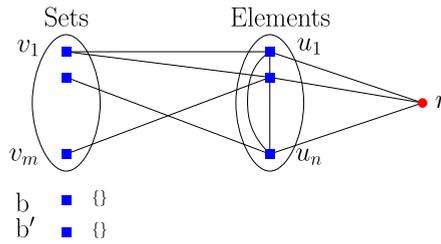
Fig. 2. Reduction from SET COVER to RED-BLUE SEPARATION of Theorem 1. Vertex r is red, the others are blue.
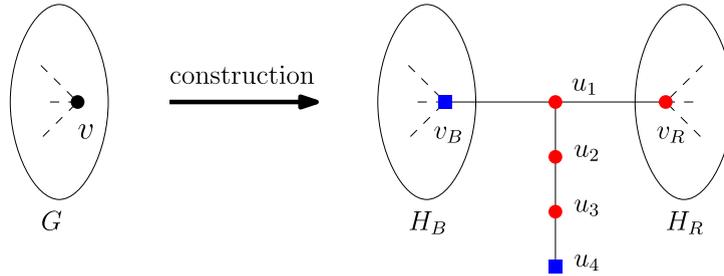
Fig. 3. Reduction from DOMINATING SET to RED-BLUE SEPARATION of Theorem 2. Vertices v_B and u_4 are blue, the others are red.

Claim 2.1. *The instance (G, k) is a YES-instance of* DOMINATING SET *if and only if* $\mathrm{sep}_{\mathrm{RB}}(H, c) \leq k' = k + 1$.

Proof of claim. ($\Longrightarrow$) Let D be a dominating set of G of size k. We construct a red-blue separating set S of (H, c) of size at most $k + 1$ as follows. For each vertex in D, include its corresponding vertex in H_R in S. Also include in S, the vertex $u_2 \in V(H)$. Observe that all blue vertices have the empty code and all red vertices have some non-empty code. Therefore S is a separating set of (H, c) of size at most $k + 1$.

($\Longleftarrow$) Consider a red-blue separating set S of (H, c) of size k'. Let us assume that the red vertices of (H, c) did not get the empty code. The argument when the blue vertices do not get the empty code is similar. Then, the set S also dominates $V(H_R) \cup \{u_1, u_2, u_3\}$. Since u_3 can only be dominated by a vertex x in $\{u_2, u_3, u_4\}$, the vertices in $V(H_R)$ are dominated by $S' = S \setminus \{x\}$. If $S' = S' \cap V(H_G)$, then the set D formed by choosing the corresponding vertices of S' in G is a dominating set of size $k = k' - 1$. Otherwise, $S' \setminus \{u_1\} = S' \cap V(H_R)$, and the set D, formed by choosing the corresponding vertices of $S' \setminus \{u_1\}$ together with x, in G is a dominating set of G of size at most $k = k' - 1$. $(\Box)$

This completes the proof of the theorem. $\square$

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 parameterized by the size of the smallest color class [3], and by the solution size [24]. It is also 2-approximable [5].

We can also provide another reduction, as follows.

Theorem 3. RED-BLUE SEPARATION *is NP-hard even when the input is a subcubic planar bipartite graph of girth at least 12 and the size of the smallest color class is $k + 1$.*

Proof. We reduce from DOMINATING SET.

DOMINATING SET is NP-hard when the input graph is a subcubic planar bipartite graph of girth at least 12 [34]. We reduce an instance (G, k) of the DOMINATING SET to an instance $((H, c_{k+1}), k)$ of RED-BLUE SEPARATION where c_{k+1} is a coloring of H with the minimum color class being of size at most $k + 1$.

Construction Without loss of generality let us assume the smaller sized color class to be red. For H we create a copy G_H of G with all its vertices colored blue and an independent set I of size $k + 1$ all its vertices colored red vertices (see Fig. 4). We now make the following claim.

Claim 3.1. (G, k) *is a YES-instance of* DOMINATING SET *if and only if* $((H, c_{k+1}), k)$ *is a YES-instance of* RED-BLUE SEPARATION.

Proof of claim. ($\Longrightarrow$) Let D be a dominating set of G. We construct a red-blue separating set of (H, c_{k+1}) by selecting the corresponding vertices of D in G_H. Since D is a dominating set of G, all blue vertices receive a non-empty code. The red vertices on the other hand receive the empty code and we have a valid separating set.
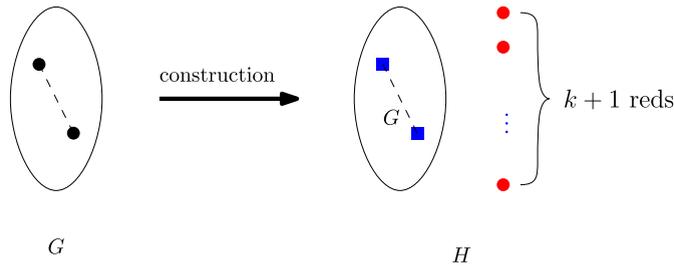
Fig. 4. Construction from an instance of DOMINATING SET to an instance of RED-BLUE SEPARATION where size of the smaller color class is bounded. In H, the square vertices are blue, the round ones are red.
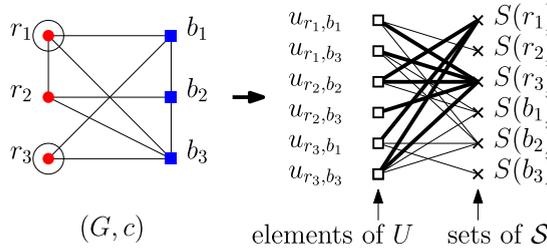
Fig. 5. A reduction instance from RED-BLUE SEPARATION of (G, c) (where the square vertices are blue, the round ones are red) to SET COVER of (U, S). The separating set in the colored graph (G, c) and the corresponding set cover in the set system (U, S) has been highlighted.

(⟸) Let (H, c_{k+1}) have a separating set C of size at most k. Since there are $k + 1$ independent red vertices, there will be at least one red vertex which receives the empty code. So, in order for C to be a valid separating set, all blue vertices must receive some non-empty code. This implies that $C \cap V(G_H)$ is a dominating set of G_H of size at most k and which implies a dominating set of G of size at most k. (□)

This completes the proof. □

2.2. Positive algorithmic results

In this subsection, we design an approximation algorithm for RED-BLUE SEPARATIONwhich runs in polynomial-time, some bounds for the problem based on the size of the color classes, and finally show that if the graph is triangle-free or has a small maximum degree Δ, then there exists a polynomial-time algorithm if the size of the smaller color class is constrained by a constant.

