

CAHIER DU LAMSADE

268

Juillet 2007

Probabilistic graph-coloring in bipartite and split graphs
(Extended version of Cahier du LAMSADE 218)

Nicolas Bourgeois, Federico Della Croce, Bruno Escoffier,
Cécile Murat, Vangelis Th. Paschos

Probabilistic graph-coloring in bipartite and split graphs

Extended version of Cahier du LAMSADE 218

N. Bourgeois¹ F. Della Croce^{2*} B. Escoffier¹ C. Murat¹ V. Th. Paschos¹

¹ LAMSADE, CNRS UMR 7024 and Université Paris-Dauphine, France

`nbourgeo@clipper.ens.fr`, `{escoffier,murat,paschos}@lamsade.dauphine.fr`

² D.A.I., Politecnico di Torino, Italy, `federico.dellacroce@polito.it`

July 14, 2007

Abstract

We revisit in this paper the stochastic model for minimum graph-coloring introduced in (C. Murat and V. Th. Paschos, *On the probabilistic minimum coloring and minimum k -coloring*, Discrete Applied Mathematics 154, 2006), and study the underlying combinatorial optimization problem (called PROBABILISTIC COLORING) in bipartite and split graphs. We show that the obvious 2-coloring of any connected bipartite graph achieves standard-approximation ratio 2, that when vertex-probabilities are constant PROBABILISTIC COLORING is polynomial and, finally, we propose a polynomial algorithm achieving standard-approximation ratio $8/7$. We also handle the case of split graphs. We show that PROBABILISTIC COLORING is **NP**-hard, even under identical vertex-probabilities, that it is approximable by a polynomial time standard-approximation schema but existence of a fully a polynomial time standard-approximation schema is impossible, even for identical vertex-probabilities, unless **P** = **NP**. We finally study differential-approximation of PROBABILISTIC COLORING in both bipartite and split graphs.

1 Preliminaries

In *minimum graph-coloring problem*, the objective is to color the vertex-set V of a graph $G(V, E)$ with as few colors as possible so that no two adjacent vertices receive the same color. The decision version of this problem, called *graph k -colorability* in [13] and defined as: “given a graph $G(V, E)$ and a positive integer $k \leq |V|$, is G , k -colorable?” was shown to be **NP**-complete in Karp’s seminal paper ([21]) and remains **NP**-complete even restricted to graphs of constant (independent on n) chromatic number at least 3 ([13]). Since adjacent vertices are forbidden to be colored with the same color, a feasible coloring is a partition of V into vertex-sets such that, for any such set, no two of its vertices are mutually adjacent. Such sets are usually called *independent sets*. So, the optimal solution of minimum coloring is a *minimum-cardinality partition into independent sets*. The *chromatic number* of a graph is the smallest number of colors that can feasibly color its vertices.

In this paper, we use the following stochastic model for combinatorial optimization problems. Consider a generic instance I of a combinatorial optimization problem Π . Assume that Π is not

*Part of this research has been performed while the first author was in visit at the LAMSADE on a research position funded by the CNRS

to be necessarily solved on the whole I , but rather on a (unknown a priori) sub-instance $I' \subseteq I$. Suppose that any datum d_i in the data-set describing I has a probability p_i , indicating how d_i is likely to be present in the final sub-instance I' . Consider finally that once instance I' is specified, the solver has no opportunity to solve directly instance I' . In this case, there certainly exist many ways to proceed. Here we deal with a simple and natural way where one computes an initial solution S for Π in the entire instance I and, once I' becomes known, one removes from S those elements of S that do not belong to I' (providing that this deletion results in a feasible solution for I') returning so a solution S' fitting I' . The objective is to determine an initial solution S for I such that, for any sub-instance $I' \subseteq I$ presented for optimization, the solution S' obtained as described just above, respects some pre-defined quality criterion (for example, optimal for I' , or achieving, say, constant approximation ratio, or ...). In what follows we apply this model to the *minimum coloring problem*.

Let us first note that, given a graph $G(V, E)$ and a coloring C for V , in any subgraph $G' = G[V']$ of G induced by some subset $V' \subseteq V$, the coloring C' consisting of the restriction of C to V' (i.e., of moving absent vertices out of the colors in C and of taking into account the non-empty surviving colors) is feasible for G' . In this paper we consider the following version of minimum coloring, called **PROBABILISTIC COLORING** in what follows, dealing with the robustness model drawn just above. We are given a graph $G(V, E)$, an n -vector $\mathbf{Pr} = (p_1, \dots, p_n)$ of vertex-probabilities, any of them representing how likely is that the corresponding vertex will be present in the final instance $I' \subseteq I$ on which the coloring problem will be really solved. We are also given a coloring $C = (S_1, \dots, S_k)$ for V . The objective is to determine a coloring C^* of G minimizing the quantity (called functional) $E(G, C, \mathbf{Pr}) = \sum_{V' \subseteq V} \Pr[V'] |C(V')|$, where $C(V')$ denotes the restriction of C to V' and $\Pr[V'] = \prod_{i \in V'} p_i \prod_{i \in V \setminus V'} (1 - p_i)$. Coloring C will be sometimes called a *a priori solution* or a *a priori coloring*; C^* is then the optimal a priori solution.

Notice that, since there exist 2^n distinct sets V' , any of them inducing a distinct subgraph $G[V']$ of G , both polynomial computation of $E(G, C, \mathbf{Pr})$ and tight combinatorial characterization of the optimal a priori solution are not always obvious or easy to perform.

Set $k' = |C(V')|$, and consider the facts F_j : color S_j has at least a vertex and $\bar{F}_j$: there is no vertex in color S_j . Then, denoting by $\mathbf{1}_{F_j}$ and $\mathbf{1}_{\bar{F}_j}$, respectively, their indicator functions, k' can be written as $k' = \sum_{j=1}^k \mathbf{1}_{F_j} = \sum_{j=1}^k (1 - \mathbf{1}_{\bar{F}_j})$, and $E(G, C, \mathbf{Pr})$ can be written as:

$$\begin{aligned}
E(G, C, \mathbf{Pr}) &= \sum_{V' \subseteq V} \mathbf{Pr}[V'] \left(\sum_{j=1}^k (1 - \mathbf{1}_{\bar{F}_j}) \right) \\
&= \sum_{V' \subseteq V} \mathbf{Pr}[V'] \sum_{j=1}^k 1 - \sum_{V' \subseteq V} \mathbf{Pr}[V'] \sum_{j=1}^k \mathbf{1}_{S_j \cap V' = \emptyset} \\
&= \sum_{j=1}^k \sum_{V' \subseteq V} \mathbf{Pr}[V'] - \sum_{j=1}^k \sum_{V' \subseteq V} \mathbf{Pr}[V'] \mathbf{1}_{S_j \cap V' = \emptyset} = k - \sum_{j=1}^k \prod_{v_i \in S_j} (1 - p_i) \\
&= \sum_{j=1}^k \left(1 - \prod_{v_i \in S_j} (1 - p_i) \right) \tag{1}
\end{aligned}$$

It is easy to see that computation of $E(G, C, \mathbf{Pr})$ needs at most $O(n^2)$ arithmetical operations. Furthermore, (1) provides a closed characterization of the optimal a priori solution C^* for **PROBABILISTIC COLORING**: if the value of an independent set S_j of G is $1 - \prod_{v_i \in S_j} (1 - p_i)$ then *the optimal a priori coloring for G is the partition into independent sets for which the sum of their values is the smallest over all such partitions*. So, **PROBABILISTIC COLORING** can be equivalently

stated as a “deterministic combinatorial optimization problem” as follows: *given a graph $G(V, E)$, and a vertex-probability vector $\mathbf{Pr}$, determine a coloring $C^* = (S_1, S_2, \dots)$ minimizing quantity $f(G, C^*, \mathbf{Pr}) = \sum_{S_j \in C^*} (1 - \prod_{v_i \in S_j} (1 - p_i))$, where $p_i = \mathbf{Pr}[v_i]$ denotes the probability of vertex $v_i \in V$.*

Supposing that binary encoding of any component of $\mathbf{Pr}$ uses a polynomial number of bits, one can immediately see that **PROBABILISTIC COLORING** $\in$ **NPO**, the class of optimization problems the decision counterparts of which are in **NP**.

A priori optimization for combinatorial optimization problems is a quite active research area that has started to be systematically studied by the end of 80’s. In [1, 4, 5, 6, 7, 16, 17, 18, 19], restricted versions of routing and network-design probabilistic minimization problems (in complete graphs) have been studied under the robustness model dealt here (called *a priori optimization*). Recently, in [8], the analysis of the probabilistic minimum travelling salesman problem, originally performed in [4, 16], has been revisited and refined. Furthermore, in [24] the minimum vertex covering problem is dealt on general and bipartite graphs. The same model is also used in [22, 23] for the study of probabilistic maximization problems, namely, the longest path and the maximum independent set, respectively. An early survey about a priori optimization can be found in [2].

PROBABILISTIC COLORING has been originally studied in [25], where complexity and approximation issues have been considered for general graphs and several special configuration graphs such as bipartite graphs, complements of bipartite graphs and others. Dealing with bipartite graphs, the results of [25] left, however, several open questions. For instance, “what is the complexity of **PROBABILISTIC COLORING** in bipartite graphs or even when we further restrict inputs, say in paths, or trees, or cycles, or stars, ...?”, etc. In this paper, we try to give some further answers. We first prove that if vertex probabilities are bounded below by a fixed constant, **PROBABILISTIC COLORING** is polynomial in bipartite graphs. We next prove that, under non-identical vertex-probabilities of any value, **PROBABILISTIC COLORING** is polynomial for stars and for trees with bounded degree and a fixed number of distinct vertex-probabilities and we deduce as a corollary that it is polynomial also for paths with a fixed number of distinct vertex-probabilities. Then, we show that, assuming identical vertex-probabilities, the problem is polynomial for paths, for even and odd cycles and for trees all leaves of which are either at even or at odd levels. We finally focus on split graphs and show that, in such graphs, **PROBABILISTIC COLORING** is **NP**-hard, even assuming identical vertex probabilities.

Let **A** be a polynomial time approximation algorithm for an **NP**-hard minimization graph-problem Π , let $m(G, S)$ be the value of the solution S provided by **A** on an instance G of Π , and $\text{opt}(G)$ be the value of the optimal solution for G (following our notation for **PROBABILISTIC COLORING**, $\text{opt}(G) = f(G, C^*)$). Finally, let $\omega(G)$ be the value of a worst solution of G defined as the value of an optimal solution for $\bar{\Pi}$, the combinatorial problem having the same constraints as Π but instead of minimizing the objective function of Π we wish to maximize it. Given an **NP**-hard problem Π , solving $\bar{\Pi}$ can be very different with respect to solving Π . For instance, if Π is the minimum traveling salesman problem, $\bar{\Pi}$ is the maximum traveling salesman. So, a worst solution for minimum traveling salesman in some graph G is as hard to compute as an optimal one since it corresponds to an optimal solution of maximum traveling salesman in G . For minimum coloring things with worst solution are easier since, following the informal definition of the worst solution given just above, such a solution consists of coloring each vertex of the input graph by a proper color; hence determining a worst solution for minimum coloring is easy.

The standard-approximation ratio $\rho_A(G)$ of an approximation algorithm **A** on G is defined as $\rho_A(G) = m(G, S)/\text{opt}(G)$. An approximation algorithm achieving standard ratio, at most, ρ on any instance G of Π will be called ρ -standard-approximation algorithm.

The differential-approximation ratio $\delta_A(G)$ of an approximation algorithm **A** on G is defined as

$\delta_A(G) = (\omega(G) - m(G, S)) / (\omega(G) - \text{opt}(G))$. By symmetry, an approximation algorithm achieving differential ratio, at least, δ on any instance G of Π will be called δ -differential-approximation algorithm. For both ratios, the closer to 1, the better the approximability quality of an algorithm.

A polynomial time standard- (resp., differential-) approximation schema (PTAS, resp., DPTAS) is a sequence A_ϵ of polynomial time approximation algorithms which when they run with inputs a graph G (instance of Π) and any fixed constant $\epsilon > 0$, they produce a solution S such that $\rho_{A_\epsilon}(G) \leq 1 + \epsilon$ (resp., $\delta_{A_\epsilon}(G) \geq 1 - \epsilon$). A fully polynomial time standard- (resp., differential-) approximation schema (FPTAS, resp., DFPTAS) is a PTAS (resp., DPTAS) where A_ϵ is polynomial not only with the size of the instance but also with $1/\epsilon$.

Dealing with approximation issues, we show that the 2-coloring (U, D) of a (connected) bipartite graph $B(U, D, E)$ achieves standard-approximation ratio 2, under any system of vertex-probabilities. Furthermore, we propose a polynomial algorithm achieving standard-approximation ratio $8/7$ under any system of vertex-probabilities. We also provide a polynomial time standard-approximation schema, under any system of vertex probabilities, for split graphs. Finally, we show that, even under identical vertex-probabilities, PROBABILISTIC COLORING cannot be solved by a fully polynomial time standard-approximation schema. On the other hand, dealing with differential approximation, we show that the differential ratio of the 2-coloring in a bipartite graph can be unbounded below, i.e., it can tend to 0, when we consider any system of probabilities, but it is bounded below (tightly) by $1/2$ when vertex-probabilities are identical. We also prove that $8/7$ -standard-approximation algorithm for PROBABILISTIC COLORING achieves tight differential-approximation ratio $4/5$, while under identical vertex-probabilities this algorithm achieves tight differential-approximation ratio $9/10$. We finally show that under identical vertex-probabilities PROBABILISTIC COLORING admits a DPTAS in split graphs.

In what follows, in Section 2 some general properties of probabilistic colorings that will be used later are given. In Section 3 complexity and (standard- and differential-) approximation issues in general bipartite graphs are presented, while in Section 4 complexity results dealing with particular classes of bipartite and almost bipartite graphs such as trees and cycles are shown. Finally, in Section 5 PROBABILISTIC COLORING in split graphs is tackled.

2 Properties of probabilistic colorings

2.1 General graphs

We give in this section some general properties about probabilistic colorings, applying in any graph, upon which we will be based later in order to achieve our results. In what follows, given an a priori k -coloring $C = (S_1, \dots, S_k)$ we will set for simplicity $f(C) = E(G, C, \mathbf{Pr})$, where $E(G, C, \mathbf{Pr})$ is given by (1), and, for $i = 1, \dots, k$, $f(S_i) = 1 - \prod_{v_j \in S_i} (1 - p_j)$.

