Probabilistic coloring of bipartite and split graphs

Federico Della Croce, Bruno Escoffier, Cecile Murat, Vangelis Paschos

Probabilistic coloring of bipartite and split graphs

Federico Della Croce[1] Bruno Escoffier[2] Cécile Murat[2] Vangelis Th. Paschos[2]

[1] D.A.I., Politecnico di Torino, Italy, federico.dellacroce@polito.it

[2] LAMSADE, Université Paris-Dauphine, France, {escoffier,murat,paschos}@lamsade.dauphine.fr

June 18, 2004

Abstract

We revisit in this paper the probabilistic coloring problem ( ) and focus ourselves on bipartite and split graphs. We first give some general properties dealing with the optimal solution. We then show that the unique 2-coloring achieves approximation ratio 2 in bipartite graphs under any system of vertex-probabilities and propose a polynomial algorithm achieving tight approximation ratio 8/7 under identical vertex-probabilities. Then we deal with restricted cases of bipartite graphs. Main results for these cases are the following. Under non-identical vertex-probabilities - is polynomial for stars, for trees with bounded degree and a fixed number of distinct vertex-probabilities, and, consequently, also for paths with a fixed number of distinct vertex-probabilities. Under identical vertex-probabilities, is polynomial for paths, for even and odd cycles and for trees whose leaves are either at even or at odd levels. Next, we deal with split graphs and show that is **NP**-hard, even under identical vertex-probabilities. Finally, we study approximation in split graphs and provide a 2-approximation algorithm for the case of distinct probabilities and a polynomial time approximation schema under identical vertex-probabilities.

1 Preliminaries

In *minimum 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. Since adjacent vertices are forbidden to be colored with the same color, a feasible coloring can be seen as a partition of V into *independent sets*. So, the optimal solution of minimum coloring is a *minimum-cardinality partition into independent sets*. The decision version of this problem was shown to be **NP**-complete in Karp's seminal paper ([13]). The *chromatic number* of a graph is the smallest number of colors that can feasibly color its vertices.

In the probabilistic version of minimum coloring, denoted by , we are given:

- a graph $G(V, E)$ of order n, and an n-vector $\mathbf{Pr} = (p_1, \ldots, p_n)$ of vertex-probabilities; in other words, an instance of is a pair $(G, \mathbf{Pr})$;
- a *modification strategy* $\mathtt{M}$, i.e., an algorithm that when receiving a coloring $C = (S_1, \ldots, S_k)$ for V, called *a priori solution*, and a subgraph $G' = G[V']$ of G induced by a sub-set $V' \subseteq V$ as inputs, it modifies C in order to produce a coloring C' for G'.

The objective is to determine a coloring C^* (called optimal a priori solution) of G minimizing the quantity (commonly called functional) $E(G, C, \mathtt{M}) = \sum_{V' \subseteq V} \Pr[V'] |C(V', \mathtt{M})|$ where $C(V', \mathtt{M})$ is the solution computed by $\mathtt{M}(C, V')$ (i.e., by $\mathtt{M}$ when executed with inputs the a priori solution C and the subgraph of G induced by V') and $\Pr[V'] = \prod_{i \in V'} p_i \prod_{i \in V \setminus V'} (1 - p_i)$ (there exist 2^n distinct sets V'; therefore, explicit computation of $E(G, C, \mathtt{M})$ is, a priori, not polynomial). The complexity of is the complexity of computing C^*.

In this paper, we study under the following simple but intuitive modification strategy M: given an a priori solution C, take the set $C \cap V'$ as solution for $G[V']$, i.e., remove the absent vertices from C. Let us note that motivation of by two real-world applications, the former dealing with timetabling and the latter with planning, is given in [17]. Since the modification strategy M is fixed for the rest of the paper we will simplify notations by using $E(G, C)$ instead of $E(G, C, \mathtt{M})$ and $C(V')$ instead of $C(V', \mathtt{M})$. Set $k' = |C(V')|$, and consider the facts $\mathsf{F_j}$: color $\mathsf{S_j}$ has at least a vertex and $\bar{\mathsf{F}}_\mathsf{j}$: there is no vertex in color $\mathsf{S_j}$. Then, denoting by $\mathbf{1}_{\mathsf{F_j}}$ and $\mathbf{1}_{\bar{\mathsf{F}}_\mathsf{j}}$, respectively, their indicator functions, k' can be written as $k' = \sum_{j=1}^{k} \mathbf{1}_{\mathsf{F_j}} = \sum_{j=1}^{k}(1 - \mathbf{1}_{\bar{\mathsf{F}}_\mathsf{j}})$, and $E(G, C)$ can be written as:

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

It is easy to see that computation of $E(G, C)$ can be performed in at most $O(n^2)$ steps, consequently, $\in \mathbf{NP}$. On the other hand, from (1), we can easily characterize the optimal a priori solution C^* for : 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*.

has been originally studied in [17, 18], 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.