We start with a reduction to SET COVER implying an approximation algorithm.

Proposition 4. RED-BLUE SEPARATION *has a polynomial-time* $(2 \ln n)$*-factor approximation algorithm, where n is the input graph's order.*

Proof. We reduce RED-BLUE SEPARATION to SET COVER. Let $((G, c), k)$ be an input instance of RED-BLUE SEPARATION. We reduce it to an instance $((U, S), k)$ of SET COVER. For each red-blue vertex pair (r, b) in G, create an element in U. For each vertex v in G create a set in S with elements (r, b) in U such that v is in the closed neighborhood of exactly one of r and b in G. See Fig. 5. Observe that a set cover C of size k corresponds to a separating set S of size at most k and vice versa. The greedy algorithm for SET COVER has an approximation factor of $\ln |U| + 1$. Since, in our case $|U| \leq n^2/4$, the resulting approximation factor for RED-BLUE SEPARATION is at most $\ln(n^2/4) + 1 \leq 2 \ln n$. □

Proposition 5. *Let (G, c) be a red-blue colored triangle-free and twin-free graph with the two color classes R and B. Then,* $\text{sep}_{RB}(G, c) \leq 3 \min\{|R|, |B|\}$.

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 (see vertex v_1 in Fig. 6). 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 (see vertex v_2 in Fig. 6). Thus, we have built a red-blue separating set S of size at most $3|R|$. □
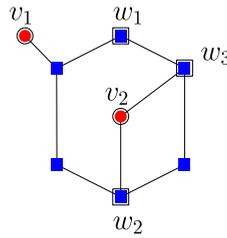
Fig. 6. Illustration of Proposition 5: here v_1 has just one blue neighbor hence w_1 is added in S. v_2's neighbors w_2 and w_3 are also included in S. The square vertices are blue, the round ones are red.
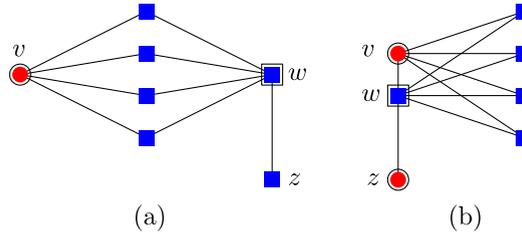
Fig. 7. Illustration of Proposition 6: the two cases when the vertices v and w are (a) not adjacent and (b) adjacent. The square vertices are blue, the round ones are red.

Proposition 6. *Let (G, c) be a red-blue colored twin-free graph with maximum degree $\Delta \geq 3$. Then, $\mathrm{sep}_{\mathrm{RB}}(G, c) \leq \Delta \min\{|R|, |B|\}$.*

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 (see Fig. 7(a)). 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 (see Fig. 7(b)). 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 7. 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$_{\mathrm{RB}}$

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 $\mathrm{max\text{-}sep}_{\mathrm{RB}}(G) \leq \mathrm{sep}(G)$, which clearly holds for every twin-free graph G.

Subsection 3.1 contains lower bounds for $\mathrm{max\text{-}sep}_{\mathrm{RB}}(G)$, subsection 3.2 contains general upper bounds and subsection 3.3 gives (almost tight) bounds for trees.

3.1. Lower bounds for general graphs

In this subsection, we give a tight general lower bound for $\mathrm{max\text{-}sep}_{\mathrm{RB}}$ and consider the relationship between $\mathrm{sep}(G)$ and $\mathrm{max\text{-}sep}_{\mathrm{RB}}(G)$.

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

Proof. Let G be a twin-free graph of order n with $\mathrm{max\text{-}sep}_{\mathrm{RB}}(G) = k$. There are 2^n different red-blue colorings of G. For each such coloring c, we have $\mathrm{sep}_{\mathrm{RB}}(G, c) \leq k$. 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. Hence, we have

$$
\begin{aligned}
2^n &\leq \tbinom{n}{k} 2^{2^k}, \text{ which implies} \\
2^n &\leq n^k 2^{2^k}, \text{ which implies} \\
n &\leq k \log_2(n) + 2^k
\end{aligned}
$$

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:

$$
\begin{aligned}
n &< \log_2(n - \log_2(n) \log_2(n)) \log_2(n) + n - \log_2(n) \log_2(n) \\
n &< \log_2(n) \log_2(n) + n - \log_2(n) \log_2(n)
\end{aligned}
$$

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 $2^n - \tbinom{n}{k} 2^{2^k}$ 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 max-sep$_{RB}(G)$ as long as $n \notin \{8, 9, 16, 17\}$. $\square$

The bound of Theorem 8 is tight for infinitely many values of n.

Proposition 9. *For any integers $k \geq 1$ and $n = 2^k$, there exists a graph G of order n with* max-sep$_{RB}(G) = k$.

Proof. We build G as follows. Let $S = \{s_1, \ldots, s_k\}$ be a set of k vertices. Let T be another set of vertices (disjoint from S). For every subset S' of S of size at least 2, we add a vertex $v_{S'}$ to T and we join it to all vertices of S' and T. The set T induces a clique. Finally, we add an isolated vertex $v_\emptyset$ to G. To see that max-sep$_{RB}(G) \leq k$, notice that S is a separating set of G (regardless of the coloring).

If $k = 1$, the coloring with s_1 Red and $v_\emptyset$ Blue shows that max-sep$_{RB}(G) \geq 1$.

If $k = 2$, color s_1 and s_2 Red and $v_{\{1,2\}}$ and $v_\emptyset$ Blue. To separate $v_{\{1,2\}}$ from s_1 (respectively s_2), s_2 must belong to any separating set (respectively s_1), which shows that max-sep$_{RB}(G) \geq 2$.

If $k \geq 3$, color v_S Blue and the other vertices Red. For each subset S' of S with $|S'| = k - 1$, in order to separate v_S from $v_{S'}$, any separating set needs to contain s_i where $\{s_i\} = S \setminus S'$. This shows that max-sep$_{RB}(G) \geq k$. $\square$

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

Theorem 10. *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 $|S_i| \leq$ max-sep$_{RB}(G)$ for each i. Let $S = \bigcup_{i=1}^k S_i$. Now, $|S| \leq k \cdot$ max-sep$_{RB}(G) = \lceil \log_2(n) \rceil \cdot$ max-sep$_{RB}(G)$. 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$_{RB}(G)$, we get the second bound. $\square$

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$_{RB}(G)$.

Proposition 11. *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$_{RB}(G) = (n - t)/2$.