Proposition 1. *Let $C = (S_1, \dots, S_k)$ be a k -coloring and assume that colors are numbered so that $f(S_i) \leq f(S_{i+1})$, $i = 1, \dots, k - 1$. Consider a vertex x (of probability p_x) colored with S_i and a vertex y (of probability p_y) colored with S_j , $j > i$, such that $p_x \geq p_y$. If swapping colors of x and y leads to a new feasible coloring C' , then $f(C') \leq f(C)$.*

Proof. Between colorings C and C' the only colors changed are S_i and S_j . Then:

$$f(C') - f(C) = f(S'_i) - f(S_i) + f(S'_j) - f(S_j) \quad (2)$$

Set now

$$\begin{aligned} S'_i &= (S_i \setminus \{x\}) \cup \{y\} \\ S'_j &= (S_j \setminus \{y\}) \cup \{x\} \\ S''_i &= S_i \setminus \{x\} = S'_i \setminus \{y\} \\ S''_j &= S_j \setminus \{y\} = S'_j \setminus \{x\} \end{aligned} \quad (3)$$

Then, using notations of (3), we get:

$$\begin{aligned} f(S'_i) - f(S_i) &= 1 - (1 - p_y) \prod_{v_h \in S''_i} (1 - p_h) - 1 + (1 - p_x) \prod_{v_h \in S''_i} (1 - p_h) \\ &= (p_y - p_x) \prod_{v_h \in S''_i} (1 - p_h) \end{aligned} \quad (4)$$

$$\begin{aligned} f(S'_j) - f(S_j) &= 1 - (1 - p_x) \prod_{v_h \in S''_j} (1 - p_h) - 1 + (1 - p_y) \prod_{v_h \in S''_j} (1 - p_h) \\ &= (p_x - p_y) \prod_{v_h \in S''_j} (1 - p_h) \end{aligned} \quad (5)$$

Using (4) and (5) in (2), we get:

$$f(C') - f(C) = (p_y - p_x) \left(\prod_{v_h \in S''_i} (1 - p_h) - \prod_{v_h \in S''_j} (1 - p_h) \right) \quad (6)$$

Recall that, by hypothesis, we have $f(S_i) \leq f(S_j)$ and $p_x \geq p_y$; consequently, by some easy algebra, we achieve $\prod_{v_h \in S''_i} (1 - p_h) - \prod_{v_h \in S''_j} (1 - p_h) \geq 0$ and, since $p_y - p_x \leq 0$, we conclude that the right-hand-side of (6) is negative, implying that coloring C' is better than C . ■

In the case of identical vertex-probabilities, Proposition 1 has a natural counterpart expressed as follows.

Proposition 2. *Let $C = (S_1, \dots, S_k)$ be a k -coloring and assume that colors are numbered so that $|S_i| \leq |S_{i+1}|$, $i = 1, \dots, k-1$. If it is feasible to inflate a color S_j by “emptying” another color S_i with $i < j$, then the new coloring C' , so created, satisfies $f(C') \leq f(C)$.*

For the proof, simply remark that if $|S_i| \leq |S_j|$, then $f(S_i) \leq f(S_j)$ and apply the same proof as for Proposition 1.

Since, in the proof of Proposition 2, only the cardinalities of the colors intervene, the following corollary-property consequently holds.

Proposition 3. *Let $C = (S_1, \dots, S_k)$ be a k -coloring and assume that colors are numbered so that $|S_i| \leq |S_{i+1}|$, $i = 1, \dots, k-1$. Consider two colors S_i and S_j , $i < j$, and a vertex-set $X \subset S_j$ such that, $|S_i| + |X| \geq |S_j|$. Consider (possibly unfeasible) coloring $C' = (S_1, \dots, S_i \cup X, \dots, S_j \setminus X, \dots, S_k)$. Then, $f(C') \leq f(C)$.*

With very similar arguments and operations as for Proposition 1, the following property, also holds.

Proposition 4. *Let $C = (S_1, \dots, S_k)$ be a k -coloring and assume that colors are numbered so that $f(S_i) \leq f(S_{i+1})$, $i = 1, \dots, k-1$. Consider a vertex x colored with color S_i . If it is feasible to color x with another color S_j , $j > i$, (by keeping colors of the other vertices unchanged), then the new feasible coloring C' satisfies $f(C') \leq f(C)$.*

Propositions 1, or 2, or 3, or 4, describe a process of achieving “locally optima” colorings by local swaps of vertices aiming to “reinforce” the heavier (larger, in the case of identical probabilities) colors. In the sequel, a coloring for which no swaps as the ones described in the statements of these properties is possible, will be called *locally optimal*. Obviously, for a non locally optimal coloring C , there exists a coloring C' , better than C , obtained as described in Propositions 1, or 2, or 3, or 4. Hence, the following Proposition immediately holds.

Proposition 5. *For any non locally optimal coloring, there exists a locally optimal one dominating it.*

Another obvious consequence of Propositions 1, 2, 3 and 4 is that, for any optimal coloring $C^* = (S_1^*, \dots, S_k^*)$ and for any $i = 1, \dots, k$, S_i^* is maximal (for the inclusion) in $G[V \setminus \cup_{j < i} S_j^*]$.

The following proposition gives a measure of the difficulty of solving PROBABILISTIC COLORING.

Proposition 6. *For any graph $G(V, E)$ and for any coloring $C = (S_1, \dots, S_k)$, the following properties are equivalent:*

1. *for any $i \in \{1, \dots, k\}$, S_i is maximal for the inclusion in $G[V \setminus \cup_{j < i} S_j]$;*
2. *there exists some probability vector $\mathbf{Pr}$ such that C is the only optimal coloring in $(G, \mathbf{Pr})$.*

Proof. Implication $2 \Rightarrow 1$ is a straightforward consequence of Proposition 1. Assume that there exists an optimal probabilistic coloring where S_i is not maximal in $G[V \setminus \cup_{j < i} S_j]$. We can strictly decrease its value by coloring with color S_i any vertex belonging to the nonempty set $V \setminus (\cup_{j \leq i} S_j \cup \Gamma(S_i))$ (where $\Gamma(S_i)$ denotes the set of neighbors of the vertices in S_i), a contradiction.

We now show that $1 \Rightarrow 2$. For any $i \in S_k$, set $p_i = 1/n^{3k}$ and observe that, for any $\epsilon \in [0, 1/2]$:

$$\log(1 - \epsilon) = - \sum_{k \geq 1} \frac{\epsilon^k}{k} \geq -\epsilon - \epsilon^2 \frac{\sum_{k \in \mathbb{N}} \epsilon^k}{2} \geq -\epsilon - \frac{\epsilon^2}{2(1 - \epsilon)} \geq -\epsilon - \epsilon^2$$

Furthermore, $e^{-\epsilon} \geq 1 - \epsilon$. Hence, for any k , the following inequalities hold:

$$\begin{aligned} f(S_k) &= 1 - \left(1 - \frac{1}{n^{3k}}\right)^{|S_k|} \leq 1 - e^{-|S_k|(\frac{1}{n^{3k}} + \frac{1}{n^{6k}})} \leq \frac{|S_k|}{n^{3k}} + \frac{|S_k|}{n^{6k}} \\ \sum_{j > k} f(S_j) &\leq 2 \sum_{j > k} \frac{|S_j|}{n^{3j}} \leq \sum_{j > k} \frac{2}{n^{3j-1}} \leq \frac{2}{n^{3k+1}} \end{aligned}$$

Assume that there exists a coloring $\hat{C} = (\hat{S}_1, \dots, \hat{S}_{\hat{k}})$ such that $f(\hat{C}) < f(C)$. Suppose first that $\hat{S}_1 \neq S_1$. According to our hypothesis, $|S_1|$ is maximal. Then, there exists $l > 1$ such that $S_1 \cap \hat{S}_l \neq \emptyset$. Hence:

$$f(\hat{C}) \geq f(\hat{S}_1) + f(\hat{S}_l) \geq \frac{2}{n^3} \geq f(C)$$

that leads to a contradiction. Thus, $\hat{S}_1 = S_1$. Fix i_0 the smallest index such that $\hat{S}_{i_0} \neq S_{i_0}$. According to our hypothesis, S_{i_0} is maximal in $G[V \setminus \cup_{j < i_0} S_j] = G[V \setminus \cup_{j < i_0} \hat{S}_j]$; then, $\exists r > i_0$ such that $S_{i_0} \cap \hat{S}_r \neq \emptyset$. Hence,

$$f(\hat{C}) \geq \sum_{j < i_0} f(\hat{S}_j) + f(\hat{S}_{i_0}) + f(\hat{S}_r) \geq \sum_{j < i_0} f(S_j) + \frac{2}{n^{3i_0}} \geq f(C)$$

once again a contradiction that achieves the proof. ■

Proposition 7. *In any graph of maximum degree Δ , the optimal solution of PROBABILISTIC COLORING contains at most $\Delta + 1$ colors.*

Proof. If an optimal coloring C uses $\Delta + k$ colors, $k > 0$, then, by emptying the least-value color (thing always possible as there are at least $\Delta + 1$ colors) and due to Proposition 4, we achieve a $\Delta + 1$ -coloring feasible for G with value better than the one of C . ■

2.2 General bipartite graphs

We consider in this section general bipartite graphs. We notice that, whenever B is connected its obvious 2-coloring (U, D) is unique. On the other hand, if B is not connected, 2-colorings in the several connected components can be polynomially merged in such a way that the best 2-coloring for the initial graph is produced. Indeed, supposing that a bipartite graph B has k connected components $B_1(U_1, D_1), \dots, B_k(U_k, D_k)$, one simply has to put together the heaviest among the U_i 's and D_i 's, for $i = 1, \dots, k$, in order to form the first color class, say U , leaving the lightest the U_i 's and D_i 's to form the second color class, say D . By the local optimality argument of Proposition 1, the so obtained coloring (U, D) is the best 2-coloring of B . In what follows, when speaking about 2-coloring (U, D) of a bipartite graph $B(U, D)$ we will refer exactly to this coloring.

Consider a bipartite graph $B(U, D, E)$ and, without loss of generality, assume $|U| \geq |D|$. Also, denote by $\alpha(B)$ the cardinality of a maximum independent set of B . Then the following property holds.

Proposition 8. *If $\alpha(B) = |U|$, then 2-coloring $C = (U, D)$ is optimal.*

Proof. Suppose a contrario that C is not optimal, then the optimal coloring C' uses exactly $k \geq 3$ colors and its largest cardinality color S'_1 has cardinality β . Consider the following exhaustive two cases:

$\alpha(B) = \beta$: then, it is sufficient to aggregate all the vertices not belonging to S'_1 into another color, say S'_2 ; this would lead to a – possibly unfeasible – solution C'' which improves upon C' (due to Proposition 5) and whose value coincides with the value of C ;

$\alpha(B) > \beta$: assume adding to color S'_1 exactly $\alpha(B) - \beta$ vertices from the other colors neglecting possible unfeasibilities; the resulting solution C'' dominates C' (due to Proposition 5); but then, the largest cardinality color S''_1 has in solution C'' exactly $\alpha(B)$ vertices; hence, as for case $\alpha(B) = \beta$, the 2-coloring C is feasible, and dominates both C'' and C' . ■

2.3 Trees

In this section, we restrict ourselves to trees and give a sufficient condition for optimality of the natural 2-coloring in such graphs. We first prove an easy lemma that will help us to get our result.

Lemma 1. *For any connected graph $G(V, E)$ with maximum degree Δ , the size of a maximal independent set is at most $((\Delta - 1)n/\Delta) + 1$.*

Proof. Fix a maximal independent set S and partition V into two sets S and $S' = V \setminus S$. Consider a tree of G rooted at some vertex of S . Each vertex of S' has at most $\Delta - 1$ children. Since S is an independent set, each of its vertices that is not the root is a child of a vertex of S' . Hence, $|S| \leq (\Delta - 1)|S'| + 1 = (\Delta - 1)(n - |S|) + 1$, i.e., $|S| \leq ((\Delta - 1)n + 1)/\Delta \leq ((\Delta - 1)n/\Delta) + 1$ as claimed. ■

We are ready now to prove the main property of this section.

Proposition 9. *Consider a tree $T(V, E)$ of maximum degree Δ . If, for any $v_i \in V$, $p_i \geq \Delta \log n/n$, then 2-coloring is optimal.*

Proof. Fix an optimal coloring $C^* = (S_1^*, S_2^*, \dots, S_k^*)$, $k \geq 2$, ordered by decreasing value of S_i^* 's, $i = 1, \dots, k$. By what has been discussed previously, S_1^* is a maximal independent set and, by Lemma 1, $|S_1^*| \leq ((\Delta - 1)n/\Delta) + 1$. Based upon Propositions 1 and 4, the value $f(C^*)$ of C^* is

greater than the value of a (possibly unfeasible) coloring (S'_1, S'_2, S'_3) where S'_1 would contain S_1^* plus $((\Delta - 1)n/\Delta + 1) - |S_1^*|$ additional vertices, S'_2 would contain the rest of the vertices of G but one of minimum probability, and S'_3 would contain only a minimum-probability vertex. In other words based upon the hypotheses of the proposition and the local optimality arguments developed in Section 2.1:

$$f(S^*) \geq 1 - \left(1 - \frac{\Delta \log n}{n}\right)^{\frac{(\Delta-1)n}{\Delta}+1} + 1 - \left(1 - \frac{\Delta \log n}{n}\right)^{\frac{n}{\Delta}-2} + \frac{\Delta \log n}{n} \quad (7)$$

Let us first consider the case where $4 \leq n \leq 7$. Then, by local optimality arguments, the best 3-coloring is worse than a (probably infeasible) 3-coloring where the two lightest colors are singletons. In other words,

$$f(C^*) \geq 1 - \left(1 - \frac{\Delta \log n}{n}\right)^{n-2} + 2\frac{\Delta \log n}{n}$$

An exhaustive study of all the possible pairs (n, Δ) for this case shows that $f(C^*) > 2$.

For $8 \leq n \leq 11$ also, an exhaustive study of (7) leads to the same result.

Assume now $n \geq 11$. For any $\epsilon \in [0, 1[$, $\log(1 - \epsilon) \leq -\epsilon$. Applying this inequality with $\epsilon = \Delta \log n/n$, we get from (7):

$$\left(1 - \frac{\Delta \log n}{n}\right)^{\frac{(\Delta-1)n}{\Delta}+1} = \exp \left\{ \left(\frac{(\Delta-1)n}{\Delta} + 1 \right) \log \left(1 - \frac{\Delta \log n}{n} \right) \right\} \leq \frac{1}{n^{\Delta-1}} \quad (8)$$

$$\left(1 - \frac{\Delta \log n}{n}\right)^{\frac{n}{\Delta}-2} = \exp \left\{ \left(\frac{n}{\Delta} - 2 \right) \log \left(1 - \frac{\Delta \log n}{n} \right) \right\} \leq \frac{n^{\frac{2\Delta}{n}}}{n} \quad (9)$$

By convexity, $e^{2\epsilon} \leq \epsilon e^2 + (1 - \epsilon)$, $\forall \epsilon \in [0, 1]$. Hence:

$$\frac{n^{\frac{2\Delta}{n}}}{n} \leq \frac{\Delta(e^2 - 1) \log n}{n^2} + \frac{1}{n} \quad (10)$$

Expressions (7), (8), (9) and (10) derive:

$$f(S^*) \geq 2 + \frac{\Delta \log n}{n} \left(1 - \frac{(e^2 - 1)}{n} - \frac{2}{\Delta \log n} \right) > 2$$

while the natural 2-coloring of T has value less than 2 and the proof of the property is completed. ■

Corollary 1. *If $\forall v_i \in V$, $p_i \geq 2 \log n/n$, then 2-coloring is optimal on paths.*

The bound given in Corollary 1 is the best possible for paths as the following proposition shows. Before introducing it recall that, by Proposition 7, an optimal coloring on paths uses either 2 or 3 colors. Remark, furthermore, that if the optimal coloring on a path uses three colors, then determining the heaviest one thoroughly determines the whole coloring. Indeed, removal of the heaviest color from the path, produces a set of connected sub-paths. Then, proceeding as described in the beginning of Section 2.2 optimally completes the coloring.