Besides , restricted versions of routing and network-design probabilistic minimization problems defined on complete graphs have been studied in ([2, 4, 5, 6, 8, 9, 10, 11]). In [16] the minimum vertex covering problem in general and in bipartite graphs is studied, while in [14, 15] the longest path and the maximum independent set, respectively, are tackled.

Dealing with in bipartite graphs, it is shown in [17] that it is **NP**-hard even if the input has only four distinct vertex-probabilities with one of them being equal to 0. Moreover, a polynomial algorithm was devised, achieving approximation ratio bounded above by 2.773. The **NP**-hardness result of [17] left, however, several open questions. For instance, "what is the complexity of when we further restrict inputs, say in paths, or trees, or cycles, or stars, ...?", etc. In this paper, we prove that, under non-identical vertex-probabilities, 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 ourselves on split graphs and show that, in such graphs, 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 $\mathrm{opt}(G)$ be the value of the optimal solution for G (following our notation for , $\mathrm{opt}(G) = E(G, C^*)$).

Graph-classes	Complexity	Approximation ratio
Bipartite	?	2
Bipartite, $p_i \geqslant 0.5$	Polynomial	
Bipartite, p_i identical	?	8/7
Trees	?	
Trees, bounded degree, k distinct probabilities	Polynomial	
Trees, all leaves exclusively at even or odd level, identical p_i's	Polynomial	
Stars	Polynomial	
Paths	?	
Cycles	?	
Even or odd cycles, p_i identical	Polynomial	
Split	**NP**-complete	2
Split, p_i identical	**NP**-complete	$1+\epsilon$, for any $\epsilon > 0$

Table 1: Summary of the main results of the paper.

The approximation ratio $\rho_{\mathtt{A}}(G)$ of the algorithm $\mathtt{A}$ on G is defined as $\rho_{\mathtt{A}}(G) = m(G,S)/\mathrm{opt}(G)$. An approximation algorithm achieving ratio, at most, ρ on any instance G of Π will be called ρ-approximation algorithm. A polynomial time approximation schema is a sequence $\mathtt{A}_\epsilon$ 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 $\rho_{\mathtt{A}_\epsilon}(G) \leqslant 1+\epsilon$.

Dealing with approximation issues, we show that the unique 2-coloring (where all nodes of each partition share the same color) achieves approximation ratio 2 in bipartite graphs under any system of vertex-probabilities. Furthermore, we propose a polynomial algorithm achieving approximation ratio 8/7 under identical vertex-probabilities. Both results importantly improve the 2.773 bound of [17]. We also provide a 2-approximation polynomial time algorithm for split graphs under distinct vertex-probabilities and a polynomial time approximation schema when vertex-probabilities are identical.

Table 1 summarizes the main results and open questions arising from the paper. Obviously, some of these results have several important corollaries. For instance, the fact that is polynomial in trees with bounded degrees and a fixed number of distinct probabilities, has as consequence that it is also polynomial in paths with a fixed number of distinct probabilities. Also, since is approximable within ratio 2 in general (i.e., under any system of vertex-probabilities) bipartite graphs, it is so in general trees, paths and even cycles, also.

2 Properties

2.1 Properties under non-identical vertex-probabilities

We give in this section some general properties about probabilistic colorings, 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: $f(C) = E(G,C)$, where $E(G,C)$ is given by (1), and, for $i = 1, \dots, k$, $f(S_i) = 1 - \prod_{v_j \in S_i}(1-p_j)$.

Property 1. Let $C = (S_1, \dots, S_k)$ be a k-coloring and assume that colors are numbered so that $f(S_i) \leqslant f(S_{i+1})$, $i = 1, \dots, k-1$. 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 $p_x \geqslant p_y$. If swapping colors of x and y leads to a new feasible coloring C', then $f(C') \leqslant f(C)$.