Proof. Let us have t parts $G_1, \ldots, G_t$ with vertex sets $V(G_i) = \{v_1^i, \ldots, v_{k_i}^i\}$. Let c be such a coloring in G that sep$_{RB}(G, c) =$ max-sep$_{RB}(G)$. Observe that S is a separating set in G if and only if $|S \cap V(G_i)| \geq k_i - 1$. Indeed, if we have $v_j^i, v_h^i \notin S$, then

$N[v_j^i] \cap S = N[v_h^i] \cap S$ since $N(v_j^i) = N(v_h^i)$. Moreover, each vertex in S is separated from vertices not in S and each vertex in $V(G_i)$ is separated from vertices in $V(G_j)$. Thus, $\text{sep}(G) = n - t$.

Next we form a red-blue separating set S' for coloring c. For each part, we choose to set S' every vertex which has the less common color in that part (however, we choose at least two vertices to S' from each part). Observe that $|S'| \leq \sum_{i=1}^{t} \frac{k_i - 1}{2} = (n-t)/2$. Moreover as above, we can see that S' is a red-blue separating set with the help of fact that if $v_j^i \notin S'$ and $v_h^i \notin S'$, then v_j^i and v_h^i have the same color.

Finally, we show that $\text{max-sep}_{\text{RB}}(G) \geq (n-t)/2$. Observe that if we have $u, v \in V(G_i)$ and $u, v \notin S$, then $N[u] \cap S = N[v] \cap S$. Thus, any two vertices $u, v \in V(G_i) \setminus S$ must have the same color. If each part has $(k_i + 1)/2$ red vertices and $(k_i - 1)/2$ blue vertices, then we require at least $\sum_{i=1}^{t} (k_1 - 1)/2 = (n-t)/2$ vertices in S. $\square$

3.2. Upper bound for general graphs

In this subsection, we first show that in every twin-free graph we have $\text{max-sep}_{\text{RB}}(G) \leq n - 1$ and then show that this bound is tight.

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

Proof. Regardless of the coloring, by Bondy's theorem, we can always find a set of size $n - 1$ that separates *all* pairs of vertices. $\square$

This bound is tight for every even n for complements of *half-graphs* (studied in the context of identifying codes in [16]).

Definition 14 *(Half-graph [15]). 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 15. *For every $k \geq 1$, we have $\text{max-sep}_{\text{RB}}(\overline{H_k}) = 2k - 1$.*

Proof. The upper bound follows from Corollary 13.

Consider the red-blue coloring c such that v_i is Blue whenever i is odd and Red, otherwise. If k is odd, w_i is Red whenever i is odd and Blue, otherwise. If k is even, w_i is Blue whenever i is odd and Red, otherwise.

For any integer i between 1 and $k - 1$, v_i and v_{i+1} have different colors and can only be separated by w_i. Likewise, w_i and w_{i+1} have different colors and can only be separated by v_{i+1}. This shows that $\{w_1, \ldots, w_{k-1}\}$ and $\{v_2, \ldots, v_k\}$ must belong to any separating set of $(\overline{H_k}, c)$. Finally, consider w_1 and v_k. They also have different colors and can only be separated by either v_1 or w_k. This shows that we need at least $n - 1$ vertices in any separating set. $\square$

3.3. Upper bound for trees

In this subsection, we show that a much better upper bound holds for trees. In particular, we offer three general upper bounds for $\text{max-sep}_{\text{RB}}(T)$ when T is a tree. Furthermore, we also show that one of our bounds is almost tight for an infinite family of 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 by $s_i(T), \ell_i(T), s_+(T)$ and $\ell_+(T)$, respectively.

For trees, we can show the following result, which is in contrast with the situation for general graphs (or even split graphs, as highlighted by Theorem 1).

Theorem 16. *Let T be a tree on $n \geq 3$ vertices and let c be a coloring with exactly one red (or blue) vertex. We have $\text{sep}_{\text{RB}}(T, c) \leq 2$.*

Proof. Let T be a tree on at least three vertices with coloring c such that there is exactly one red vertex $v \in V(T)$.
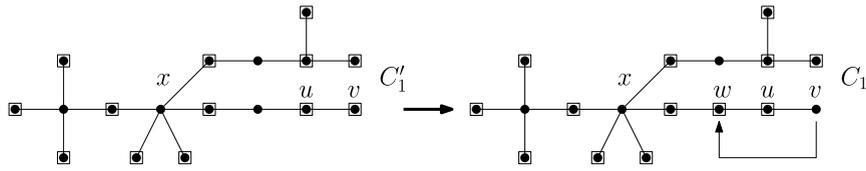
Fig. 8. Construction of C_1 from C_1' where the boxed elements represent members of the set.
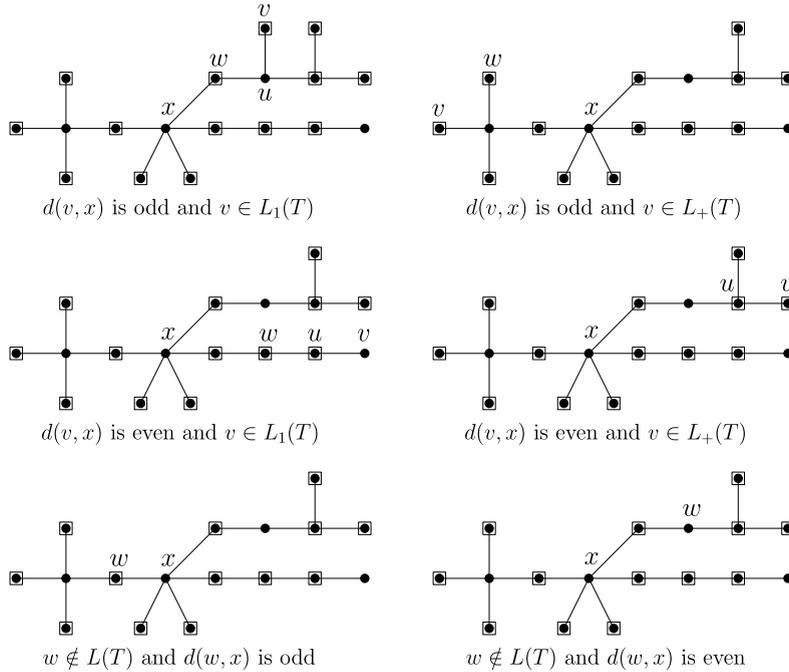
Fig. 9. Illustration of various cases for C_1 in Claim 17.1.

Let us assume first that $v \notin L(T)$. Thus, v has at least two neighbors w and u. If we now include w and u in the separating set S, then v is the only vertex in T which has two adjacent vertices in S and hence, S is a red-blue separating set in T for coloring c.