Proposition 10. *If we only assume that, for any $v_i \in V$ $p_i \geq \lambda \log n/n$, with $\lambda < 2$ then there exist paths where the optimal coloring uses three colors and the least-value color is arbitrarily large.*

Proof. Fix $m = n^{1-\lambda/2}/(2\lambda \log n)$. Consider a path of length $n - m$, where odd vertices have probability 1, while even vertices have probability $\alpha = \lambda \log n/n$. Now select at random m vertices of probability 1 and, for each of them, insert just behind it a new vertex with presence-probability α . The size of the so-obtained path is n .

Obviously, the natural 2-coloring of this path has value 2, while the 3-coloring C induced by setting $S_1^* = \{i : p_i = 1\}$ (see the remark just before the statement of the proposition) has value $f(C) = 3 - (1 - \alpha)^{(n-m)/2} - (1 - \alpha)^m$. Since $\log(1 - \alpha) = -\alpha + o(\alpha) = -\lambda \log n/n + o(\log n/n)$, we get:

$$(1 - \alpha)^{\frac{n-m}{2}} = e^{\log(1-\alpha) \frac{n-m}{2}} = e^{-\frac{\lambda \log n}{2} + o(\log n)} = \frac{1 + o(1)}{n^{\frac{\lambda}{2}}} \quad (11)$$

$$(1 - \alpha)^m = e^{-\frac{\lambda m \log n}{n} + o(\frac{m \log n}{n})} = e^{-\frac{\lambda}{2}(1+o(1))} = 1 - \frac{1}{2n^{\frac{\lambda}{2}}} + o\left(\frac{1}{2n^{\frac{\lambda}{2}}}\right) \quad (12)$$

Combination of (11) and (12) leads, after some easy algebra, to $f(C) < 2$ and completes the proof. ■

Remark 1. Similar, though more complicated constructions show that if we only assume that, $\forall v_i \in V$, $p_i \geq \lambda \log n/n$, with $\lambda < \Delta$, then there exist trees of maximum degree Δ where the optimal coloring uses three colors and the least-value color is arbitrarily large. ■

Proposition 11. *On general bipartite graphs, even on trees, it is impossible to find any lower bound $\epsilon_n \rightarrow 0$, such that, if $\forall v_i \in V$, $p_i \geq \epsilon_n$, then 2-coloring is optimal.*

Proof. Suppose, a contrario, there exists such an ϵ_n and consider the following tree:

- $V = (A, B) = \{a_i\}_{i \leq n} \cup \{b_i\}_{i \leq n}$;
- $E = \{\{a_i, b_1\}_{i \leq n}\} \cup \{\{b_i, a_1\}_{i \leq n}\}$;
- $\forall i < n$, $p(a_i) = p(b_i) = \epsilon_n$ and $p(a_n) = p(b_n) = 1$.

Clearly, $f(A, B) = 2$ while $f(\{a_i, b_i\}_{i \geq 2}, \{a_1\}, \{b_1\}) = 1 + 2\epsilon_n$. ■

3 PROBABILISTIC COLORING in general bipartite graphs

We first prove a complexity result claiming that PROBABILISTIC COLORING is polynomial in bipartite graphs when vertex-probabilities are bounded below by any fixed constant.

Proposition 12. *If vertex-probabilities are bounded below by a fixed constant, then PROBABILISTIC COLORING is polynomial in bipartite graphs.*

Proof. Consider a bipartite graph $B(U, D)$ of order n , and denote by $C = (U, D)$, the natural 2-coloring of B . Denote by $C' = (S_1, S_2, \dots, S_k)$ any k -coloring (a fortiori an optimal one). As previously, we denote by $f(C)$ and $f(C')$ the values of C and C' , respectively. The proof of the proposition is essentially based upon the following claims.

Claim 1. The size k of C' verifies $k \leq \beta = \frac{2}{p_{\min}}$.

Proof of Claim 1. For any color S_i , $f(S_i) \geq p_{\min}$; hence, $f(C') \geq kp_{\min}$. Since $f(C) \leq 2$, the result follows. □

Claim 2. At most two colors of C' can have a size greater than $\alpha = \ln(3)/-\ln(1 - p_{\min})$.

Proof of Claim 2. A color S_i of size greater than α satisfies $f(S_i) > 1 - (1 - p_{\min})^\alpha = 1 - 1/3 = 2/3$. Since $f(C) \leq 2$, the result follows. $\square$

Henceforth, one can solve the problem with the following algorithm:

- consider all the possibilities of putting at most α vertices in each (eventually empty) color S_i , for $i = 3, \dots, \beta$ and, for each of these possibilities, color optimally the remaining (uncolored) vertices;
- return the best solution computed in the previous step.

This algorithm is optimal thanks to Claims 1 and 2. There exist at most $O(n^\alpha)$ choices for each color S_i , $i \geq 3$, hence $O(n^{\alpha(\beta-2)})$ choices for colors $S_3, \dots, S_\beta$. Since α and β are fixed constants, and since one can optimally 2-color a bipartite graph in polynomial time, the result of the proposition follows. $\blacksquare$

3.1 Standard approximation of PROBABILISTIC COLORING in bipartite graphs

We now settle the general case for stochastic bipartite graphs where there exist vertex-probabilities arbitrarily small (i.e., depending on n).

Proposition 13. *In any bipartite graph $B(U, D, E)$, its 2-coloring $C = (U, D)$ achieves standard-approximation ratio bounded by 2. This bound is tight even on paths.*

Proof. Consider a bipartite graph $B(U, D, E)$. A trivial lower bound on the optimal solution cost (due to Proposition 1) is given by the unfeasible 1-coloring $U \cup D$ with all the vertices having the same color. Hence, denoting by C^* , an optimal coloring of B , we have:

$$f(U \cup D) \leq f(C^*) \quad (13)$$

Assume that $f(U) \leq f(D)$. Then, since $D \subseteq U \cup D$, $f(D) \leq f(U \cup D)$. Therefore, using (13) $f(C) = f(U) + f(D) \leq 2f(D) \leq 2f(U \cup D) \leq 2f(C^*)$, qed.

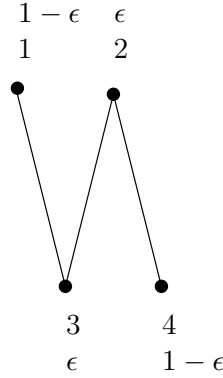


Figure 1: Ratio 2 is tight for the 2-coloring of a bipartite graph.

For tightness, consider the 4-vertex path of Figure 1. The 2-coloring has value $2 - 2\epsilon + 2\epsilon^2$, while the 3-coloring $\{1, 4\}, \{2\}, \{3\}$ has value $1 + 2\epsilon - \epsilon^2$. For $\epsilon \rightarrow 0$, the latter is the optimal solution and the standard-approximation ratio of the two coloring tends to 2. $\blacksquare$

We now improve the previous result, thanks to the following algorithm, denoted by 3-COLOR in what follows:

1. compute and store the 2-coloring $C_0 = (U, D)$;

2. compute a maximum weight independent set S of B , where the weight of a vertex of probability p is $-\log(1-p)$ (and the weight of an independent set is the sum of the weights of its vertices);
3. output the best coloring (break ties at random) among C_0 and $C_1 = (S, U \setminus S, D \setminus S)$ (in the case where C_0 and C_1 are equally good randomly output one of them).

Obviously, 3-COLOR is polynomial, since computation of a maximum weight independent set can be performed in polynomial time in bipartite graphs ([10]).

Proposition 14. *Algorithm 3-COLOR achieves a standard-approximation ratio bounded above by $8/7$ in bipartite graphs.*

Proof. Note first that the independent set found in step 2 maximizes the quantity:

$$\sum_{v_i \in S} -\log(1-p_i) = \log \left(\frac{1}{\prod_{v_i \in S} (1-p_i)} \right)$$

and hence, it maximizes $f(S)$.

Consider an optimal solution $C^* = (S_1^*, S_2^*, \dots, S_k^*)$, and assume w.l.o.g that $f(S_1^*) \geq f(S_2^*) \geq \dots \geq f(S_k^*)$. Set $\gamma = \sqrt{1-f(U \cup D)} = \sqrt{\prod_{v_i \in B} p_i}$, $\alpha = \sqrt{1-f(S)} = \sqrt{\prod_{v_i \in S} p_i}$ and $\beta = \sqrt{\prod_{v_i \notin S} p_i} = \gamma/\alpha$.

Since S is a maximum weight independent set, $f(S) \geq \max\{f(U), f(D)\} \geq 1 - \gamma$, hence $\alpha \leq \beta$.

Based upon Proposition 2, the worst case for C_0 is reached when it is “highly” non locally optimal, i.e., when $f(U) = f(D)$. In other words:

$$f(C_0) = f(U) + f(D) \leq 2(1 - \gamma) \quad (14)$$

By exactly the same reasoning:

$$f(C_1) = f(S) + f(U \setminus S) + f(D \setminus S) \leq 1 - \alpha^2 + 2(1 - \beta) \quad (15)$$

Remark also that $f(S_1^*) \leq f(S_1) = 1 - \alpha^2$. If this inequality is strict, then, applying Proposition 2, one, by (virtually) emptying some colors S_j^* , $j > 1$, can obtain a (probably infeasible) coloring C' such that $f(C') \leq f(C^*)$ and the heaviest color of C' has value $1 - \alpha^2$; in other words:

$$f(C^*) \geq f(C') \geq 1 - \alpha^2 + 1 - \beta^2 \quad (16)$$

Using (14), (15) and (16), we get (omitting, for simplicity, to index ρ by 3-COLOR):

$$\rho(B) = \min \left\{ \frac{f(C_0)}{f(C^*)}, \frac{f(C_1)}{f(C^*)} \right\} \leq \min \left\{ \frac{2(1 - \alpha\beta)}{2 - \alpha^2 - \beta^2}, \frac{3 - \alpha^2 - 2\beta}{2 - \alpha^2 - \beta^2} \right\} \quad (17)$$

Recall that $0 \leq \alpha \leq \beta < 1$. We now show that function $f_1(x) = 2(1 - \alpha x)/(2 - x^2 - \alpha^2)$ is decreasing with x in $[\alpha, 1]$, while function $f_2(x) = (3 - \alpha^2 - 2x)/(2 - x^2 - \alpha^2)$ is increasing with x in the same interval. Indeed, by elementary algebra, one immediately gets:

$$f_1'(x) = \frac{-2\alpha(x - \alpha) \left(x - \left(\frac{2 - \alpha^2}{\alpha} \right) \right)}{(2 - x^2 - \alpha^2)^2} \quad (18)$$

$$f_2'(x) = \frac{-2(x - 1) \left(x - (2 - \alpha^2) \right)}{(2 - x^2 - \alpha^2)^2} \quad (19)$$

In (18), $(2 - \alpha^2)/\alpha \geq 1$; so, $f'_1(x)$ is nonnegative for $x \in [\alpha, 1[$ and, consequently, f_1 is non-decreasing with x in this interval. On the other hand, in (19), since $x < 1$ and $\alpha < 1$, $x - 1 < 0$ and $x - (2 - \alpha^2) < 0$. So, $f'_2(x)$ is negative for $x \in [\alpha, 1[$ and, consequently, f_2 is decreasing with x in this interval.

In all, quantity $\min\{f_1(\beta), f_2(\beta)\}$ achieves its maximum value for β verifying $f_1(\beta) = f_2(\beta)$, or when $2(1 - \alpha\beta) = 3 - \alpha^2 - 2\beta$, i.e., when $\beta = (1 + \alpha)/2$. In this case (17) becomes (for $\alpha \leq 1$):

$$\rho(B) \leq \frac{2(1 - (\frac{1+\alpha}{2})\alpha)}{2 - (\frac{1+\alpha}{2})^2 - \alpha^2} = \frac{8 - 4\alpha - 4\alpha^2}{7 - 2\alpha - 5\alpha^2} \leq \frac{8}{7}$$

and the proposition is proved. ■

Notice that this bound is tight, even in paths. For instance, consider the case where the graph is a path on 7 vertices, $v_1, \dots, v_7$, and the vertex-probabilities are $p_1 = p_4 = 1 - \varepsilon$, $p_2 = p_3 = \varepsilon$ and $p_5 = p_6 = p_7 = 1/2$, for some $\varepsilon > 0$ arbitrarily small. Then, $f(C_0) \geq 2 - 2\varepsilon$. On the other hand, $S = \{v_1, v_4, v_7\}$ is a maximum weight independent set. Hence, 3-COLOR can output the 3-coloring $(S, \{v_2, v_5\}, \{v_3, v_6\})$ of value greater than $1 - \varepsilon^2 + 2 \times 1/2 = 2 - \varepsilon^2$, while the solution given by $S_1^* = \{v_1, v_4, v_6\}$, $S_2^* = \{v_2, v_5, v_7\}$ and $S_3^* = \{v_3\}$ has value at most $1 + (1 - (1 - \varepsilon)/4) + \varepsilon \leq 7/4 + 5\varepsilon/4$.

It seems also natural to wonder whether the ratio achieved by 3-COLOR is improvable or not when dealing with the identical probability case. Unfortunately, there exist arbitrarily large instances in which, if 3-COLOR is allowed to arbitrarily choose some maximum independent set, then it achieves standard-approximation ratio asymptotically equal to $8/7$. For instance, fix an $n \in \mathbb{N}$ and consider the following bipartite graph $B(U, D, E)$ consisting of:

- an independent set S_1 on $2n^2$ vertices; n^2 of them, denoted by $v_U^1, \dots, v_U^{n^2}$ belong to U and the n^2 remaining ones, denoted by $v_D^1, \dots, v_D^{n^2}$ belong to D ;
- n paths $P_1, \dots, P_n$ of size 4 (i.e. on 3 edges); set, for $i = 1, \dots, n$, $P_i = (r_i^1, r_i^2, r_i^3, r_i^4)$, where $r_i^1, r_i^3 \in U$ and $r_i^2, r_i^4 \in D$; S_1 and the n paths P_i are completely disjoint;
- two vertices $u \in U$ and $v \in D$; u is linked to all the vertices of D and v to all the vertices of U ;
- for any $v_i \in U \cup D$, $p_i = p = \ln 2/n$.

The graph so-constructed is balanced (i.e., $|U| = |D|$) and has size $2n^2 + 4n + 2$. Figure 2 shows such a graph for $n = 2$.