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

$$f\left(C'\right) - f(C) = f\left(S_i'\right) - f\left(S_i\right) + f\left(S_j'\right) - f\left(S_j\right) \qquad (2)$$

Set now

$$\begin{array}{rcl} S'_i &=& (S_i \setminus \{x\}) \cup \{y\} \\ S'_j &=& (S_j \setminus \{y\}) \cup \{x\} \\ S''_i &=& S_i \setminus \{x\} \;=\; S'_i \setminus \{y\} \\ S''_j &=& S_j \setminus \{y\} \;=\; S'_j \setminus \{x\} \end{array} \tag{3}$$

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

$$\begin{aligned} f\left(S'_i\right) - f\left(S_i\right) &= 1 - (1-p_y) \prod_{v_h \in S''_i} (1-p_h) - 1 + (1-p_x) \prod_{v_h \in S''_i} (1-p_h) \\ &= (p_y - p_x) \prod_{v_h \in S''_i} (1-p_h) \qquad (4) \\ f\left(S'_j\right) - f\left(S_j\right) &= 1 - (1-p_x) \prod_{v_h \in S''_j} (1-p_h) - 1 + (1-p_y) \prod_{v_h \in S''_j} (1-p_h) \\ &= (p_x - p_y) \prod_{v_h \in S''_j} (1-p_h) \qquad (5) \end{aligned}$$

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

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

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

With very similar arguments and operations as for Property 1, the following property, that is a particular case of Property 1, also holds.

Property 2. Let $C = (S_1, \ldots, S_k)$ be a k-coloring and assume that colors are numbered so that $f(S_i) \leqslant f(S_{i+1})$, $i = 1, \ldots, k-1$. 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' verifies $f(C') \leqslant f(C)$.

Property 3. In any graph of maximum degree Δ, the optimal solution of 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 Property 2, we achieve a $\Delta + 1$-coloring feasible for G with value better than the one of C. ∎

2.2 Properties under identical vertex-probabilities

Properties seen until now in this section work for any graph and for any vertex-probability system. Let us now focus ourselves on the case of identical vertex-probabilities. Remark first that, for this case, Property 2 has a natural counterpart expressed as follows.

Property 4. Let $C = (S_1, \ldots, S_k)$ be a k-coloring and assume that colors are numbered so that $|S_i| \leqslant |S_{i+1}|$, $i = 1, \ldots, k-1$. 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, verifies $f(C') \leqslant f(C)$.

Proof. Simply remark that if $|S_i| \leqslant |S_j|$, then $f(S_i) \leqslant f(S_j)$ and apply the same proof as for Property 1. ∎

Since, in the proof of Property 4, only the cardinalities of the colors intervene, the following corollary-property consequently holds.

Property 5. Let $C = (S_1, \ldots, S_k)$ be a k-coloring and assume that colors are numbered so that $|S_i| \leqslant |S_{i+1}|$, $i = 1, \ldots, k-1$. Consider two colors S_i and S_j, $i < j$, and a vertex-set $X \subset S_j$ such that, $|S_i| + |X| \geqslant |S_j|$. Consider (possibly unfeasible) coloring $C' = (S_1, \ldots, S_i \cup X, \ldots, S_j \setminus X, \ldots, S_k)$. Then, $f(C') \leqslant f(C)$.

From now on we define those colorings C such that Properties 1, or 2, or 4 hold, as "balanced colorings". In other words, for a balanced coloring C, there exists a coloring C', better than C, obtained as described in Properties 1, or 2, or 4. On the other hand, colorings for which transformations of the properties above cannot apply will be called "unbalanced colorings".

From the above definition, the following Proposition immediately holds.

Proposition 1. *For any balanced coloring, there exist an unbalanced one dominating it.*

Let us further restrict ourselves to bipartite graphs. Remark first that the cases of vertex-probability 0 or 1 are trivial: for the former any a priori solution has value 0; for the latter, coincides with the classical (deterministic) coloring problem where the (unique) 2-coloring is the best one.

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

Property 6. 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 $k \geqslant 3$ colors and its largest cardinality color S'_1 has cardinality β. Consider the following exhaustive two cases:

$\boldsymbol{\alpha(B) = \beta}$: then, it is su cient 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 1) and whose value coincides with the value of C;

$\boldsymbol{\alpha(B) < \beta}$: 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 1); 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'. ∎

3 General bipartite graphs

We first give an easy result showing that the hard cases for are the ones where vertex-probabilities are "small". Consider a bipartite graph $B(U, D, E)$ and denote by $p_{\min}$ its smallest vertex-probability.