On the other hand, if $v \in L(T)$, $u \in N(v)$ and $w \in N(u) \setminus \{v\}$, then $S = \{v, w\}$ is a red-blue separating set in T for coloring c, and $\mathrm{sep}_{RB}(T, c) \leq 2$. $\square$

To prove our upper bound for trees, we need Theorems 17 and 18.

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

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 $\max\text{-}\mathrm{sep}_{RB}(T) = \mathrm{sep}_{RB}(T, c)$.

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' (including x itself) and do the shifting when $u \in S_1(T)$ has even distance to x. Sets C_1 and C_2 have been previously considered in [17, Theorem 6]. See Fig. 8.

Claim 17.1. *Both C_1 and C_2 are separating sets.*

Proof of claim. Let us first consider set C_1. See Fig. 9 for helpful illustrations of different cases we go through in the following arguments. Let us first show that each leaf $v \in L(T)$ is separated from all other vertices by C_1. Let $u \in N(v)$ be the adjacent support vertex. If $d(x, v)$ is odd, then $v \in C_1$ and $|N(v) \cap C_1| \geq 2$. Hence, v is separated from every other vertex. If $d(v, x)$ is even and $v \in L_+(T)$, then $v \in C_1$ and $|N[u] \cap C_1| \geq 3$. Hence, v is again separated from other vertices. If $d(v, x)$ is even and
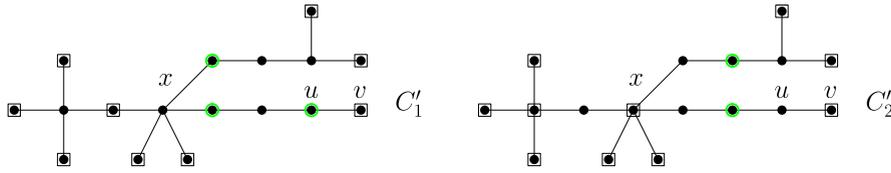
Fig. 10. Comparison of the sets C_1' and C_2' where the vertices highlighted in green belong to the sets $C_1' \setminus (L(T) \cup S_+(T) \cup NS_3(T))$ and $C_2' \setminus (L(T) \cup S_+(T) \cup NS_3(T))$, respectively.

$v \in L_1(T)$, then $v \notin C_1$, $u \in C_1$ and $|N[u] \cap C_1| \geq 2$ due to the shift when we form C_1 from C_1'. If we have $N[w] \cap C_1 = \{u\}$ for some $w \neq u$ and $w \notin L(T)$, then $d(w,x)$ is even and we have $|N[w] \cap C_1| \geq 2$, a contradiction. Thus, any leaf v is separated from each other vertex.

Let us then consider a vertex $w \notin L(T)$ with odd distance $d(w,x)$, then $w \in C_1$. Moreover, each neighbor $u \in N(w)$ has even distance to x and is either a leaf which is separated from w or has at least two neighbors with odd distances to x and hence, $|C_1 \cap N(u)| \geq 2$ and u is separated from w. Finally, if $w \notin L(T)$ and $d(w,x)$ is even, then $w \notin C_1'$ and $|N(w) \cap C_1| \geq 2$. Since there are no 3-cycles nor 4-cycles in a tree, w is separated from all other vertices.

The proof that also C_2 is a separating set is similar. We only swap evens and odds. Hence, the claim follows. □

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 $|C_a' \setminus (L(T) \cup S_+(T) \cup NS_3(T))| \leq (n - \ell(T) - s_+(T) - |NS_3(T)|)/2$. In Fig. 10, a comparison of the sets C_1' and C_2' is shown. Next, 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 $|C \cap (L_3(T) \cup NS_3(T))|/s_3(T) \leq 2$.

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 of the considered vertices in $N(u)$, and at most one other 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 17.2. *C is a red-blue separating set for coloring c.*

Proof of claim. Since C_a is a separating set and $C_a \setminus C \subseteq L_+(T)$, if two vertices w, v are not separated by C, then they were separated by a leaf in $L_+(T)$ in C_a. Moreover, there exists a support vertex $u \in S_+(T)$ such that $v, w \in N(u)$. Recall that $S_+(T) \subseteq C$ and hence, $u \in C$. If w and v are both leaves, then they have the same color and do not need to be separated. In the following, we go through all the other possibilities for w and v.

Assume first that $w, v \notin L(T)$ and $d(w,v) = 1$ and let us say, without loss of generality, that $d(w,x)$ is of such parity that $w \in C_a'$ (and $w \in C$). Notice that we have $w = u$ or $v = u$ in this case. Moreover, since there are no cycles in T, the parities of $d(w,x)$ and $d(v,x)$ differ. Thus, there exists some vertex $b \in N(v) \setminus N[w]$ such that the parity of $d(b,x)$ equals to the parity of $d(w,x)$. Thus, $b \in C_a'$. If b is not a leaf, then $b \in C$ and b separates w and v. Thus, $b \in L(T)$. However, if $b \in L_1(T)$, then $b \in C$ and if $b \in L_+(T)$, then $v \in S_+(T)$ and there exists a leaf in C adjacent to v separating v and w, a contradiction. When $d(w,v) = 2$, neither of w or v are in C since they cannot be separated. Again, we can find some vertex b in C that will separate them as in the previous case.

Moreover, since there are no cycles in T, the parities of $d(w,x)$ and $d(v,x)$ differ. Thus, there exists some vertex $b \in N(v) \setminus N[w]$ such that the parity of $d(b,x)$ equals to the parity of $d(w,x)$. If b is not a leaf, then $b \in C$ and b separates w and v. Thus, $b \in L(T)$. However, if $b \in L_1(T)$, then $b \in C$ and if $b \in L_+(T)$, then $v \in S_+(T)$ and there exists a leaf in C adjacent to v separating v and w, a contradiction. When $d(w,v) \geq 2$, we notice that this case cannot occur since $v, w \in N(u)$ and at least one of these two vertices is adjacent to a leaf in $N(u)$.