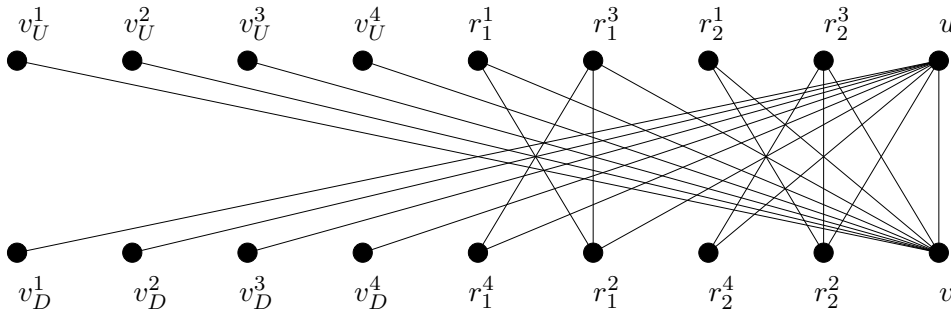


Figure 2: An $8/7$ instance for 3-COLOR with $n = 2$.

Apply algorithm 3-COLOR to the so-constructed graph B . Coloring $C_0 = (U, D)$ has value:

$$f(C_0) = 2 \left(1 - (1 - p)^{n^2 + 2n + 1} \right) \quad (20)$$

On the other hand, one can see that several maximum independent sets of B exist, each consisting of the $2n^2$ vertices of S_1 plus two vertices per any of the n paths P_i , $i = 1, \dots, n$. Assume that the maximum independent set computed in step 2 of algorithm 3-COLOR is $S = S_1 \cup_{i=1, \dots, n} \{r_i^1, r_i^4\}$. In this case, $|S| = 2n^2 + 2n$, and $|U \setminus S| = |D \setminus S| = n + 1$; hence, the value of the coloring $C_1 = (S, U \setminus S, D \setminus S)$ examined in step 3 has value:

$$f(C_1) = 1 - (1 - p)^{2n^2+2n} + 2(1 - (1 - p)^{n+1}) \quad (21)$$

Finally, consider the coloring $\hat{C} = (\hat{S}_1, \hat{S}_2, \hat{S}_3)$ of B where:

- $\hat{S}_1 = S_1 \cup_{i=1, \dots, n} \{p_i^1, p_i^3\}$;
- $\hat{S}_2 = \{v\} \cup_{i=1, \dots, n} \{p_i^2, p_i^4\}$;
- $\hat{S}_3 = \{u\}$.

Obviously:

$$f(\hat{C}) = 1 - (1 - p)^{2n^2+2n} + 1 - (1 - p)^{2n+1} + p \quad (22)$$

One can easily see that, for $n \rightarrow \infty$ and for $p = \ln 2/n$, (20), (21) and (22) give respectively: $f(C_0) \rightarrow 2$, $f(C_1) \rightarrow 2$ and $f(C^*) \leq f(\hat{C}) \rightarrow 7/4$.

Notice that the tightness of the bound $8/7$ can be shown (under identical probabilities and under the same hypothesis on the way it works) for algorithm 3-COLOR also on trees by means of the following instance T presented in Figure 3, for $n = 2$. There, the root-vertex a_0 of T has $n^2 + 1$ children $a_1, \dots, a_{n^2}, b_0$. Vertices $\{a_1, \dots, a_{n^2}\}$ have no children, while vertex b_0 has $n^2 + 1$ children $b_1, \dots, b_{n^2}, c_0$. Again, vertices $b_1, \dots, b_{n^2}$ have no children, while vertex c_0 has $2n$ children $c_1, \dots, c_{2n}$. Finally, vertex c_{2n} has no children while any vertex c_i , with $i = 1, \dots, 2n - 1$, has a single child-vertex d_i . Assume that all the vertices have probability $\ln 2/n$.

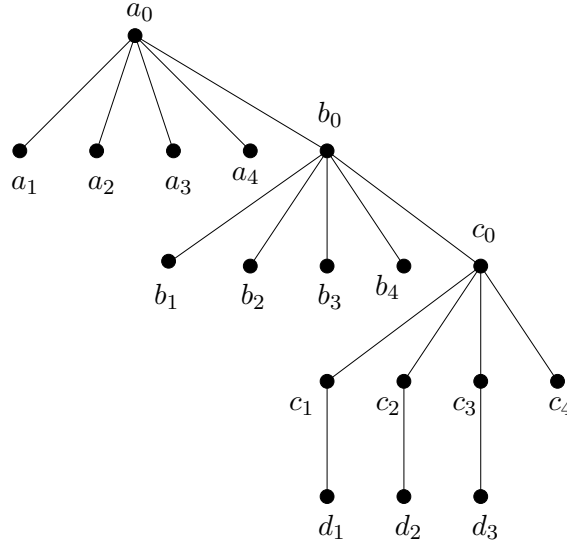


Figure 3: Lower bound $8/7$ is attained for 3-COLOR even in trees ($n = 4$).

The tree T so-constructed gives, as in the previous example, a balanced bipartite graph (i.e., $|U| = |D|$) and has size $2n^2 + 4n + 2$. Apply algorithm 3-COLOR to T and set $C'_0 = (U, D)$. Assume that the maximum independent set computed in step 2 of algorithm 3-COLOR is $S' =$

$\{a_1, \dots, a_{n^2}, b_1, \dots, b_{n^2}, c_{n+1}, \dots, c_{2n}, d_1, \dots, d_n\}$. Then the coloring $C' = (S', U \setminus S', D \setminus S')$ is also examined in step 3. Besides, coloring $\hat{C}' = (\hat{S}'_1, \hat{S}'_2, \hat{S}'_3)$ with

$$\begin{aligned}\hat{S}'_1 &= \{a_1, \dots, a_{n^2}, b_1, \dots, b_{n^2}, c_1, \dots, c_{2n}\} \\ \hat{S}'_2 &= \{a_0, c_0, d_1, \dots, d_{2n-1}\} \\ \hat{S}'_3 &= \{b_0\}\end{aligned}$$

is the best one. Some easy algebra derives then ratio $8/7$ for 3-COLOR when running on the considered tree.

Algorithm 3-COLOR is a simplified version of the following algorithm, called MASTER-SLAVE¹:

1. compute and store the natural 2-coloring (U, D) ;
2. set $B_1(U_1, D_1) = B(U, D)$;
3. set $i = 1$;
4. repeat the following steps until possible:
 - (a) compute some maximum independent set S_i of B_i ;
 - (b) set $(U_{i+1}, D_{i+1}) = (U_i \setminus S_i, D_i \setminus S_i)$;
 - (c) compute and store coloring $(S_1, \dots, S_i, U_{i+1}, D_{i+1})$;
5. compute and store coloring $(S_1, S_2, \dots)$, where S_i 's are the independent sets computed during the executions of step 4a;
6. output C , the best among the colorings computed in steps 1, 4c and 5.

This algorithm, obviously provides solutions that are at least as good as the ones provided by 3-COLOR. Therefore its standard-approximation ratio for PROBABILISTIC COLORING is at most $8/7$. We show that it cannot do better (always as it is, i.e., allowing it to arbitrarily choose the consecutive independent set S_i in step 4a), even in the case of identical probabilities. Indeed, consider the counter-example after the proof of Proposition 14. After computation of S the surviving graph consists of the vertex-set $\cup_{i=1, \dots, n} \{r_i^2, r_i^3\} \cup \{u, v\}$. In this graph, the maximum independent set is of size $n + 1$ (say the vertices of the surviving subset of U). In other words, colorings C_i computed, for $i \geq 2$ by MASTER-SLAVE are the same as coloring C_1 computed by 3-COLOR.

Notice, however, that the counter-example on trees, presented just above, does not work if algorithm MASTER-SLAVE is applied instead of 3-COLOR.

Algorithm 3-COLOR colors any bipartite graph with three colors. We show in the sequel that it is “optimal” in the sense that no polynomial time algorithm that 3-colors the vertices of a bipartite graph can guarantee a better standard-approximation ratio.

Proposition 15. *The problem of finding the best 3-coloring is NP-hard in bipartite graphs, even under identical probabilities. Moreover, it is not $(8/7 - \varepsilon)$ -approximable, for any $\varepsilon > 0$, unless $P = NP$.*

¹This kind of algorithms approximately solving a “master” problem (COLORING in this case) by running a subroutine for a maximization “slave” problem (MAX INDEPENDENT SET here) appears for first time in [20]; appellation “master-slave” for these algorithms is due to [26].

Proof. The reduction is from the precoloring extension problem on bipartite graphs that is shown to be **NP**-complete in [9] and is defined as follows: given a bipartite graph $B(U, D, E)$ and three vertices v_1, v_2 and v_3 in U , we wish to determine if there exists a 3-coloring (S_1, S_2, S_3) of B such that $v_i \in S_i, i = 1, 2, 3$.

Consider a bipartite graph $B(U, D, E)$ instance the precoloring extension problem. Set $U_1 = U \setminus \{v_1, v_2, v_3\}$ and let $\varepsilon > 0$. We construct the following bipartite graph $B'(U', D', E')$ as instance of our problem:

- start with the bipartite graph B ;
- replace vertex v_1 by a set S_U^1 of k_1 copies of v_1 and vertex v_2 by a set S_U^2 of k_2 copies of v_2 ; for commodity define $S_U^3 = \{v_3\}$;
- a vertex in S_U^i is linked in B' to a vertex u if and only if v_i is linked to u in B ;
- add to D' a set S_D^1 of k_1 vertices, a set S_D^2 of k_2 vertices, and a singleton S_D^3 ;
- add all the edges between S_U^i and S_D^j , for $j \neq i$;
- fix $p = 1/n^2$ and set $k_1 = \lceil \ln(\varepsilon)/\ln(1-p) \rceil$, and $k_2 = \lceil \ln(1/2)/\ln(1-p) \rceil$.

Note that the whole transformation of B into B' is polynomial as $k_1 = \theta(n^2)$ and $k_2 = \theta(n^2)$.

We show that if the answer to the precoloring extension problem is *yes*, then $\text{opt}(B') \leq 7/4 + O(\varepsilon)$ while, if this answer is *no*, then $\text{opt}(B') \geq 2 - O(\varepsilon)$.

First, let us consider that the answer to the precoloring extension problem is *yes*, i.e., we have a 3-coloring (S_1, S_2, S_3) of B with $v_i \in S_i$. Then consider the 3-coloring of B' where one considers the same colors for vertices in U_1 and D , and colors vertices in S_U^i and S_D^i with color i . This is obviously a proper 3-coloring for B' . Moreover, by local optimality arguments of Proposition 2, the worst case is achieved when all the vertices in U_1 and D are in the third color (since the sizes of the two first ones are much bigger). Hence: $\text{opt}(B') \leq 3 - (1-p)^{2k_1} - (1-p)^{2k_2} - (1-p)^{n-1} \leq 7/4 + O(1/n^2)$, as $(1-p)^{2k_1} \geq 0$, $(1-p)^{2k_2} \geq 1/4 - O(1/n^2)$, and $(1-p)^{n-1} \geq 1 - O(1/n^2)$.

Now, suppose that the answer to the precoloring extension problem is *no*. We show that $\text{opt}(B')$ is at least (nearly) equal to 2. Remark first that, for any i , all the vertices in S_U^i have the same neighborhood (this is also true for S_D^i). Hence, there exists an optimal solution in which all these vertices will receive the same color (it easily follows from local optimality arguments). We reason with respect to this optimal solution of B' . Now, consider the following cases:

- if S_U^1 and S_D^1 are not in the same color, then the value of the coloring is at least $2 - 2\varepsilon$; indeed, the value of S_U^1 is already $1 - (1-p)^{k_1} \geq 1 - \varepsilon$ (the same holds for S_D^1);
- if S_U^1 and S_D^1 are in the same color S_1 , then S_U^2 and S_D^2 are not in S_1 ; if they are not in the same color, then the value of the coloring is at least $1 - \varepsilon^2 + 1/2 + 1/2 = 2 - \varepsilon^2$; indeed, the value of S_1 is at least $1 - (1-p)^{2k_1} \geq 1 - \varepsilon^2$, and the value of S_U^2 (and S_D^2) is at least $1 - (1-p)^{k_2} \geq 1/2$;
- if S_U^1 and S_D^1 are in the same color S_1 , and if S_U^2 and S_D^2 are in the same color S_2 , then S_U^3 and S_D^3 are in S_3 ; but this is impossible since this gives a 3-coloring of the initial graph and we have assumed that such coloring does not exist.

Discussion of the two cases above, implies that, for n sufficiently large, we have a lower bound of $8/7 - O(\varepsilon)$. ■

3.2 Differential approximation

In what follows, we denote by δ_A , the differential-approximation ratio of an algorithm A . When it becomes clear by the context, index A will be omitted.

Worst solution for minimum graph-coloring consists of giving to each vertex of the input graph an unused color ([11, 12, 14, 15]). For PROBABILISTIC COLORING, the value $\omega(G)$ of such solution is equal to $\sum_{v_i \in V} p_i$.

Let us first notice that on the contrary to the standard ratio, the differential ratio of the natural 2-coloring for bipartite graphs is unbounded below. Indeed, consider the following path on vertex-set $\{v_1, v_2, \dots, v_{n+2}\}$ with $p_1 = p_4 = 1$ and $p_i = 1/n^2$, $i = 2, 3, 5, \dots, n+2$. Then,

$$\delta = \frac{(2 + \frac{1}{n}) - 2}{(2 + \frac{1}{n}) - \left(2 - (1 - \frac{1}{n^2})^{\frac{n+1}{2}} + \frac{1}{n^2}\right)} \xrightarrow{n \rightarrow \infty} 0$$

Furthermore, one can easily show that neither δ is a refinement of ρ , nor ρ is a refinement of δ . In other words, it is possible to find sequences of paths where 2-coloring is not optimal and verifies:

1. $\delta \rightarrow 1$ and $\rho \rightarrow 1$;
2. $\delta \rightarrow 0$ and $\rho \rightarrow 1$;
3. $\delta \rightarrow 1$ and $\rho \rightarrow 2$ (the standard-approximation ratio achieved by the 2-coloring even in paths);
4. $\delta \rightarrow 0$ and $\rho \rightarrow 2$.

Indeed, consider the following paths:

- for item 1: $V = \{v_i : i = 1, \dots, n\}$, $p_i = 1$, for $i \equiv 1 \pmod{3}$, $p_i = \log n/n$, otherwise;
- for item 2: $V = \{v_1, v_2, v_3, v_4\}$, $p_1 = p_4 = 1/n$, $p_2 = p_3 = 1/n^2$;
- for item 3: $V = \{v_i : i = 1, \dots, n\}$, $p_i = 1$, for $i \equiv 1 \pmod{3}$, $p_i = 1/n^2$, otherwise;
- for item 4: $V = \{v_i : i = 1, \dots, n\}$, $p_1 = p_4 = 1$, $p_i = 1/(n-2)^2$, otherwise.

We now prove a positive result for 2-coloring that only works under the assumption that vertices have identical presence-probabilities.