Proposition 2. *If* $p_{\min} \geqslant 0.5$*, then the unique 2-coloring* $C = (U, D)$ *is optimal for* B*.*

Proof. If $p_{\min} \geqslant 0.5$, then, for any color S_i of any coloring C' of B, $1 > f(S_i) \geqslant 0.5$. Hence, for any feasible coloring C' of B, $f(C') \geqslant 0.5|C'| > 0.5$. On the other hand, as $f(C) < 2$, the optimal coloring can never use more than 3 colors. So, at a first time, an optimal coloring of B uses either 2, or 3 colors.

Consider any 3-coloring C' of B. Due to Properties 1 and 2, the best 3-coloring ever reachable (and possibly unfeasible) is coloring $C'' = (S''_1, S''_2, S''_3)$ assigning color S''_1 to a vertex of B with lowest probability (denote by v such a vertex), color S''_2 to a vertex with the second lowest probability (denote by $p'_{\min}$ this probability and by v' such a vertex) and color S_3'' to all the other vertices of B. It is easy to see that $f(S_3'') > f(S_2'') \geqslant f(S_1'')$. More precisely,

$$\begin{aligned} f\left(S_1''\right) &= p_{\min} && (7)\\ f\left(S_2''\right) &= p'_{\min} \geqslant p_{\min} && (8)\\ f\left(S_3''\right) &\geqslant p'_{\min} \geqslant p_{\min} \end{aligned}$$

Using (7) and (8) and the fact that $p_{\min} \geqslant 0.5$, we get:

$$f\left(S_1''\right) + f\left(S_2''\right) \geqslant 2p_{min} \geqslant 1 \tag{9}$$

We will prove that $f(U) + f(D) \leqslant f(S_1'') + f(S_2'') + f(S_3'')$. For this, we distinguish the following four exhaustive cases, depending on the fact that v and v' belong to U, or to D:

1. $v \in U$ and $v' \in D$;
2. $v \in D$ and $v' \in U$;
3. $v, v' \in U$;
4. $v, v' \in D$.

We will examine Cases 1 and 3 as Case 2 is exactly specular to the former and Case 4 to the latter.

For Case 1, using (7), (8) and (9), one has to show that

$$\begin{aligned} & 1 + 1 - \prod_{v_i \in (U \cup D)\setminus\{v,v'\}} (1-p_i) = 2 - \prod_{v_i \in (U \cup D)\setminus\{v,v'\}} (1-p_i) \\ & \geqslant 1 - \prod_{v_i \in U} (1-p_i) + 1 - \prod_{v_i \in D} (1-p_i) = 2 - \prod_{v_i \in U} (1-p_i) - \prod_{v_i \in D} (1-p_i) \end{aligned} \tag{10}$$

or, equivalently,

$$\prod_{v_i \in (U \cup D)\setminus\{v,v'\}} (1-p_i) - (1-p_{\min}) \prod_{v_i \in U\setminus\{v\}} (1-p_i) - \left(1-p'_{\min}\right) \prod_{v_i \in D\setminus\{v'\}} (1-p_i) \leqslant 0 \tag{11}$$

Set $\Gamma_1 = \prod_{v_i \in U\setminus\{v\}}(1-p_i)$ and $\Gamma_2 = \prod_{v_i \in D\setminus\{v'\}}(1-p_i)$. Then, (11) becomes:

$$\Gamma_1\Gamma_2 - (1-p_{\min})\,\Gamma_1 - \left(1-p'_{\min}\right)\Gamma_2 \leqslant 0 \tag{12}$$

Taking into account that $1 - p_{min} \geqslant \Gamma_1$ and $1 - p'_{min} \geqslant \Gamma_2$, (12) becomes $\Gamma_1^2 + \Gamma_2^2 - \Gamma_1\Gamma_2 = (\Gamma_1 - \Gamma_2)^2 + \Gamma_1\Gamma_2 \geqslant 0$, that is always true. The proof of Case 1 is complete.