Finally, we have the case where exactly one of the two vertices is a leaf, let us say $v \in L(T)$ and since $v \in N(u)$, we have $v \in L_+(T)$. Assume first that $d(v, w) = 2$. Thus, $v, w \notin C$. However, then we again have $b \in C$ such that $b \in N(w) \setminus \{u\}$, either due to the parity of $d(x, b)$ or because b is a leaf. As the last case we have $u = w$ and $v \in L_+(T) \cap N(u)$. If $u \notin C'_a$, then, by parity, u has an adjacent non-leaf vertex in C since T is not a star. On the other hand, if $u \in C'_a$, then if $u \in S_2(T)$ and the two leaves have the same color, there is again a non-leaf neighbor in C. If $u \in S_2(T)$ and the two leaves have different colors, then there is a leaf of the same color, in $N(u)$, as u which is in C. If $u \in S_3(T)$, then u has a non-leaf neighbor in $NS_3(T) \cap C$ or in $S_+(T) \cap C$. If $u \in S_i(T)$ for $i \geq 4$, then u has at least two adjacent leaves which are in C. Hence, w and v are either separated or they have the same color. $\square$

This completes the proof of the theorem. $\square$

The upper bound of Theorem 17 is tight. Consider, for example, a path on eight vertices. Also, the trees presented in Proposition 20 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. The following theorem has been previously considered for total dominating identifying codes (under the term differentiating-total dominating set) in [21].

Theorem 18. *For any tree T of order $n \geq 5$, $\mathrm{sep}(T) \leq n - s(T)$.*

Proof. Let us choose for each support vertex $u \in S(T)$ exactly one adjacent leaf $v \in L(T)$ and say that these vertices form the set S'. Next, we form the separating set $S = V(T) \setminus S'$. Notice that $|S| = n - s(T)$. In the following, we show that S is a separating set in T.

Observe that if $v \notin S$, then v is a leaf and no support vertex has two adjacent leaves which do not belong to S. Thus, vertices which do not belong to S are pairwise separated. Since $S' \subseteq L(T)$, $T[S]$ is a connected induced subgraph of T. Moreover, as $n \geq 5$, we have $|N[w] \cap S| \geq 2$ for each vertex $w \in S$. Thus, vertices in S are separated from vertices which are not in S. Finally, any two vertices $w, w' \in S$ are separated since $|V(T[S])| \geq 3$; hence, each closed neighborhood is unique in $T[S]$. $\square$

The following corollary is a direct consequence of Theorems 17 and 18. Indeed, we have $\mathrm{max\text{-}sep}_{\mathrm{RB}}(T) \leq \min\{n - s(T), (n + s(T))/2\}$.

Corollary 19. *For any tree T of order $n \geq 5$, we have $\mathrm{max\text{-}sep}_{\mathrm{RB}}(T) \leq \frac{2n}{3}$.*

We next show that Corollary 19 (and Theorem 17) is not far from tight.

Proposition 20. *For any $k \geq 1$, there is a tree T of order $n = 5k + 1$ with $\mathrm{max\text{-}sep}_{\mathrm{RB}}(T) = \frac{3(n-1)}{5} = \frac{n+s(T)-1}{2}$.*

Proof. Consider the tree T formed by taking k disjoint copies $P^1, \ldots, P^k$ of a path of order 6 and identifying one endpoint of each path into one single vertex x. We consider the coloring that colors x red, and all other vertices are colored red-blue-red... following the bipartition of the tree. Let $v_1^i, \ldots, v_5^i$ be the vertices of P^i distinct from x, where x is adjacent to v_1^i. In order to separate v_5^i from v_4^i, we need v_3^i in any red-blue separating set. To separate v_4^i from v_3^i, we need either v_2^i or v_5^i. To separate v_3^i from v_2^i, we need either v_1^i or v_4^i. Thus, we need at least three vertices of P^i in any red-blue separating set, which shows that $\mathrm{max\text{-}sep}_{\mathrm{RB}}(T) \geq 3k$. To see that $\mathrm{max\text{-}sep}_{\mathrm{RB}}(T) \leq 3k$, one can see that the set consisting of all vertices v_1^i, v_3^i, v_5^i separates all pairs of vertices. Finally, since $s(T) = k$ and $n - 1 = 5k$, we get that $\frac{n+s(T)-1}{2} = 3k = \frac{3(n-1)}{5}$. $\square$

4. Complexity and approximation for MAX RED-BLUE SEPARATION

We first design an approximation algorithm for MAX RED-BLUE SEPARATION using Theorem 10 and an approximation algorithm from [19]. After that we show that MAX RED-BLUE SEPARATION is NP-hard.

Theorem 21. MAX RED-BLUE SEPARATION *can be approximated within a factor of $O(\ln^2 n)$ on graphs of order n in polynomial time.*

Proof. Let A be a polynomial-time $(2 \ln n + 1)$-approximation algorithm for the SEPARATION problem [19]. For any graph G, let $S(G)$ denote the separating set returned by A on the input graph G. Using Theorem 10, we have

$$|S(G)| \leq (2 \ln n + 1) \cdot \mathrm{sep}(G)$$

$$\leq (2 \ln n + 1) \cdot \lceil \log_2 n \rceil \cdot \mathrm{max\text{-}sep}_{\mathrm{RB}}(G)$$

Hence, algorithm A is a polynomial-time $O(\ln^2 n)$-approximation algorithm for MAX RED-BLUE SEPARATION. $\square$
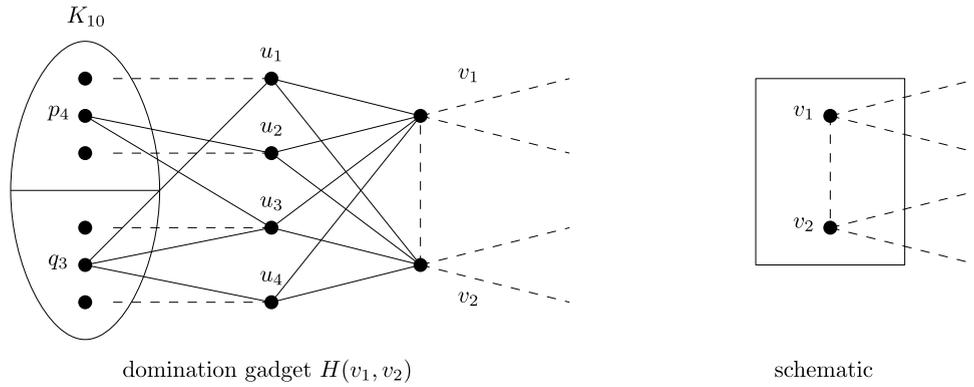
Fig. 11. A domination gadget on vertices v_1 and v_2 and its schematic representation.

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.

Theorem 22. Max Red-Blue Separation *is NP-hard even for graphs of maximum degree 12.*

Proof. We reduce from the following NP-hard version of 3-SAT [31].

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?