Proposition 16. *In bipartite graphs, under identical vertex-probabilities, the differential ratio of 2-coloring is bounded below by $1/2$, and this bound is tight even in trees.*

Proof. Denote by S^* a maximum-cardinality independent set in a bipartite graph $B(U, D)$. By Propositions 1 and 4, the worst case for 2-coloring (U, D) is when $|U| = |D|$, while the best 3-coloring C^* is never better than the (possibly unfeasible) coloring $(S^*, U \setminus S^*, D \setminus S^*)$, where $|U \setminus S^*| = |D \setminus S^*| = 1$. Thus, the following inequalities hold:

$$f(U, D) \leq 2 \left(1 - (1-p)^{\frac{n}{2}}\right) \tag{23}$$

$$f(C^*) \geq 1 + 2p - (1-p)^{n-2} \tag{24}$$

Combining (23) and (24) we get:

$$1 \geq \delta \geq \frac{np - 2 \left(1 - (1-p)^{\frac{n}{2}}\right)}{np - (1 + 2p - (1-p)^{n-2})}$$

In order to prove that δ is bounded below by $1/2$, it suffices to show that φ is non-negative, where φ is defined as follows:

$$\begin{aligned}\varphi(p) &= 2 \left(np - 2 \left(1 - (1-p)^{\frac{n}{2}} \right) \right) - (np - (1 + 2p - (1-p)^{n-2})) \\ &= (n+2)p - 3 - (1-p)^{n-2} + 4(1-p)^{\frac{n}{2}}\end{aligned}$$

Function φ belongs to $\mathcal{C}^2[0, 1]$ and its derivatives are:

$$\begin{aligned}\varphi'(p) &= n + 2 - 2n(1-p)^{\frac{n}{2}-1} + (n-2)(1-p)^{n-3} \\ \varphi''(p) &= n(n-2)(1-p)^{\frac{n}{2}-2} - (n-2)(n-3)(1-p)^{n-4} \geq 0\end{aligned}$$

that involves $\varphi'(p) \geq \varphi'(0) = 0$ and, finally, $\varphi(p) \geq \varphi(0) = 0$, which is the expected result.

In order to show tightness, consider the following bipartite graph $B(D, U, E)$ (that is, in fact, a tree) with:

- $U = \{u_i : i \leq n/2\}$ and $D = \{d_i : i \leq n/2\}$;
- $E = \{(a_1, b_i) : i \leq n/2\} \cup \{(a_i, b_1) : i \leq n/2\}$;
- $p = 1/(n \log n)$.

The optimal coloring C^* of B is $C^* = (\{a_i, b_j : i, j \geq 2\}, a_1, b_1)$. So:

$$\begin{aligned}f(U, D) &= 2 \left(1 - \left(1 - \frac{1}{n \log n} \right)^{\frac{n}{2}} \right) = \frac{1}{\log n} - \frac{1}{4 \log^2 n} + o\left(\frac{1}{\log^2 n}\right) \\ f(C^*) &= 1 - \left(1 - \frac{1}{n \log n} \right)^{n-2} + \frac{2}{n \log n} = \frac{1}{\log n} - \frac{1}{2 \log^2 n} + o\left(\frac{1}{\log^2 n}\right) \\ \delta &= \frac{np - f(U, D)}{np - f(C^*)} \rightarrow \frac{1}{2}\end{aligned}$$

The proof of the proposition is now complete. ■

Let us now revisit algorithm **3-COLOR** already studied in Section 3.1 under the standard-approximation paradigm. In what follows, we analyze its differential-approximation ratio and show the following result.

Proposition 17. *The differential ratio of 3-COLOR in bipartite graphs is bounded below by $4/5$ and this bound is tight.*

In order to prove Proposition 17, let us at first introduce some further notations. For a bipartite graph $B(U, D, E)$, we set $V = U \cup D$. For any $H \subset V$ we set:

$$\begin{aligned}\sigma(H) &= \sum_{v_i \in H} p_i \\ \pi(H) &= \prod_{v_i \in H} (1 - p_i)\end{aligned}$$

Let S be the maximum weight independent set computed at step 2 of **3-COLOR** in $B(U, D, E)$. Using notations above, the differential-approximation ratios of the two candidate solutions C_0 and C_1 compared at step 3 of the algorithm are, respectively,

$$\delta_0 = \frac{\sigma(V) - 2 + \pi(U) + \pi(D)}{\sigma(V) - 2 + \pi(S) + \pi(V \setminus S)} \quad (25)$$

$$\delta_1 = \frac{\sigma(V) - 3 + \pi(S) + \pi(U \setminus S) + \pi(D \setminus S)}{\sigma(V) - 2 + \pi(S) + \pi(V \setminus S)} \quad (26)$$

Denote by C the solution that algorithm 3-COLOR returns at step 3 and recall that $C = \operatorname{argmin}\{|C_0|, |C_1|\}$. In other words, the differential ratio of the solution returned at this step is $\delta = \max\{\delta_0, \delta_1\}$. Finally, set $\mathcal{H} = \{S \cap D, S \cap U, D \setminus S, U \setminus S\}$. Recall also that, by local optimality arguments, the value of an optimal coloring C^* is never better than the value of the (generally unfeasible) coloring $S, V \setminus S$. Then:

$$\begin{aligned} f(C) &= \min\{f(U, D), f(S, U \setminus S, D \setminus S)\} \\ &= \min\{2 - \pi(U) - \pi(D), 3 - \pi(S) - \pi(U \setminus S) - \pi(D \setminus S)\} \\ f(C^*) &\geq f(S, V \setminus S) = 2 - \pi(S) - \pi(V \setminus S) \end{aligned}$$

Lemma 2. *Algorithm 3-COLOR reaches its worst ratio when, for any $H \in \mathcal{H}$, all but one vertices of H have probability 0.*

Proof. Notice first that, since for every $A, B \subset V$, $\pi(A \cup B) = \pi(A)\pi(B)$, once values of $\pi(H)$ are fixed for some $H \in \mathcal{H}$, then $f(U, D)$, $f(S, U \setminus S, D \setminus S)$ and $f(S, V \setminus S)$ are also fixed, regardless how probabilities are dispatched inside these sets.

When a, b are fixed constants such that $0 < a < b < x$, then function $(x - b)/(x - a) = 1 - (b - a)/(x - a)$ increases with x . Thus, once, for some $H \in \mathcal{H}$, $\pi(H)$ is fixed, the differential ratio δ reaches its worst value when $\sigma(V)$ is minimum. i.e., since elements from $\mathcal{H}$ are disjoint for every $H \in \mathcal{H}$, δ reaches its worst value when $\sigma(H)$ is minimum and this happens when $\sum_{v_i \in H} (1 - p_i)$ is maximized. It is easy to verify that quantity $\sum_{v_i \in H} (1 - p_i)$ is maximized when $|H| - 1$ vertices of H have probability 0 and only one has probability equal to $1 - \pi(H)$. This concludes the proof of the lemma. ■

Proof. (Proposition 17) Using Lemma 2, we identify, for simplicity, any $H \in \mathcal{H}$ with its only vertex of probability $1 - \pi(H) > 0$. Consider the following variables:

$$x = \frac{\pi(S \cap D) - \pi(S \cap U)}{2} \quad (27)$$

$$y = \frac{\pi(S \cap D) + \pi(S \cap U)}{2} \quad (28)$$

$$z = \frac{\pi(D \setminus S) - \pi(U \setminus S)}{2} \quad (29)$$

$$t = \frac{\pi(D \setminus S) + \pi(U \setminus S)}{2} \quad (30)$$

Using (27), (28), (29) and (30), ratios δ_0 and δ_1 in (25) and (26), respectively, can be rewritten as:

$$\begin{aligned} \delta_0 &= \frac{2 - 2y - 2t + (y + x)(t + z) + (y - x)(t - z)}{2 - 2y - 2t + (y^2 - x^2) + (t^2 - z^2)} \\ &= \frac{2 - 2y - 2t + 2yt + 2xz}{2 - 2y - 2t + (y^2 - x^2) + (t^2 - z^2)} \end{aligned} \quad (31)$$

$$\delta_1 = \frac{1 - 2y + y^2 - x^2}{2 - 2y - 2t + (y^2 - x^2) + (t^2 - z^2)} \quad (32)$$

Notice that the sign of x and z is indifferent to δ_1 , while δ_0 is minimal when these two variables have opposed sign. So, let us suppose, w.l.o.g, that x is nonnegative while z is nonpositive. Our goal is now to minimize $\max\{\delta_0, \delta_1\}$, under the following constraints:

$$\begin{aligned} y, t &\in]0, 1[\\ 0 &\leq x \leq y \text{ and } x \leq (1 - y) \end{aligned} \quad (33)$$

$$\begin{aligned} t - 1 &\leq z \leq 0 \text{ and } -t \leq z \\ 0 &\leq t - y \end{aligned} \quad (34)$$

The last constraint is a direct consequence of the definition of S to be a maximum-weight independent set; the rest of the above constraints result from definition of x, y, z and t .

We first focus on case where the 3-coloring C_1 computed by **3-COLOR** is better than the 2-coloring C_0 , in other words, $\delta_1 \geq \delta_0$. Then:

$$\begin{aligned}
\delta_0 \leq \delta_1 &\Leftrightarrow 2t(1-y) \geq (1-y^2) + x^2 + 2xz \\
&\Rightarrow 4(1-t)^2 \leq (1-y)^2 - 2x(x+2z) + \frac{x^2(x+2z)^2}{(1-y)^2} \\
&\Leftrightarrow (1-y)^2 - x^2 - 4(1-t)^2 \geq -(x-(2z+x))^2 + (2z+x)^2 \left(1 - \frac{x^2}{(1-y)^2}\right) \\
&\Leftrightarrow (1-y)^2 - x^2 - 4(1-t)^2 + 4z^2 \geq (2z+x)^2 \left(1 - \frac{x^2}{(1-y)^2}\right) \tag{35}
\end{aligned}$$

For convenience, let us introduce a new function $\Phi_1(x, y, z, t) = \delta_1 - (4/5)$, that is negative if and only if δ_1 is smaller than $4/5$. From (32), this function is defined by:

$$\begin{aligned}
\Phi_1 &= 5(1-2y+y^2-x^2) - 4(2-2y-2t+(y^2-x^2)+(t^2-z^2)) \\
&= -4(1-t)^2 + 4z^2 + (1-y)^2 - x^2 \tag{36}
\end{aligned}$$

Combining (33), (35) and (36) we get:

$$\Phi_1 \geq (2z+x)^2 \left(1 - \frac{x^2}{(1-y)^2}\right) \geq 0 \tag{37}$$

Let us now assume that $|C_0| \leq |C_1|$. We show that function $\Phi_0 = \delta_0 - (4/5)$ is also nonnegative, where, by (31):

$$\begin{aligned}
\Phi_0 &= 5(2-2y-2t+2yt+2xz) - 4(2-2y-2t+(y^2-x^2)+(t^2-z^2)) \\
&= 2-2y-2t-4y^2-4t^2+10yt+10xz+4x^2+4z^2
\end{aligned}$$

Function Φ_0 is $\mathcal{C}^\infty$ and, by (34), $\partial\Phi_0/\partial t = -2-8t+10y \leq 0$. Hence, Φ_0 is decreasing with t as long as $t \geq y$. Following assumption that $\delta_0 \geq \delta_1$, we get:

$$\begin{aligned}
\delta_1 \leq \delta_0 &\Leftrightarrow 2t(1-y) \leq (1-y^2) + x^2 + 2xz \\
&\Rightarrow \Phi_0(x, y, z, t) \geq \Phi_0\left(x, y, z, \frac{(1-y^2) + x^2 + 2xz}{2(1-y)}\right) \\
&\Rightarrow \Phi_0(x, y, z, t) \geq \Phi_1\left(x, y, z, \frac{(1-y^2) + x^2 + 2xz}{2(1-y)}\right) \geq 0 \tag{38}
\end{aligned}$$

Putting (37) and (38) together, we conclude that $\delta = \max\{\delta_0, \delta_1\} \geq 4/5$.

In order to show tightness, consider the following graph:

- $U = \{u_0, u_1, u_2\}$, $D = \{d_0, d_1, d_2\}$, $E = \{(u_1, d_2), (u_2, d_1)\} \cup (\{u_0\} \times D) \cup (\{d_0\} \times U)$;
- $p_{u_1} = p_{d_1} = 1/2$, $p_{u_2} = p_{d_2} = 1/4$, $p_{u_0} = p_{d_0} = \epsilon$, for a positive ϵ arbitrarily small.

Then, $S = \{u_1, d_1\}$ and:

$$\begin{aligned}
f(C) &= \min\left\{2\left(1 - \frac{1}{2} \times \frac{3}{4}(1-\epsilon)\right), 1 - \left(\frac{1}{2}\right)^2 + 2\left(1 - \frac{3}{4}(1-\epsilon)\right)\right\} = \frac{5}{4} + O(\epsilon) \\
f(C^*) &\leq 1 - \left(\frac{1}{2}\right)^2 + 1 - \left(\frac{3}{4}\right)^2 + 2\epsilon = \frac{19}{16} + O(\epsilon) \\
\delta &\leq \frac{\frac{3}{2} - \frac{5}{4} + O(\epsilon)}{\frac{3}{2} - \frac{19}{16} + O(\epsilon)} = \frac{4}{5} + O(\epsilon)
\end{aligned}$$

The proof of Proposition 17 is now complete. ■

We conclude this section by showing that the differential-approximation ratio of 3-COLOR is even better if one assumes identical vertex-probabilities, as shows the proposition below.

Proposition 18. *Under identical vertex-probabilities, the differential ratio of 3-COLOR in bipartite graphs is bounded below by 0.9 and this ratio is tight.*

Since enumeration of any possible coloring for graphs of bounded-size can be performed in polynomial time, we can assume, without loss of generality, that n is arbitrarily large. In order to prove the proposition, let us introduce the following variables:

$$\begin{aligned} x &= \frac{|S^*|}{n} \in \left[\frac{1}{2}, 1 \right[\\ \alpha &= (1-p)^{\frac{(1-x)n}{2}} \\ \beta &= (1-p)^{\frac{xn}{2}} \end{aligned}$$

It is easy to see that $(x, p) \mapsto (\alpha, \beta)$ is bijective and infinitely differentiable on $[1/2, 1[\times]0, 1[\rightarrow \{(a, b) : 0 < b \leq a < 1\}$.