We now analyze Case 3. By analogy with (11), we have to show that

$$\prod_{v_i \in (U \cup D)\setminus\{v,v'\}} (1-p_i) - (1-p_{\min})\left(1-p'_{\min}\right) \prod_{v_i \in U\setminus\{v,v'\}} (1-p_i) - \prod_{v_i \in D} (1-p_i) \leqslant 0 \tag{13}$$

Set this time $\Gamma_1 = \prod_{v_i \in U\setminus\{v,v'\}}(1-p_i)$ and $\Gamma_2 = \prod_{v_i \in D}(1-p_i)$. Then, (13) becomes:

$$\Gamma_1\Gamma_2 - (1-p_{\min})\left(1-p'_{\min}\right)\Gamma_1 - \Gamma_2 \leqslant 0 \tag{14}$$

or, equivalently $\Gamma_2(\Gamma_1 - 1) \leqslant (1-p_{\min})(1-p'_{\min})\Gamma_1$, which is always true since the left-hand quantity is negative and right-hand one is positive. This completes the proof of Case 3 and of the proposition. ∎

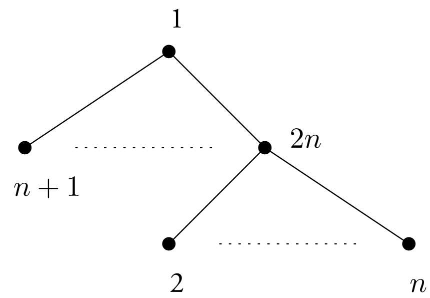

Figure 1: A tree with a 3-coloring of better value than the one of its 2-coloring.

When vertex-probabilities are generally and typically smaller than 0.5, the situation completely changes with respect the result of Proposition 2. Indeed, in this case, it is possible to provide instances, even with identical vertex-probabilities, where the 2-coloring does not provide the optimal solution. For instance, consider the tree T of Figure 1, where vertex 1 (the tree's root) is linked to vertices $n+1, \ldots, 2n$ and vertex $2n$ is linked to vertices $1, \ldots, n$.

Assume that vertex-probabilities of the vertices of T are all equal to $p \ll 0.5$. Then, the 2-coloring $\{\{1, \ldots, n\}, \{n+1, \ldots, 2n\}\}$ has value $f_2 = 2(1-(1-p)^n)$, while the 3-coloring $\{\{1\}, \{2, \ldots, 2n-1\}, \{2n\}\}$ has value $f_3 = 2(1-(1-p)) + (1-(1-p)^{2n-2})$. For p small enough and n large enough, we have $f_2 \approx 2$ and $f_3 \approx 1$.

The example of Figure 1 generalizes the counter-example of [17], dealing only with bipartite graphs, and shows that not only in general bipartite graphs but even in trees (that are restricted cases of bipartite graphs) the obvious 2-coloring is not always the optimal solution of .

In [17], it is shown that the natural 2-coloring is a 2.773-approximation of in bipartite graphs. In the following proposition, based upon Property 1, we improve this bound to 2.

Proposition 3. *In any bipartite graph B(U,D,E), its unique 2-coloring $C = (U, D)$ achieves approximation ratio bounded by 2. This bound is tight.*

Proof. Consider a bipartite graph $B(U, D, E)$. A trivial lower bound on the optimal solution cost (due to Property 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:

$$f(U \cup D) \leqslant f(C^*) \tag{15}$$

Assume that $f(U) \leqslant f(D)$. Then, since $D \subseteq U \cup D$, $f(D) \leqslant f(U \cup D)$. Therefore, using (15) $f(C) = f(U) + f(D) \leqslant 2f(D) \leqslant 2f(U \cup D) \leq 2f(C^*)$, qed.

For tightness, consider the 4-vertex path of Figure 2. 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 $\epsilon \to 0$, the latter is the optimal solution and the approximation ratio of the two coloring tends to 2. ∎

From the tightness of the bound provided in Proposition 3, the following corollary holds immediately.

Corollary 1. *The natural 2-coloring is not always optimal even when dealing with paths (or trees), under distinct vertex-probabilities. This coloring constitutes a tight 2-approximation for all these graph-families.*

We now restrict ourselves to the case of identical vertex-probabilities and consider the following algorithm, denoted by `3-COLOR` in what follows:

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

2. compute a maximum independent set S of B;

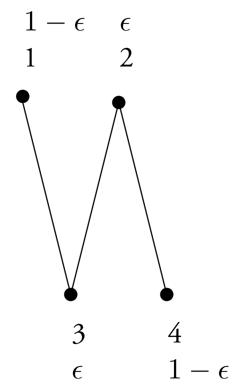

Figure 2: Ratio 2 is tight for the 2-coloring of a bipartite graph.