Construction   Given an instance σ of 3-SAT-2L with m clauses and n variables, we create an instance (H, k) of Max Red-Blue Separation, where $k = 4m + 9n$. We first explain the construction of a domination gadget.

A *domination gadget* on vertices v_1 and v_2, denoted by $H(v_1, v_2)$, consists of 16 vertices including v_1 and v_2. This gadget will ensure that v_1, v_2 are dominated by the local solution of the gadget (and thus, separated from the rest of the graph). See Fig. 11 for reference. The vertices v_1 and v_2 may be connected to each other or to some other vertices outside the domination gadget as represented by the dashed edges incident to them. Both v_1 and v_2 are also connected to the vertices u_1 to u_4 as shown in the figure. Next we have a clique K_{10} consisting of the vertices $\{p_1, p_2, \ldots, p_6, q_1, q_2, \ldots, q_4\}$. 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.

The graph H consists of one *variable gadget* per variable and one *clause gadget* per clause. The variable gadget for a variable x consists of the graph $H(x^a, x^b)$ and $H(x, \overline{x})$ with additional edges (x^a, x^b), (x^a, x) and $(x^a, \overline{x})$. The clause gadget for a clause $C = (x \vee \overline{y} \vee z)$ (say) is $H(c^a, c^b)$, where c^a is connected to the vertices $c^b, x, \overline{y}$ and z. See Fig. 12 for an illustration. Observe that since each literal can appear at most twice therefore, the maximum degree of H is 12, determined by the q_i's in each domination gadget.

Let us illustrate an example instance for the reduction (see Fig. 13). In this example, we have the formula as $(\overline{w} \vee \overline{x} \vee y) \wedge (x \vee \overline{y} \vee z)$. Thus, we have two clauses $c_1 = \overline{w} \vee \overline{x} \vee y$ and $c_2 = x \vee \overline{y} \vee z$ and four variables w, x, y and z. Corresponding to each clause and each variable, we have a clause gadget and a variable gadget respectively (as described earlier). The domination gadget attached with both the clause and variable gadgets is shown below to increase readability. We have shown one possible coloring in this instance with a large solution size.

Claim 22.1. $\text{max-sep}_{\text{RB}}(H) = \text{sep}(H)$.

Proof of claim. It is clear that $\text{max-sep}_{\text{RB}}(H) \leq \text{sep}(H)$, since any set that separates *all* pairs of vertices also separates all red-blue pairs, for any possible coloring.

We will now prove that $\text{sep}(H) \leq \text{max-sep}_{\text{RB}}(H)$ by providing a specific coloring, and showing that for this coloring, any (optimal) solution also separates all pairs of vertices.

Consider a coloring c of H, where for each domination gadget $H(v_1, v_2)$, all p_i's are colored blue and all q_j's are colored red, or vice versa. For each variable x, the vertices x^a and x^b are given different colors. For each clause C, the vertices c^a and c^b are also colored differently. The rest of the vertices are colored arbitrarily.
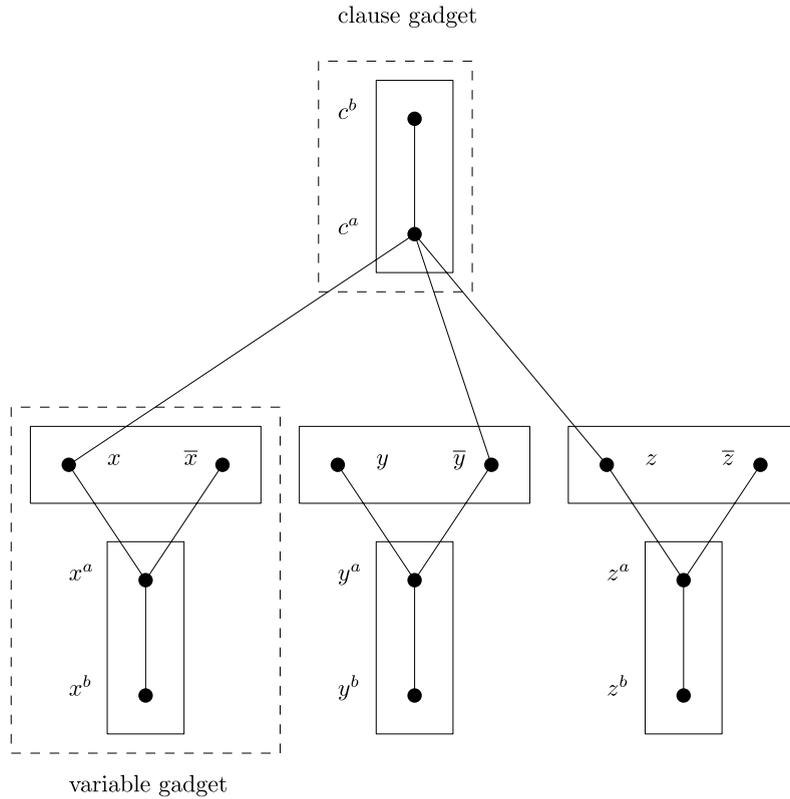
Fig. 12. Variable and clause gadgets.

Consider a (minimum) red-blue separating set S of H with respect to the coloring c. For each domination gadget $H(v_1, v_2)$, the vertices u_1 to u_4 must be included in S. This is because for each $u_h \in H(v_1, v_2)$, there exist $p_i, q_j \in H(v_1, v_2)$ such that $N[p_i] \triangle N[q_j] = \{u_h\}$, where $\triangle$ denotes the symmetric difference between two sets. Since p_i and q_j are given different colors, u_h is the only vertex which separates them and has to be included in S.

Observe that, for each variable x, the vertices u_1 to u_4 of the domination gadget $H(x^a, x^b)$ separate the vertices x^a and x^b from the rest of H. Therefore, they only need to be separated between themselves. The same reasoning holds for the vertices x and $\overline{x}$, for each variable x, as well as for the vertices c^a and c^b, for each clause C.

For each variable x, at least one of the vertices x or $\overline{x}$ must be included in S in order to separate the vertices x^a and x^b as they are colored differently. This also separates vertices x and $\overline{x}$, if they are assigned different colors. For each clause $C = (x \vee \overline{y} \vee z)$ (say), at least one of the vertices $x, \overline{y}$ or z needs to be included in S (if not already included) in order to separate the vertices c^a and c^b as they are also colored differently. The set S must contain a subset of vertices from the set $\{x, \overline{x} \mid x \in X\}$ which separates the vertices y^a and y^b for all variables y, and the vertices c^a and c^b for all clauses C.