Algorithm 3-COLOR produces coloring $(S, U \setminus S, D \setminus S)$. By Propositions 1 and 4, this coloring attains its worst possible value when $|U \setminus S| = |D \setminus S| = n(1-x)/2$; hence:

$$f(S, U \setminus S, D \setminus S) \leq 3 - (1-p)^{xn} - 2(1-p)^{\frac{(1-x)n}{2}} \quad (39)$$

As previously, let us denote by δ the differential ratio of 3-COLOR. Combining (23), (39) and using the inequality $f(C^*) \geq 2 - (1-p)^{nx} - (1-p)^{n(1-x)}$ (obtained by local optimality arguments and the fact that $|S^*| = nx$), we get:

$$\begin{aligned} 1 &\geq \delta \geq \frac{np - 3 + (1-p)^{xn} + 2(1-p)^{\frac{(1-x)n}{2}}}{np - 2 + (1-p)^{nx} + (1-p)^{n(1-x)}} = \frac{1}{\Phi(x, p)} \\ &\geq \frac{n \left(1 - (\alpha\beta)^{\frac{2}{n}} \right) - 3 + \beta^2 + 2\alpha}{n \left(1 - (\alpha\beta)^{\frac{2}{n}} \right) - 2 + \alpha^2 + \beta^2} \\ 1 &\geq \delta \geq \frac{np - 2 \left(1 - (1-p)^{\frac{n}{2}} \right)}{np - 2 + (1-p)^{nx} + (1-p)^{n(1-x)}} = \frac{1}{\Xi(x, p)} \\ &\geq \frac{n \left(1 - (\alpha\beta)^{\frac{2}{n}} \right) - 2 + 2\beta\alpha}{n \left(1 - (\alpha\beta)^{\frac{2}{n}} \right) - 2 + \alpha^2 + \beta^2} \end{aligned} \quad (40)$$

Lemma 3. *Ratio δ reaches its worst value on line $\beta = 2\alpha - 1$.*

Proof. The following hold for functions Ξ and Φ :

$$\frac{\partial \Xi(x, p)}{\partial x} = \frac{n \log(1-p) \left((1-p)^{xn} - (1-p)^{(1-x)n} \right)}{np - 2 - 2(1-p)^{\frac{n}{2}}} \quad (41)$$

$$\frac{\partial \Phi(x, p)}{\partial x} = \frac{n \log(1-p) \nu(x, p)}{\left(np - 3 + (1-p)^{nx} + 2(1-p)^{\frac{n(1-x)}{2}} \right)^2} \quad (42)$$

where:

$$\begin{aligned}
\nu(x, p) &= \left((1-p)^{nx} - (1-p)^{n(1-x)} \right) \left(np - 3 + (1-p)^{nx} + 2(1-p)^{\frac{n(1-x)}{2}} \right) \\
&\quad - \left((1-p)^{nx} - (1-p)^{\frac{n(1-x)}{2}} \right) \left(np - 2 + (1-p)^{nx} + (1-p)^{n(1-x)} \right) \\
&= AB - CD
\end{aligned}$$

Since $(1-x)/2 < 1-x \leq x$, and $p < 1$, it is easy to see that the derivative in (41) is nonnegative, while (42) has the same sign as $-\nu$. Notice that $C \leq A \leq 0 \leq B \leq D$, that means $\nu \geq 0$. Thus, when p is fixed, $\min\{\Phi, \Xi\}$ is maximized for some x satisfying:

$$\begin{aligned}
\Phi(x, p) = \Xi(x, p) &\Leftrightarrow 2(1-p)^{\frac{n}{2}} + 1 - 2(1-p)^{\frac{(1-x)n}{2}} - (1-p)^{xn} \\
&\Leftrightarrow 2\alpha\beta + 1 - 2\alpha - \beta^2 = 0 \Leftrightarrow \beta = 2\alpha - 1
\end{aligned}$$

that completes the proof of the lemma. ■

To get the ratio claimed, we bound by below the value of the worst solution $f(C_w) = n(1 - (\alpha\beta)^{2/n})$.

First, if $(\alpha\beta)^{2/n} \leq 1 - 20/n$, then $f(C_w) \geq 20$. Since the 2-coloring has value at most 2, we obtain a ratio greater than $(20 - 2)/20 = 9/10$.

Otherwise, $\alpha\beta \geq (1 - 20/n)^{n/2}$. For a sufficiently large n , we get $\alpha\beta \geq e^{-M}$ for some constant $M > 10$. Hence, $0 \geq \log(\alpha\beta) \geq -M$.

Using the fact that, for any $x \leq 0$, $e^x \leq 1 + x + x^2/2$, we get: $(\alpha\beta)^{2/n} \leq 1 + 2\log(\alpha\beta)/n + 2\log^2(\alpha\beta)/n^2$. Then, $f(C_w) \geq -2\log(\alpha\beta) - 2\log^2(\alpha\beta)/n$.

By Lemma 3, the worst case is obtained for $\beta = 2\alpha - 1$. Let $u = 1 - \alpha$. Then $\beta = 1 - 2u$. Since $\log(1 - t) \leq -t - t^2/2$, for any $t < 0$, we derive $\log(\alpha\beta) = \log(\alpha) + \log(\beta) \leq -3u - 5u^2/2$.

Finally, note that $x - x^2/n$ is increasing with x , for $x \leq n/2$. For n large enough, $-\log(\alpha\beta) \leq M \leq n/2$. So:

$$f(C_w) \geq 2 \left(3u + \frac{5u^2}{2} \right) - \frac{2 \left(3u + \frac{5u^2}{2} \right)^2}{n} = 6u + 5u^2 - u^2 \times O(n^{-1}) \quad (43)$$

Then, using (43) in (40) leads to:

$$\begin{aligned}
\delta &\geq 1 - \frac{(\alpha - \beta)^2}{f(C_w) - 2 + \alpha^2 + \beta^2} \\
&\geq 1 - \frac{u^2}{6u + 5u^2 - u^2 \times O(n^{-1}) - 2 + (1 - 2u + u^2) + (1 - 4u + 4u^2)} \\
&\geq 1 - \frac{1}{10 - O(n^{-1})}
\end{aligned}$$

In order to show tightness, consider the following bipartite graph $G(U, D)$ illustrated in Figure 4 for $n = 2$:

- an independent set S on $2n$ vertices, n of them belonging to U and the n remaining ones belonging to D ;
- n paths of size 4, $\{(r_i^1, r_i^2, r_i^3, r_i^4) : i \leq n\}$, such that, for any of them, r_i^1 and r_i^3 belong to U , while r_i^2 and r_i^4 belong to D ;
- one vertex $u \in U$ linked to each vertex of D and, symmetrically, one vertex $d \in D$ adjacent to each vertex of U ;

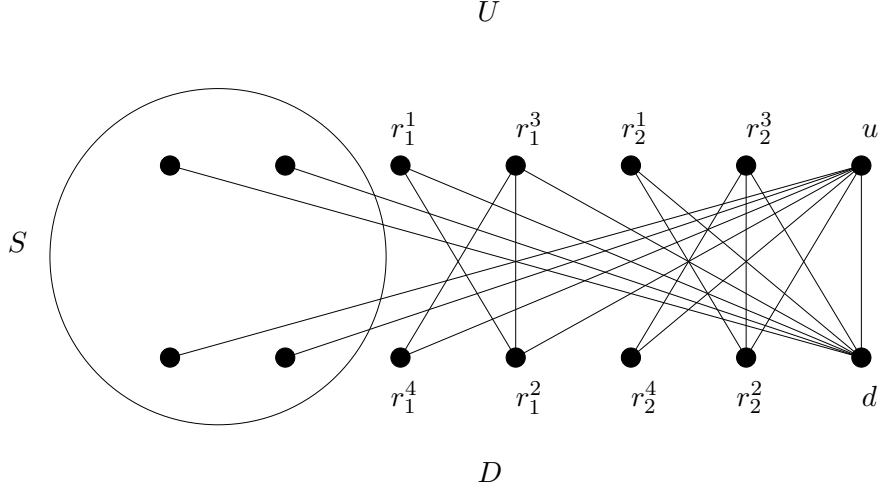


Figure 4: The 9/10-tightness example.

- $p = 1/n \log n$

An optimal coloring is $C^* = (S \cup \{r_i^1, r_i^3 : i \leq n\}, \{r_i^2, r_i^4 : i \leq n\} \cup \{d\}, \{u\})$, while 3-COLOR may produce, at worst, coloring $C = (S \cup \{r_i^1, r_i^4 : i \leq n\}, \{r_i^2 : i \leq n\} \cup \{d\}, \{r_i^3, i \leq n\} \cup \{u\})$. Hence:

$$\begin{aligned}
f(U, D) &= 2 \left(1 - \left(1 - \frac{1}{n \log n} \right)^{3n+1} \right) = \frac{6}{\log n} - \frac{9}{\log^2 n} + o\left(\frac{1}{\log^2 n}\right) \\
f(C) &= 1 - \left(1 - \frac{1}{n \log n} \right)^{4n} + 2 \left(1 - \left(1 - \frac{1}{n \log n} \right)^{n+2} \right) \\
&= \frac{6}{\log n} - \frac{9}{\log^2 n} + o\left(\frac{1}{\log^2 n}\right) \\
f(C^*) &= 1 - \left(1 - \frac{1}{n \log n} \right)^{4n} + 1 - \left(1 - \frac{1}{n \log n} \right)^{2n+1} + \frac{1}{n \log n} \\
&= \frac{6}{\log n} - \frac{10}{\log^2 n} + o\left(\frac{1}{\log^2 n}\right) \\
\delta &= \frac{(6n+2)p - f(U, D)}{(6n+2)p - f(C^*)} \rightarrow 9/10
\end{aligned}$$

This completes the proof of Proposition 18.

4 Particular families of bipartite and “almost” bipartite graphs: trees and cycles

Let us first note that for “trivial” families of bipartite graphs, as graphs isomorphic to a perfect matching, or to an independent set (i.e., collection of isolated vertices), PROBABILISTIC COLORING is polynomial, under any system of vertex-probabilities. In fact, for the former case, the optimal solution is given by a 2-coloring where for each pair of matched vertices, the one with largest probability is assigned to the first color, while the other one is assigned to the second color. For the latter case, trivially, the 1-coloring is optimal.

Also, under any vertex-probability system 2-coloring is optimal for stars. Indeed, the center of the star constitutes a color per se in any feasible coloring. Then, Proposition 4 applied on star's leaves suffices to conclude the proof.

4.1 Trees

Recall that the counter-example of Figure 1 shows that the natural 2-coloring is not always optimal in paths under distinct vertex-probabilities. In what follows, we study PROBABILISTIC COLORING on trees. As previously, we assume, that $|U| \geq |D|$.

Proposition 19. PROBABILISTIC COLORING can be optimally solved in trees with complexity bounded above by $(n+1)^{\Delta(k\Delta+k+1)+1}$ where Δ denotes the maximum degree of the tree and k the number of distinct vertex-probabilities.

Proof. Consider a tree $T(N, E)$ of order n . Let $p_1, \dots, p_k$ be the k distinct vertex-probabilities in T , n_i be the number of vertices of T with probability p_i and set $M = \prod_{i=1}^k \{0, \dots, n_i\}$. Recall finally that, from Proposition 7, any optimal solution of PROBABILISTIC COLORING in T uses at most $\Delta + 1$ colors.

Consider a vertex $v \in N$ with δ children and denote them by $v_1, \dots, v_\delta$. Let $c \in \{1, \dots, \Delta+1\}$ and $Q = \{q_1, \dots, q_{\Delta+1}\} \in M^{\Delta+1}$ where, for any $j \in \{1, \dots, \Delta+1\}$, $q_j = (q_{j1}, \dots, q_{jk}) \in M$. We search if there exists a coloring of $T[v]$, i.e., of the sub-tree of T rooted at v verifying both of the following properties:

- v is colored with color c ;
- q_{ij} vertices with probability p_i are colored with color j .

For this, let us define predicate $P_v(c, Q)$ with value **true** if such a coloring exists. In other words, we consider any possible configuration (in terms of number of vertices of any probability in any of the possible colors) for all the feasible colorings for $T[v]$.

One can determine value of P_v if one can determine values of P_{v_i} , $i = 1, \dots, \delta$. Indeed, it suffices that one looks-up the several alternatives, distributing the q_{ij} vertices (of probability p_i colored with color j) over the δ children of v (q_{ij} may be $q_{ij} - 1$ if $p(v) = p_i$ and $c = j$). More formally,

$$P_v(c, Q) = \bigvee_{(c_1, \dots, c_\delta)} \bigvee_{(Q^1, \dots, Q^\delta)} \left(P_{v_1}(c_1, Q^1) \wedge \dots \wedge P_{v_\delta}(c_\delta, Q^\delta) \right) \quad (44)$$

where in the clauses of (44):

- for $j = 1, \dots, \delta$, $c_j \neq c$ (in order that one legally colors v with color c),
- for $s = 1, \dots, \delta$, $Q^s \in M^{\Delta+1}$ and
- for any pair (i, j) :

$$\sum_{s=1}^{\delta} q_{ji}^s = \begin{cases} q_{ij} - 1 & \text{if } p(v) = p_i \text{ and } c = j \\ q_{ij} & \text{otherwise} \end{cases}$$

Observe now that $|M| \leq (n+1)^k$ and, consequently, $|M^{\Delta+1}| \leq (n+1)^{k(\Delta+1)}$. For any vertex v , there exist at most $n|M^{\Delta+1}|$ values of P_v to be computed and for any of these computations, at most $(n|M^{\Delta+1}|)^\delta$ conjunctions, or disjunctions, have to be evaluated. Hence, the total complexity of this algorithm is bounded above by $n(n|M^{\Delta+1}|)^{\delta+1} \leq (n+1)^{\Delta(k\Delta+k+1)+1}$. To conclude it suffices to output the coloring corresponding to the best of the values of predicate $P_r(c, Q)$, where r is the root of T . ■

Corollary 2. *PROBABILISTIC COLORING is polynomial in trees with bounded degree and with bounded number of distinct vertex-probabilities. Consequently, PROBABILISTIC COLORING is polynomial in bounded-degree trees with identical vertex-probabilities.*

Since paths are trees of maximum degree 2, we get also the following result.

Proposition 20. *PROBABILISTIC COLORING is polynomial in paths with bounded number of distinct vertex-probabilities. Consequently, it is polynomial for paths under identical vertex-probabilities.*

Let us note that for the second statement of Proposition 20, one can show something stronger, namely that *2-coloring is optimal for paths under identical vertex-probabilities*. Indeed, this case can be seen as an application of Proposition 8. The maximum independent set in a path coincides with U as any vertex of D is adjacent (and hence cannot have the same color) to a distinct vertex of U . This suffices to prove the proposition.

Consider now two particular classes of trees, denoted by $\mathcal{T}_E$ and $\mathcal{T}_O$, where all leaves lie exclusively either at even or at odd levels, respectively (root is considered at level 0). Obviously trees in both classes can be polynomially checked. We are going to prove that, under identical vertex-probabilities, PROBABILISTIC COLORING is polynomial for both $\mathcal{T}_E$ and $\mathcal{T}_O$. To do this, we first prove the following lemma where, for a tree T , we denote by N_E (resp., N_O) the even-level (resp., odd-level) vertices of T .

Lemma 4. *Consider $T \in \mathcal{T}_O$ (resp., in $\mathcal{T}_E$). Then N_O (resp., N_E) is a maximum independent set of T .*

Proof. We prove the lemma for $T \in \mathcal{T}_O$; case $T \in \mathcal{T}_E$ is completely similar. Set $n_o = |N_O|$, $n_e = |N_E|$ and notice that $n_o > 0$ (otherwise, T consists of a single isolated vertex). We will show ab absurdo that there exists a maximum independent set S^* of T such that $S^* = N_O$ (resp., $S^* = N_E$).

Suppose a contrario that any independent set S^* satisfies $|S^*| > n_o$. Then the following two cases can occur.

$S^* \subseteq N_E$. This implies $|S^*| \leq n_e$. Since any vertex in N_E has at least a child, $n_e \leq n_o$, hence $|S^*| \leq n_o$, absurd since N_O is also an independent set and S^* is supposed to be the maximum one.