3. output the best coloring among C_0 and $C_1 = (S, U \setminus S, D \setminus S)$.

Obviously, 3-COLOR is polynomial, since computation of a maximum independent set can be performed in polynomial time in bipartite graphs ([7]).

Proposition 4. *Algorithm 3-COLOR achieves approximation ratio bounded above by 8/7 in bipartite graphs with identical vertex-probabilities. This bound is asymptotically tight.*

Proof. Consider an optimal solution $C^* = (S_1^*, S_2^*, \ldots S_k^*)$, and assume that $|S_1^*| \geqslant |S_2^*| \geqslant \ldots \geqslant |S_k^*|$. Set $n = |U \cup D|$, $n_1 = |S|$ and $n_2 = n - |S| = n - n_1$. Obviously, $n_1 \geqslant n_2$.

Based upon Property 4, the worst case for C_0 is reached when it is completely balanced, i.e., when $|U| = |D|$. In other words,

$$f\left(C_0\right) = f(U) + f(D) \leqslant 2\left(1 - (1-p)^{\frac{n_1+n_2}{2}}\right) \tag{16}$$

By exactly the same reasoning,

$$f\left(C_1\right) = f(S) + f(U \setminus S) + f(D \setminus S) \leqslant 1 - (1-p)^{n_1} + 2\left(1 - (1-p)^{\frac{n_2}{2}}\right) \tag{17}$$

Remark also that $|S_1^*| \leqslant |S_1| = n_1$. If this inequality is strict, then, applying Property 4, one, by emptying some colors S_j^*, $j > 1$, can obtain a (probably infeasible) coloring C' such that $f(C') \leqslant f(C^*)$ and the largest color of C' is of size n_1; in other words,

$$f\left(C^*\right) \geqslant f\left(C'\right) \geqslant 1 - (1-p)^{n_1} + 1 - (1-p)^{n_2} \tag{18}$$

Setting $\beta = (1-p)^{n_1/2}$, $\alpha = (1-p)^{n_2/2}$ and using (16), (17) and (18), we get (omitting, for simplicity, to index ρ by 3-COLOR):

$$\rho(B) = \min\left\{\frac{f\left(C_0\right)}{f\left(C^*\right)}, \frac{f\left(C_1\right)}{f\left(C^*\right)}\right\} \leqslant \min\left\{\frac{2(1-\alpha\beta)}{2-\alpha^2-\beta^2}, \frac{3-\beta^2-2\alpha}{2-\alpha^2-\beta^2}\right\} \tag{19}$$

Since $n_2 \leqslant n_1$, $0 \leqslant \beta \leqslant \alpha < 1$. We now show that function $f_1(x) = 2(1-\beta x)/(2-x^2-\beta^2)$ is decreasing with x in $[\beta, 1[$, while function $f_2(x) = (3-\beta^2-2x)/(2-x^2-\beta^2)$ is increasing with x in the same interval. Indeed, by elementary algebra, one immediately gets:

$$f_1'(x) = \frac{-2\beta(x-\beta)\left(x - \left(\frac{2-\beta^2}{\beta}\right)\right)}{\left(2-x^2-\beta^2\right)^2} \tag{20}$$

$$f_2'(x) = \frac{-2(x-1)\left(x - \left(2-\beta^2\right)\right)}{\left(2-x^2-\beta^2\right)^2} \tag{21}$$

In (20), $(2-\beta^2)/\beta \geqslant 1$; so, $f_1'(x)$ is positive for $x \in [\beta, 1[$ and, consequently f_1 is increasing with x in this interval. On the other hand, in (21), since $x < 1$ and $\beta < 1$, $x - 1 \leqslant 0$ and $x - (2-\beta^2) \leqslant 0$. So, $f_2'(x)$ is negative for $x \in [\beta, 1[$ and, consequently f_2 is decreasing with x in this interval.