Since the size of set S is minimum, it is a minimum red-blue separating set of H with respect to the coloring c. We will now show that S is also a separating set of (uncolored) H. Observe that for each domination gadget $H(v_1, v_2)$, the vertices from u_1 to u_4 (which are in S) separate all (uncolored) p_i's and q_j's from the rest of the vertices in H. This is because the intersection of the set $\{u_1, u_2, u_3, u_4\}$ and the closed neighborhood of each p_i or q_j is unique. The vertices u_1 to u_4 are also separated from the remaining vertices.

Also, since the vertices u_1 to u_4 are included in S for all domination gadgets of H, therefore going by the previous explanation we only need to separate the vertex pairs (x^a, x^b), $(x, \overline{x})$ and (c^a, c^b). But the construction of S is such that all these pairs of vertices are already separated by S. Therefore S is also a separating set of H. This proves our claim. ☐

Claim 22.2. *If σ is satisfiable, $\text{sep}(H) = k$ and otherwise, $\text{sep}(H) > k$, where $k = 4m + 9n$.*

Proof of claim. Consider a separating set S of H. For each domination gadget $H(v_1, v_2)$, the vertices u_1 to u_4 must be included in S as these are the only vertices which can separate all the p_i's and q_j's of $H(v_1, v_2)$ and themselves. Also, for each variable x, one of the vertices x or $\overline{x}$ must also be included in S in order to separate x^a and x^b. Thus, a total of $4(m + 2n) + n = 4m + 9n$ vertices are needed in S. This implies that $\text{sep}(H) \geq k = 4m + 9n$.
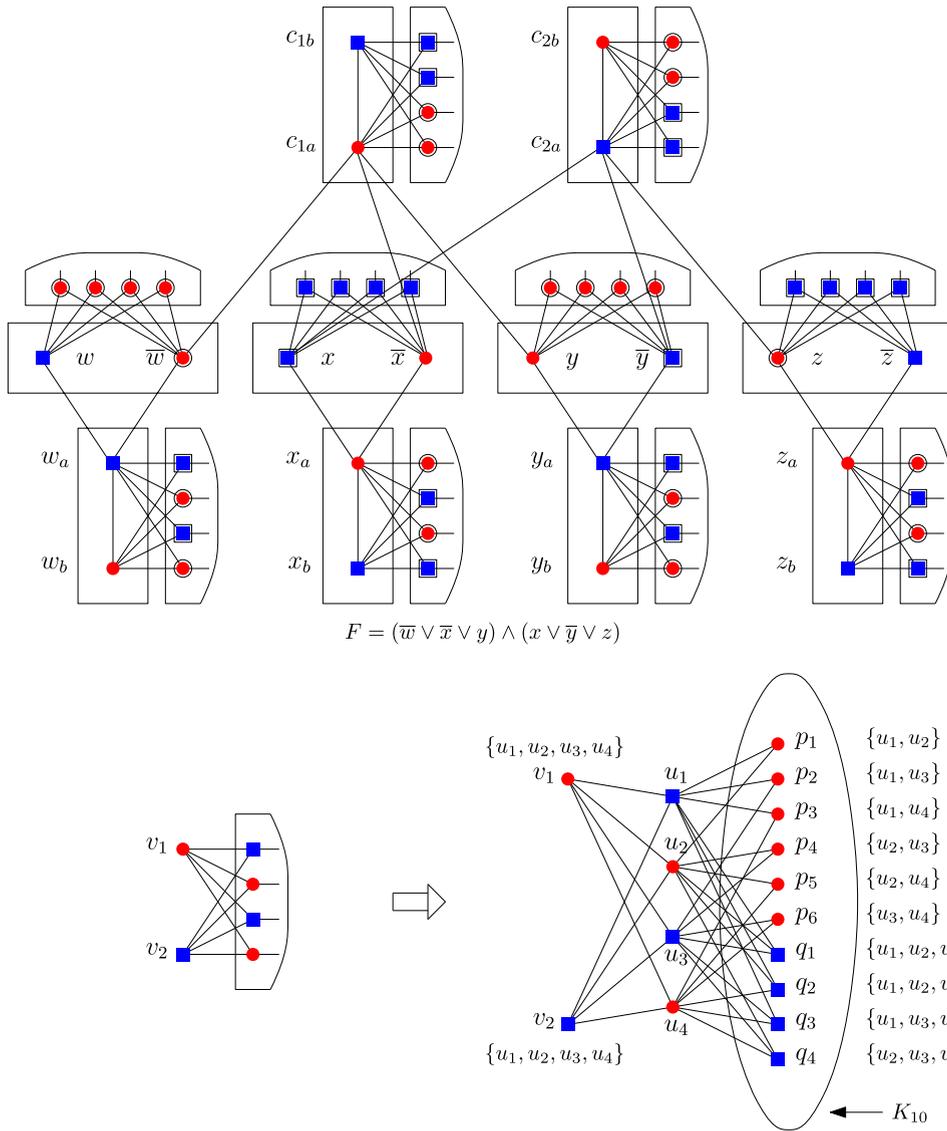
Fig. 13. An example to show the reduction from 3-SAT-2L to MAX RED-BLUE SEPARATION. The square vertices are blue, the round ones are red.

If σ is satisfiable, then with respect to a satisfying assignment, we include for each variable x, the vertex x if it is assigned true, or the vertex $\overline{x}$, if it is assigned false. These vertices also separate c^a and c^b for all clauses C. Therefore, S is a separating set of H and $\text{sep}(H) = k$.

On the contrary, if σ is not satisfiable, then S is not a separating set of H as there exists some clause C for which c^a and c^b are not separated. Since all vertices in S are necessary to be included, any separating set of H is a strict superset of S, and $\text{sep}(H) > k$. □

From Claim 22.1 and Claim 22.2 it follows that σ is satisfiable if and only if $\text{max-sep}_{\text{RB}}(H) \leq k = 4m + 9n$. Since the maximum degree of H is 12 and 3-SAT-2L is NP-hard, MAX RED-BLUE SEPARATION is also NP-hard for graphs with maximum degree 12. □

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. Moreover, max-sep$_{RB}(G)$ 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). However, the upper bound of Theorem 17, which is based on the number of support vertices, is tight for trees. 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 sep$(G) \leq \lceil \log_2(n) \rceil \cdot$ max-sep$_{RB}(G)$. Is it true that sep$(G) \leq 2$ max-sep$_{RB}(G)$? 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?

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Florent Foucaud reports financial support was provided by French National Research Agency.

Data availability

No data was used for the research described in the article.