$S^* \subseteq N_O \cup N_E$. In other words, S^* contains vertices from both N_O and N_E . Then, for any vertex $e \in N_E \cap S^*$ that is parent of a leaf, e has at least a children with no other neighbors in S^* . We can then switch between S^* and its children, obtaining so an independent set at least as large as S^* . We can iterate this argument with the vertices of this new independent set (denoted also by S^* for convenience) lying two levels above e (i.e., the great-grandparents of the leaves). Let g be such a vertex and assume that $g \in S^*$. Obviously, all its children are odd-level vertices and none of them is in S^* (a contrario, S^* would not be an independent set). Furthermore, none of these children can have a child $c \in S^*$ because e is an even-level vertex previously switched off from S^* , in order to be replaced by its children. Thus, we can again switch between g and its children, getting so a new independent set S^* larger than the previous one. We again iterate up to the root, always obtaining a new “maximum independent set” larger than the older one. Moreover, at the end, the independent set obtained will verify $S^* = N_O$. ■

Proposition 21. *Under identical vertex-probabilities, PROBABILISTIC COLORING is polynomial in $\mathcal{T}_O$ and $\mathcal{T}_E$.*

Proof. By Lemma 4, trees in $\mathcal{T}_O$ and $\mathcal{T}_E$ fit Proposition 8. So, for these trees, 2-coloring is optimal. ■

4.2 Cycles

In what follows in this section, we deal with cycles C_n of size n with identical vertex-probabilities. We will prove that in such cycles, **PROBABILISTIC COLORING** is polynomial.

Proposition 22. ***PROBABILISTIC COLORING** is polynomial in cycles with identical vertex-probabilities.*

Proof. Remark that in even cycles, Proposition 8 applies immediately; therefore, the natural 2-coloring is optimal.

Consider an odd cycle C_{2k+1} , denote by $1, 2, \dots, 2k+1$ its vertices and fix an optimal solution C^* for it. By Proposition 7, $|C^*| \leq 3$. Since C_{2k+1} is not bipartite, we can immediately conclude that $|C^*| = 3$. Set $C^* = (S_1^*, S_2^*, S_3^*)$ and denote by S^* a maximum independent set of C_{2k+1} ; assume $S^* = \{2i : i = 1, \dots, k\}$, i.e., $|S^*| = k$. By Proposition 4,

$$f(C^*) \geq f(S^*) + f_r^* = 1 - (1-p)^k + f_r \quad (45)$$

where f_r^* is the value of the best coloring in the rest of C_{2k+1} , i.e., in the sub-graph of C_{2k+1} induced by $V(C_{2k+1}) \setminus S^*$. This graph, of order $k+1$ consists of edge (v_1, v_{k+1}) and $k-1$ isolated vertices. Following once more Proposition 4, in a graph of order $k+1$ that is not a simple set of isolated vertices, the ideal coloring would be an independent set of size k and a singleton of total value $1 - (1-p)^k + p$. So, using (45), we get: $f(C^*) \geq 2 - 2(1-p)^k + p$. But the coloring $\hat{C} = (S^*, \{2i-1 : i = 1, \dots, k\}, \{2k+1\})$ attains this value; therefore it is optimal for C_{2k+1} , qed. ■

5 Split graphs

5.1 The complexity of **PROBABILISTIC COLORING**

We deal now with split graphs. This class of graphs is quite close to bipartite ones, since any split graph of order n is composed by a clique K_{n_1} , on n_1 vertices, an independent set S of size $n_2 = n - n_1$ and some edges linking vertices of $V(K_{n_1})$ to vertices of S . These graphs are, in some sense, on the midway between bipartite graphs and complements of bipartite graphs. In what follows, we first show that **PROBABILISTIC COLORING** is **NP**-hard in split graphs even under identical vertex-probabilities. For this, we prove that the decision counterpart of **PROBABILISTIC COLORING** in split graphs is **NP**-complete. This counterpart, denoted by **PROBABILISTIC COLORING**(K) is defined as follows: “given a split graph $G(V, E)$ a system of identical vertex-probabilities for G and a constant $K \leq |V|$, does there exist a coloring the functional of which is at most K ?”.

Proposition 23. ***PROBABILISTIC COLORING**(K) is **NP**-complete in split graphs, even assuming identical vertex-probabilities.*

Proof. Inclusion of **PROBABILISTIC COLORING**(K) in **NP** is immediate. In order to prove completeness, we will reduce 3-EXACT COVER ([13]) to our problem. Given a family $\mathcal{S} = \{S_1, S_2, \dots, S_m\}$ of subsets of a ground set $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_n\}$ (we assume that $\cup_{S_i \in \mathcal{S}} S_i = \Gamma$) such that $|S_i| = 3$, $i = 1, \dots, m$, we are asked if there exists a sub-family $\mathcal{S}' \subseteq \mathcal{S}$, $|\mathcal{S}'| = n/3$, such that $\mathcal{S}'$ is a partition on Γ . Obviously, we assume that n is a multiple of 3.

Consider an instance $(\mathcal{S}, \Gamma)$ of 3-EXACT COVER and set $q = n/3$. The split graph $G(V, E)$ for **PROBABILISTIC COLORING** will be constructed as follows:

- family $\mathcal{S}$ is replaced by a clique K_m (i.e., we take a vertex per set of $\mathcal{S}$); denote by $s_1, \dots, s_m$ its vertices;
- ground set Γ is replaced by an independent set $X = \{v_1, \dots, v_n\}$;

- $(s_i, v_j) \in E$ iff $\gamma_j \notin S_i$;
- $p > 1 - (1/q)$;
- $K = mp + q(1 - p) - q(1 - p)^4$.

Figure 5 illustrates the split graph obtained, by application of the three first items of the construction above, on the following 3-EXACT COVER-instance:

$$\begin{aligned}
\Gamma &= \{\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6\} \\
\mathcal{S} &= \{S_1, S_2, S_3, S_4, S_5\} \\
S_1 &= \{\gamma_1, \gamma_2, \gamma_3\} \\
S_2 &= \{\gamma_1, \gamma_2, \gamma_4\} \\
S_3 &= \{\gamma_3, \gamma_4, \gamma_5\} \\
S_4 &= \{\gamma_4, \gamma_5, \gamma_6\} \\
S_5 &= \{\gamma_3, \gamma_5, \gamma_6\}
\end{aligned} \tag{46}$$

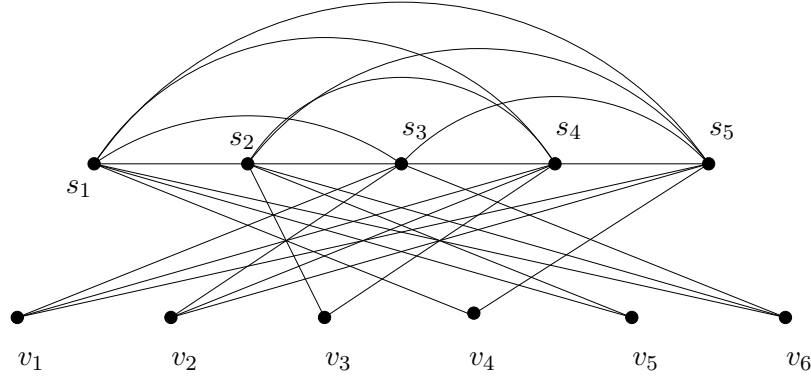


Figure 5: The split graph obtained from 3-EXACT COVER-instance described in (46).

Suppose that a partition $\mathcal{S}' \subseteq \mathcal{S}$, $|\mathcal{S}'| = q = n/3$ is given for $(\mathcal{S}, \Gamma, q)$. Order $\mathcal{S}$ in such a way that the q first sets are in $\mathcal{S}'$. For any $S_i \in \mathcal{S}'$, set $S_i = \{\gamma_{i_1}, \gamma_{i_2}, \gamma_{i_3}\}$. Then, subset $\{s_i, v_{i_1}, v_{i_2}, v_{i_3}\}$ of V is an independent set of G . Construct for G the coloring $C = (\{s_i, v_{i_1}, v_{i_2}, v_{i_3}\}_{i=1, \dots, q}, \{s_{q+1}\}, \dots, \{s_m\})$. It is easy to see that $f(C) = q(1 - (1-p)^4) + (m-q)p = mp + q(1 - p) - q(1 - p)^4 = K$.

Conversely, suppose that a coloring C is given for G with value $f(C) \leq K$. There exist, in fact, two types of feasible coloring in G :

1. C is as described just above, i.e., $C = (\{s_i, v_{i_1}, v_{i_2}, v_{i_3}\}_{i=1, \dots, q}, \{s_{q+1}\}, \dots, \{s_m\})$;
2. up to reordering of colors, C is of the form:

$$\begin{aligned}
C &= (S_1, \dots, S_{q_4}, S_{q_4+1}, \dots, S_{q_4+q_3}, S_{q_4+q_3+1}, \dots, S_{q_4+q_3+q_2}, \\
&\quad \{v_{q_4+q_3+q_2+1}\}, \dots, \{v_m\}, X')
\end{aligned} \tag{47}$$

where:

- the q_4 first sets are of the form: $\{s_i, v_{i_1}, v_{i_2}, v_{i_3}\}$, $i = 1, \dots, q_4$,
- the q_3 next sets are of the form: $\{s_i, v_{i_1}, v_{i_2}\}$, $i = q_4 + 1, \dots, q_4 + q_3$,
- the q_2 next sets are of the form: $\{s_i, v_{i_1}\}$, $i = q_4 + q_3 + 1, \dots, q_4 + q_3 + q_2$,

- the $m - (q_4 + q_3 + q_2)$ singletons are the remaining vertices of K_m which form a color per such vertex and
- X' is the subset of X not contained in the colors above;

we remark that coloring $C' = (\{s_1\}, \dots, \{s_m\}, X)$ is a particular case of (47) with $q_1 = q_2 = q_3 = 0$.

If C is of Type 1, then for any color $\{s_i, v_{i_1}, v_{i_2}, v_{i_3}\}$, $i = 1, \dots, q$, we take set S_i in $\mathcal{S}'$. By construction of G , set S_i covers elements γ_{i_1} , γ_{i_2} and γ_{i_3} of the ground set Γ . The q sets so selected form a partition on Γ of cardinality q .

Let us now assume that C is of Type 2 (see (47)). Note first that, for coloring C' mentioned at the end of Item 2 above, and for $p > 1 - (1/q)$:

$$f(C') = mp + (1 - (1 - p)^n) > mp + q(1 - p) - q(1 - p)^4 = K \quad (48)$$

Remark first that color X' (see Item 2) can never satisfy $|X'| \geq 4$; a contrario, using the local optimality argument of Proposition 2, since X' is the largest color, coloring C' would have value smaller than the one of C ; hence the latter value would be greater than K (see (48)). Therefore, we can assume $|X'| \leq 3$. In this case, one can, by keeping the q_4 colors of size 4 unchanged, progressively transform the rest of the colors by successive applications of Proposition 2 in order to create new (possibly unfeasible) 4-colors. This can be done by moving vertices from the smaller colors to the larger ones and is always possible since $n - 3q_4$ is a multiple of 3. Therefore, at the end of this processus, one can obtain exactly q (possibly unfeasible) 4-colors, the remaining vertices being colored with one color by vertex. Denoting by C'' the “coloring” so obtained, we have obviously, $f(C'') = K < f(C)$.

Therefore, by the discussion above, the only coloring having value at most K is the one of Type 1, qed. ■

Split graphs are particular cases of larger graph-family, the chordal graphs (graphs for which any cycle of length at least 4 has a chord ([3])).

Corollary 3. *PROBABILISTIC COLORING is **NP**-hard in chordal graphs even under identical vertex-probabilities.*

5.2 Standard-approximation results

For the rest of this section we deal with standard approximation of PROBABILISTIC COLORING in split graphs. Let $G(K, S, E)$ be such a graph, where K is the vertex set of the clique ($|K| = m$) and S is the independent set ($|S| = n$). Fix an optimal PROBABILISTIC COLORING-solution $C^* = (S_1^*, S_2^*, \dots, S_k^*)$ in $G(K, S, E)$.

Fact 1. $m \leq k \leq m + 1$. Indeed, since vertex-set K forms a clique, any solution in G will use at least m colors. On the other hand, if C^* uses more than m colors, this is due to the fact that there exist elements of S that cannot be included in any of the m colors associated with the vertices of K . If at least two such colors are used, then, since both of them are proper subsets of S (recall that S is an independent set), the local optimality argument of Proposition 1, would conclude the existence of a solution better than C^* , a contradiction. ■

Consider now the natural coloring, denoted by C , consisting of taking an unused color for any vertex of K and a color for the whole set S (in other words C uses $m + 1$ colors for G).

Proposition 24. *Coloring C is a 2-standard approximation for split graphs under any system of vertex-probabilities.*

Proof. Denote by $C^* = (S_1^*, S_2^*, \dots, S_k^*)$, an optimal solution in G and assume that colors are ranged in decreasing-value order, i.e., $f(S_i^*) \geq f(S_{i+1}^*)$, $i = 1, \dots, k-1$. From Fact 1, $m \leq k \leq m+1$. If $k = m+1$ and S_1^* is the color that is a subset of S , then local optimality arguments of Proposition 4 conclude that C is optimal. Hence, assume that S_1^* is a color including a vertex of K and vertices of S . For reasons of facility assume also that, upon a reordering of vertices, vertex $v_i \in K$ is included in color S_i^* ; also denote by p_i , the probability of vertex $v_i \in K$ and by q_i the probability of a vertex $v_i \in S$. Then,

$$f(C) = \sum_{i=1}^m p_i + \left(1 - \prod_{i=1}^n (1 - q_i)\right) \quad (49)$$

$$f(C^*) \geq \sum_{i=2}^m p_i + \left(1 - (1 - p_1) \prod_{i=1}^n (1 - q_i)\right) \quad (50)$$

where (50) holds thanks to local optimality arguments leading to Proposition 4, when we charge color S_1^* with all vertices of S . Observe also that:

$$1 - \prod_{i=1}^n (1 - q_i) \leq 1 - (1 - p_1) \prod_{i=1}^n (1 - q_i) \quad (51)$$

$$1 - (1 - p_1) \prod_{i=1}^n (1 - q_i) \geq p_1 \quad (52)$$

Combining (49) and (50), and using also (51) and (52), we get:

$$\begin{aligned} \frac{f(C)}{f(C^*)} &\leq \frac{p_1 + \sum_{i=2}^m p_i + \left(1 - \prod_{i=1}^n (1 - q_i)\right)}{\sum_{i=2}^m p_i + \left(1 - (1 - p_1) \prod_{i=1}^n (1 - q_i)\right)} \stackrel{(51)}{\leq} \frac{p_1 + \sum_{i=2}^m p_i + \left(1 - (1 - p_1) \prod_{i=1}^n (1 - q_i)\right)}{\sum_{i=2}^m p_i + \left(1 - (1 - p_1) \prod_{i=1}^n (1 - q_i)\right)} \\ &= 1 + \frac{p_1}{\sum_{i=2}^m p_i + \left(1 - (1 - p_1) \prod_{i=1}^n (1 - q_i)\right)} \stackrel{(52)}{\leq} 1 + \frac{p_1}{p_1 + \sum_{i=2}^m p_i} \leq 2 \end{aligned}$$

and the proof of the proposition is complete. ■

We now show the main positive standard-approximation result of this section, namely that PROBABILISTIC COLORING in split graphs can be solved by a polynomial time standard-approximation schema, under any system of vertex-probabilities.