In all, quantity $\min\{f_1(\alpha), f_2(\alpha)\}$ achieves its maximum value for α verifying $f_1(\alpha) = f_2(\alpha)$, or when $2(1-\alpha\beta) = 3 - \beta^2 - 2\alpha$, i.e., when $\alpha = (1+\beta)/2$. In this case (19) becomes (for $\beta \leqslant 1$):

$$\rho(B) \leqslant \frac{2\left(1 - \left(\frac{1+\beta}{2}\right)\beta\right)}{2 - \left(\frac{1+\beta}{2}\right)^2 - \beta^2} = \frac{8 - 4\beta - 4\beta^2}{7 - 2\beta - 5\beta^2} \leqslant \frac{8}{7}$$

and the claim about the approximation ratio is proved.

For tightness, 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 $v_U^1, \ldots, v_U^{n^2}$ belong to U and the n^2 remaining ones, denoted by $v_D^1, \ldots, v_D^{n^2}$ belong to D;
- n paths $P_1, \ldots, P_n$ of size 4 (i.e. on 3 edges); set, for $i = 1, \ldots, n$, $P_i = (p_i^1, p_i^2, p_i^3, p_i^4)$; 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 3 shows such a graph for $n = 2$.

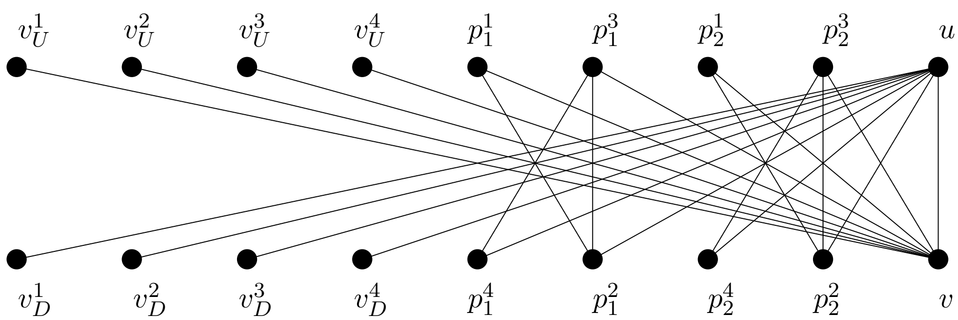

Figure 3: An 8/7 tightness instance with $n = 2$.

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

$$f(C_0) = 2\left(1 - (1-p)^{n^2+2n+1}\right) \tag{22}$$

On the other hand, one can see that a maximum independent set of B consists of the $2n^2$ vertices of S_1 plus two vertices per any of the n paths P_i, $i = 1, \ldots, n$. Assume without loss of generality that the maximum independent set computed in Step 2 of algorithm 3-COLOR is $S = S_1 \cup_{i=1,\ldots,n} \{p_i^1, p_i^4\}$. 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

$$f(C_1) = 1 - (1-p)^{2n^2+2n} + 2\left(1 - (1-p)^{n+1}\right) \tag{23}$$

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

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

Obviously,

$$f\left(\hat{C}\right) = 1 - (1-p)^{2n^2+2n} + 1 - (1-p)^{2n+1} + p \tag{24}$$

One can easily see that, for $n \to \infty$ and for $p = \ln 2/n$, (22), (23) and (24) give respectively: $f(C_0) \to 2$, $f(C_1) \to 2$ and $f(C^*) \leqslant f(\hat{C}) \to 7/4$. This proves the statement about tightness of `3-COLOR` and completes the proof of the proposition. ∎

Algorithm `3-COLOR` is a simplified version of the following algorithm, denoted by `MASTER-SLAVE`[1]:

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 a 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 $(S_1, \ldots, S_i, U_{i+1}, D_{i+1})$;
5. compute and store coloring $(S_1, S_2, \ldots)$, 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 approximation ratio for is at most $8/7$. We prove that it cannot do better. Indeed, consider the counter-example of Proposition 4. After computation of S the surviving graph consists of the vertex-set $\cup_{i=1,\ldots,n}\{p_i^2, p_i^3\} \cup \{u, v\}$. 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 $i \geqslant 2$ by `MASTER-SLAVE` are the same as coloring C_1 computed by `3-COLOR`. So, the following corollary is immediately concluded.