References

[1] B. Bollobás, A.D. Scott, On separating systems, Eur. J. Comb. 28 (2007) 1068–1071.
[2] J.A. Bondy, Induced subsets, J. Comb. Theory, Ser. B 12 (2) (1972) 201–202.
[3] É. Bonnet, P. Giannopoulos, M. Lampis, On the parameterized complexity of red-blue points separation, J. Comput. Geom. 10 (1) (2019) 181–206.
[4] N. Bousquet, A. Lagoutte, Z. Li, A. Parreau, S. Thomassé, Identifying codes in hereditary classes of graphs and VC-dimension, SIAM J. Discrete Math. 29 (4) (2015) 2047–2064.
[5] G. Călinescu, A. Dumitrescu, H.J. Karloff, P. Wan, Separating points by axis-parallel lines, Int. J. Comput. Geom. Appl. 15 (6) (2005) 575–590.
[6] E. Charbit, I. Charon, G. Cohen, O. Hudry, A. Lobstein, Discriminating codes in bipartite graphs: bounds, extremal cardinalities, complexity, Adv. Math. Commun. 2 (4) (2008) 403–420.
[7] B.S. Chlebus, S.H. Nguyen, On finding optimal discretizations for two attributes, in: Proceedings of the 1st International Conference on Rough Sets and Current Trends in Computing (RSCTC 1998), in: Lecture Notes in Computer Science, vol. 1424, 1998, pp. 537–544.
[8] R. Crowston, G. Gutin, M. Jones, G. Muciaccia, A. Yeo, Parameterizations of test cover with bounded test sizes, Algorithmica 74 (1) (2016) 367–384.
[9] E. Dahlhaus, D.S. Johnson, C.H. Papadimitriou, P.D. Seymour, M. Yannakakis, The complexity of multiterminal cuts, SIAM J. Comput. 23 (4) (1994) 864–894.
[10] S.R. Dev, S. Dey, F. Foucaud, R. Klasing, T. Lehtilä, The Red-Blue Separation problem on graphs, in: Proceedings of the 33rd International Workshop on Combinatorial Algorithms (IWOCA 2022), in: Lecture Notes in Computer Science, vol. 13270, 2022, pp. 285–298.
[11] S. Dey, F. Foucaud, S.C. Nandy, A. Sen, Discriminating codes in geometric setups, in: Proceedings of the 31st International Symposium on Algorithms and Computation (ISAAC 2020), in: Leibniz International Proceedings in Informatics, vol. 181, 2020, 24.
[12] K.M.J. De Bontridder, B.V. Halldórsson, M.M. Halldórsson, C.A.J. Hurkens, J.K. Lenstra, R. Ravi, L. Stougie, Approximation algorithms for the test cover problem, Math. Program., Ser. B 98 (2003) 477–491.
[13] I. Dinur, D. Steurer, Analytical approach to parallel repetition, ACM Symp. Theory Comput. 46 (2014) 624–633.
[14] R.G. Downey, M.R. Fellows, Parameterized Complexity, Springer Verlag, 1999.
[15] P. Erdős, Some combinatorial, geometric and set theoretic problems in measure theory, in: Measure Theory Oberwolfach 1983, Springer, 1984, pp. 321–327.
[16] F. Foucaud, E. Guerrini, M. Kovše, R. Naserasr, A. Parreau, P. Valicov, Extremal graphs for the identifying code problem, Eur. J. Comb. 32 (4) (2011) 628–638.
[17] F. Foucaud, T. Lehtilä, Revisiting and improving upper bounds for identifying codes, SIAM J. Discrete Math. 36 (4) (2022) 2619–2634.
[18] V. Gledel, A. Parreau, Identification of points using disks, Discrete Math. 342 (2019) 256–269.
[19] S. Gravier, R. Klasing, J. Moncel, Hardness results and approximation algorithms for identifying codes and locating-dominating codes in graphs, Algorithmic Oper. Res. 3 (1) (2008) 43–50.
[20] S. Har-Peled, M. Jones, On separating points by lines, Discrete Comput. Geom. 63 (2020) 705–730.
[21] T.W. Haynes, M.A. Henning, J. Howard, Locating and total dominating sets in trees, Discrete Appl. Math. 154 (8) (2006) 1293–1300.
[22] M.A. Henning, A. Yeo, Distinguishing-transversal in hypergraphs and identifying open codes in cubic graphs, Graphs Comb. 30 (4) (2014) 909–932.
[23] M.G. Karpovsky, K. Chakrabarty, L.B. Levitin, On a new class of codes for identifying vertices in graphs, IEEE Trans. Inf. Theory 44 (1998) 599–611.
[24] S. Kratsch, T. Masařík, I. Muzi, M. Pilipczuk, M. Sorge, Optimal discretization is fixed-parameter tractable, in: Proceedings of the 32nd ACM-SIAM Symposium on Discrete Algorithms (SODA 2021), 2021, pp. 1702–1719.
[25] J. Kujala, T. Elomaa, Improved algorithms for univariate discretization of continuous features, in: Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Database (PKDD 2007), in: Lecture Notes in Computer Science, vol. 4702, 2007, pp. 188–199.
[26] A. Lobstein, Watching systems, identifying, locating-dominating and discriminating codes in graphs: a bibliography, https://www.lri.fr/~lobstein/debutBIBidetlocdom.pdf.
[27] D. Marx, Parameterized graph separation problems, Theor. Comput. Sci. 351 (3) (2006) 394–406.
[28] N. Misra, H. Mittal, A. Sethia, Red-blue point separation for points on a circle, in: Proceedings of the 32nd Canadian Conference on Computational Geometry (CCCG 2020), 2020, pp. 266–272.
[29] B.M.E. Moret, H.D. Shapiro, On minimizing a set of tests, SIAM J. Sci. Stat. Comput. 6 (4) (1985) 983–1003.
[30] A. Rényi, On random generating elements of a finite Boolean algebra, Acta Sci. Math. 22 (1961) 75–81.
[31] C.A. Tovey, A simplified NP-complete satisfiability problem, Discrete Appl. Math. 8 (1) (1984) 85–89.
[32] R. Ungrangsi, A. Trachtenberg, D. Starobinski, An implementation of indoor location detection systems based on identifying codes, in: Proceedings of Intelligence in Communication Systems, INTELLCOMM 2004, in: Lecture Notes in Computer Science, vol. 3283, 2004, pp. 175–189.
[33] V.N. Vapnik, A.Ya. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, Theory Probab. Appl. 16 (2) (1971) 264–280.
[34] I.E. Zverovich, V.E. Zverovich, An induced subgraph characterization of domination perfect graphs, J. Graph Theory 20 (3) (1995) 375–395.