Proposition 25. *PROBABILISTIC COLORING in split graphs is approximable by a polynomial time standard-approximation schema.*

Proof. Consider a split graph $G(K, S, E)$ and some optimal coloring $C^* = (S_1^*, S_2^*, \dots)$ of G , with $f(S_1^*) \geq f(S_2^*) \geq \dots$. Assume, without loss of generality, that C^* contains:

- some colors built from one vertex of K and some vertices of S ;
- some singletons of vertices of K ;
- less than one color all of its vertices belong to S (by Proposition 1); we denote this color by S_r^* .

Then the following facts can be derived for the form of C^* :

1. for any $i > r$, S_i^* is a singleton $\{k_{j_i}\} \subset K$;
2. for every $i < r$, the independent set (color) S_i^* is maximal (for the inclusion) for the graph $G_i = G[V \setminus (S_1^* \cup \dots \cup S_{i-1}^*)]$ (where, for $i = 1$, $S_{i-1}^* = \emptyset$).

Indeed, for Fact 1, if there exists a color $S_i^* = \{k_{j_i}, s_i^1, s_i^2, \dots\}$, for $i > r$, then by the local optimality arguments of Proposition 1 and given the ordering assumed for the colors $S_1^*, S_2^*, \dots$, putting vertices $s_i^1, s_i^2, \dots$ in S_r^* would improve the value of C^* .

On the other hand, for Fact 2, if S_i^* is not maximal for G_i , there exists a color S_j^* , $j > i$, some vertices of which can be legally introduced in S_i^* . Given the ordering of the colors, introduction of these vertices in S_i^* would lead to improvement of the value of C^* . Note now that, one can conclude from Fact 2 that, if S_i^* is not a singleton k_i of K , but it also contains some vertices of $S \setminus (S_1^* \cup \dots \cup S_{i-1}^*)$, then it contains all the vertices of $S \setminus (S_1^* \cup \dots \cup S_{i-1}^*)$ that are not its neighbors. This implies that if one could know exactly which vertex of K belongs to color S_i^* , then one can exactly determine any color S_i^* , for $i < r$.

Now, given a sequence X of k distinct vertices of K , we denote by C_X the set of the k colors built following the rule of Fact 2. Consider also coloring C consisting of taking an unused color for any vertex of K and a color for the whole set S (i.e., the one studied in Proposition 24). Revisit the proof of Proposition 24 and note that from (50), (51) and (52):

$$f(C^*) \geq f(C) - f(S_1^*) \quad (53)$$

Consider the following algorithm **SCHEMA** (it is rather a family of algorithms parameterized by a constant $\epsilon > 0$):

1. fix an $\epsilon > 0$;
2. set $k = \lceil 1/\epsilon \rceil$;
3. build and store coloring C of Proposition 24 for G ;
4. for any $k' \in \{1, \dots, k-1\}$ and any sequence $X \subset K$ of vertices, such that $|X| = k'$:
 - (a) construct the k' -coloring C_1 derived by the vertices of X along the rules of Fact 2;
 - (b) consider the subgraph of G induced by the still uncolored vertices and build the coloring C_2 of Proposition 24 for this graph;
 - (c) build and store coloring $C' = (C_1, C_2)$;
5. output the best coloring $\hat{C}$ among coloring C and the colorings C' built in steps 3 and 4c, respectively.

All executions of step 4 need at most $O(m^k) = O(m^{\lceil 1/\epsilon \rceil})$ while an execution of steps 4a and 4b take at most $O(nm)$. So, the overall complexity of **SCHEMA** is in $O(nm^{1+(1/\epsilon)})$, polynomial if ϵ is fixed.

Note first that if $r \leq k$, then $\hat{C}$ is optimal. Indeed, any subsequence of K vertices of size $r-1$ has been processed during the iterations of step 4 and any of the colorings C_X obtained has been completed by the still uncolored part of S (constituting a color) and by as many colors as the yet uncolored vertices of K . By what has been discussed above, in Facts 1 and 2 and just after them, one of the colorings so-built and completed is optimal and has been retained by **SCHEMA**.

If, on the other hand, $r > k$, then for the set X^* , corresponding to C^* , the coloring C_{X^*} obtained is $C_{X^*} = \{S_1^*, S_2^*, \dots, S_{k-1}^*\}$. Furthermore, on the subgraph of G induced by the still uncolored vertices, the coloring C_2 obtained is such that (consider (53)):

$$f(C_2) \leq f(S_k^*, \dots, S_\ell^*) + f(S_k^*) \quad (54)$$

where ℓ denotes the number of colors in C^* . Using (54), we get:

$$f(\hat{C}) \leq f(C^*) + f(S_k^*) \leq f(C^*) \left(1 + \frac{1}{k}\right) \leq f(C^*) (1 + \epsilon)$$

In other words, for any $\epsilon > 0$: $f(\hat{C})/f(C^*) \leq 1 + \epsilon$. So, for a fixed $\epsilon > 0$, **SCHEMA** constitutes a polynomial time standard-approximation schema for **PROBABILISTIC COLORING** in split graphs. ■

We now show that the result of Proposition 25 is optimal since it is the best possible approximability result even under identical vertex-probabilities.

Proposition 26. *Unless $P = NP$, **PROBABILISTIC COLORING**, on split graphs, cannot be solved by a fully polynomial time standard-approximation schema, even if identical vertex-probabilities are assumed.*

Proof. Revisit the proof of Proposition 23 and notice that it works for any $p > 1 - (1/q)$, where $q = n/3$. Denote by $|G|$, the size of G in a suitable encoding. Notice finally that, given that $|X'| \leq 3$, application of the local optimality principle of Proposition 2, in the case where the initial instance of **3-EXACT COVER** is a *yes*-instance (see [13]), the second best solution, for G is coloring $C' = (\{s_1\}, \dots, \{s_m\}, X)$ with value $f(C') = mp + 1 - (1 - p)^n$; furthermore, C' is feasible in any split graph.

Assume that a fully polynomial time standard-approximation schema A_ϵ exists for **PROBABILISTIC COLORING** in split graphs. Consider a graph G , resulting from the transformation described in the proof of Proposition 23 from an instance $(\mathcal{S}, \Gamma)$ of **3-EXACT COVER**, with $p > 1 - (3/n)$ say $p = 1 - (1/\omega(n))$, where ω is some polynomial with positive coefficients. Apply A_ϵ to G and take as final solution the best among the solution computed by this schema and C' .

If $(\mathcal{S}, \Gamma)$ is a *no*-instance, then C' is an optimal solution for G .

Suppose now that $(\mathcal{S}, \Gamma)$ is a *yes*-instance. In this case, the best coloring for G has value K and C' achieves ratio:

$$\begin{aligned} \frac{f(C')}{K} &= \frac{mp + 1 - (1 - p)^n}{mp + q(1 - p) - q(1 - p)^4} = \frac{mp + 1 - (1 - p)^n}{mp + q(1 - p)(1 - (1 - p)^3)} \\ &\geq \frac{mp + 1 - (1 - p)^4}{mp + 1 - (1 - p)^3} = 1 + \frac{p(1 - p)^3}{mp + 1 - (1 - p)^3} \end{aligned}$$

Henceforth, execution of A_ϵ on G with $\epsilon < p(1 - p)^3/(mp + 1 - (1 - p)^3)$ will return the optimal coloring of G with value K and, in this case, one can safely answer that $(\mathcal{S}, \Gamma)$ is a *yes*-instance for **3-EXACT COVER**. Notice finally, that, since $p = 1 - (1/\omega(n))$, $\epsilon \approx 1/m\omega^3(n)$, i.e., $1/\epsilon \approx m\omega^3(n)$. So, A_ϵ becomes an optimal and polynomial algorithm correctly deciding **3-EXACT COVER**. ■

5.3 A differential-approximation result

Consider again natural coloring C in split graphs, using $m + 1$ colors and consisting of taking an unused color for any vertex of K and a color for the whole set S . We show that C can be used to build a polynomial time differential-approximation schema when vertex-probabilities are identical.

Proposition 27. *PROBABILISTIC COLORING can be solved by a polynomial time differential-approximation schema on split graphs, under identical vertex-probabilities.*

Proof. Fix $\epsilon > 0$ and consider a split graph on a clique K_m and an independent set S of cardinality n . Consider coloring C consisting of assigning an unused color for any vertex of K_m and a color for the whole set S . Then:

$$\delta \geq \frac{(m+n)p - (mp + 1 - (1-p)^n)}{(m+n)p - ((m-1)p + 1 - (1-p)^{n+1})} \geq \frac{n}{n+1}$$

If $n \leq 1/\epsilon$, then an optimal coloring can obviously be found in linear time (but exponential in $1/\epsilon$). Otherwise, coloring C achieves differential-approximation ratio $\delta \geq 1 - \epsilon$, q.e.d. ■

Let us conclude this section by showing that, on split graphs, under any distribution of probabilities, the differential ratio of coloring C of Proposition 27 may be unbounded below.

Consider a split graph G and assume that one of the vertices of K_m and one of the vertices of S ($|S| = n$) have probability 1, while any other vertex of the graph has probability p . Finally, for the cross-edges, i.e., the edges between clique- and independent set-vertices, assume that only one clique-vertex with probability p is linked with any vertex of S (i.e., there exist n cross-edges). An optimal coloring C^* for G consists of taking the vertex of K_m with probability 1 together with all the vertices of S in the same color and color any of the other vertices of K_m with an unused color. Then:

$$\begin{aligned} \omega(G) &= \sum_{v_i \in V} p_i = 2 + (n+m-2)p \\ f(C) &= 2 + (m-1)p \\ f(C^*) &\leq 1 + (n+m-2)p \\ \delta &\leq \frac{2 + (n+m-2)p - 2 - (m-1)p}{2 + (n+m-2)p - (1 + (n+m-2)p)} = (n-1)p \end{aligned}$$

It can be immediately seen that there exist values for p (for instance, $p = 1/(n-1)^2$) for which $\delta \rightarrow 0$.

6 Final remark

There exists a list of interesting open problems dealing with the results of this paper. For example, the complexity of PROBABILISTIC COLORING remains open, notably for natural graph-families as: general bipartite graphs (under “small” probabilities), paths and cycles with distinct vertex-probabilities, trees, etc. The time we have spent trying to handle these problem makes us believing that they are indeed interesting open mathematical problems that deserve further studies and research.

References

- [1] I. Averbakh, O. Berman, and D. Simchi-Levi. Probabilistic a priori routing-location problems. *Naval Res. Logistics*, 41:973–989, 1994.
- [2] M. Bellalouna, C. Murat, and V. Th. Paschos. Probabilistic combinatorial optimization problems: a new domain in operational research. *European J. Oper. Res.*, 87(3):693–706, 1995.
- [3] C. Berge. *Graphs and hypergraphs*. North Holland, Amsterdam, 1973.

- [4] D. J. Bertsimas. *Probabilistic combinatorial optimization problems*. Phd thesis, Operations Research Center, MIT, Cambridge Mass., USA, 1988.
- [5] D. J. Bertsimas. On probabilistic traveling salesman facility location problems. *Transportation Sci.*, 3:184–191, 1989.
- [6] D. J. Bertsimas. The probabilistic minimum spanning tree problem. *Networks*, 20:245–275, 1990.
- [7] D. J. Bertsimas, P. Jaillet, and A. Odoni. A priori optimization. *Oper. Res.*, 38(6):1019–1033, 1990.
- [8] L. Bianchi, J. Knowles, and N. Bowler. Local search for the probabilistic traveling salesman problem: correction to the 2-p-opt and 1-shift algorithms. *European J. Oper. Res.*, 161(1):206–219, 2005.
- [9] H. L. Bodlaender, K. Jansen, and G. J. Woeginger. Scheduling with incompatible jobs. *Discrete Appl. Math.*, 55:219–232, 1994.
- [10] J.-M. Bourjolly, P. L. Hammer, and B. Simeone. Node-weighted graphs having the König-Egervary property. *Math. Programming Stud.*, 22:44–63, 1984.
- [11] M. Demange, P. Grisoni, and V. Th. Paschos. Approximation results for the minimum graph coloring problem. *Inform. Process. Lett.*, 50:19–23, 1994.
- [12] M. Demange, P. Grisoni, and V. Th. Paschos. Differential approximation algorithms for some combinatorial optimization problems. *Theoret. Comput. Sci.*, 209:107–122, 1998.
- [13] M. R. Garey and D. S. Johnson. *Computers and intractability. A guide to the theory of NP-completeness*. W. H. Freeman, San Francisco, 1979.
- [14] M. M. Halldórsson. Approximating discrete collections via local improvements. In *Proc. Symposium on Discrete Algorithms, SODA '95*, pages 160–169, 1995.
- [15] R. Hassin and S. Khuller. z -approximations. *J. Algorithms*, 41:429–442, 2001.
- [16] P. Jaillet. Probabilistic traveling salesman problem. Technical Report 185, Operations Research Center, MIT, Cambridge Mass., USA, 1985.
- [17] P. Jaillet. A priori solution of a traveling salesman problem in which a random subset of the customers are visited. *Oper. Res.*, 36(6):929–936, 1988.
- [18] P. Jaillet. Shortest path problems with node failures. *Networks*, 22:589–605, 1992.
- [19] P. Jaillet and A. Odoni. The probabilistic vehicle routing problem. In B. L. Golden and A. A. Assad, editors, *Vehicle routing: methods and studies*. North Holland, Amsterdam, 1988.
- [20] D. S. Johnson. Approximation algorithms for combinatorial problems. *J. Comput. System Sci.*, 9:256–278, 1974.
- [21] R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of computer computations*, pages 85–103. Plenum Press, New York, 1972.
- [22] C. Murat and V. Th. Paschos. The probabilistic longest path problem. *Networks*, 33:207–219, 1999.

- [23] C. Murat and V. Th. Paschos. A priori optimization for the probabilistic maximum independent set problem. *Theoret. Comput. Sci.*, 270:561–590, 2002. Preliminary version available at <http://www.lamsade.dauphine.fr/~paschos/documents/c166.pdf>.
- [24] C. Murat and V. Th. Paschos. The probabilistic minimum vertex-covering problem. *Int. Trans. Opl Res.*, 9(1):19–32, 2002. Preliminary version available at <http://www.lamsade.dauphine.fr/~paschos/documents/c170.pdf>.
- [25] C. Murat and V. Th. Paschos. On the probabilistic minimum coloring and minimum k -coloring. *Discrete Appl. Math.*, 154:564–586, 2006.
- [26] H. U. Simon. On approximate solutions for combinatorial optimization problems. *SIAM J. Disc. Math.*, 3(2):294–310, 1990.