Corollary 2. *Algorithm `MASTER-SLAVE` achieves approximation ratio bounded above by 8/7 in bipartite graphs with identical vertex-probabilities. This bound is asymptotically tight.*

Notice that the tightness of the bound 8/7 can be shown for algorithm `3-COLOR` also on trees by means of the following instance T presented in Figure 4, for $n = 2$. There, the root-vertex a_0 of T has $n^2 + 1$ children $a_1, \ldots, a_{n^2}, b_0$. Vertices $\{a_1, \ldots, a_{n^2}\}$ have no children, while vertex b_0 has $n^2 + 1$ children $b_1, \ldots, , b_{n^2}, c_0$. Again, vertices $b_1, \ldots, b_{n^2}$ have no children, while vertex c_0 has $2n$ children $c_1, \ldots, c_{2n}$. Finally, vertex c_{2n} has no children while any vertex c_i, with $i = 1, \ldots, 2n-1$, has a single child-vertex d_i.

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. The 2-coloring $C_0' = (U, D)$ has the same value of before:

$$f\left(C_0'\right) = 2\left(1 - (1-p)^{n^2+2n+1}\right) \tag{25}$$

[1]This kind of algorithms approximately solving a "master" problem (in this case) by running a subroutine for a maximization "slave" problem (here) appears for first time in [12]; appellation "master-slave" for these algorithms is due to [19].

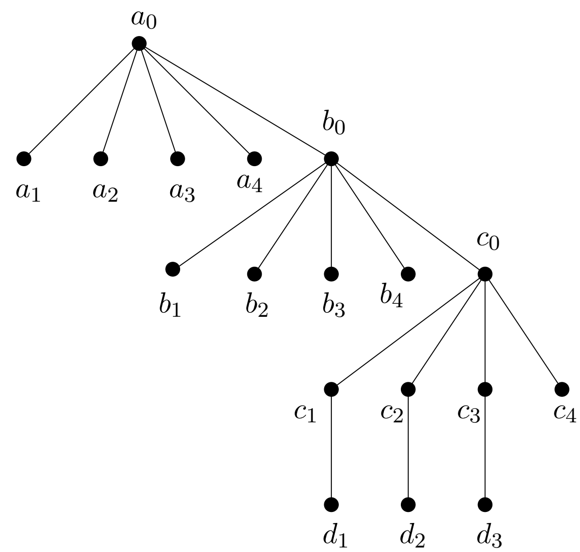

Figure 4: Lower bound 8/7 is attained for `3-COLOR` even in trees ($n = 4$).

Also in this case, a maximum independent set of T consists of $2n^2 + 2n$ vertices and we can assume, without loss of generality, that the maximum independent set computed in Step 2 of algorithm `3-COLOR` is $S' = \{a_1, \ldots, a_{n^2}, b_1, \ldots, b_{n^2}, c_{n+1}, \ldots, c_{2n}, d_1, \ldots, d_n\}$. Then the coloring $C' = (S', U \setminus S', D \setminus S')$ examined in Step 3 has value

$$f\left(C'\right) = 1 - (1-p)^{2n^2+2n} + 2\left(1 - (1-p)^{n+1}\right) \tag{26}$$

Besides, coloring $\hat{C}' = (\hat{S}'_1, \hat{S}'_2, \hat{S}'_3)$ with

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

has value

$$f\left(\hat{C}\right) = 1 - (1-p)^{2n^2+2n} + 1 - (1-p)^{2n+1} + p \tag{27}$$

and, as all values in (25), (26) and (27) correspond to those related to the example of Figure 3, we get the same 8/7 tight bound. Notice, however, that the proposed example does not guarantee the same tightness if algorithm `MASTER-SLAVE` is applied instead of algorithm `3-COLOR`.

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), 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.

4.1 Trees

Recall that the counter-example of Figure 2 shows that the natural 2-coloring is not always optimal in paths under distinct vertex-probabilities. In what follows, we exhibit classes of trees where - is polynomial. As previously, we assume, that $|U| \geqslant |D|$.

Proposition 5. *is polynomial in trees with bounded degree and with bounded number of distinct vertex-probabilities.*

Proof. Consider a tree $T(N, E)$ of order n and denote by Δ its maximum degree. Let $p_1, \ldots, p_k$ be the k distinct vertex-probabilities in T, n_i be the number of vertices of T with probability p_i and set $M = \prod_{i=1}^{k}\{0, \ldots, n_i\}$. Recall finally that, from Property 3, any optimal solution of in T uses at most $\Delta + 1$ colors.

Consider a vertex $v \in N$ with δ children and denote them by $v_1, \ldots, v_\delta$. Let $c \in \{1, \ldots, \Delta+1\}$ and $Q = \{q_1, \ldots, q_{\Delta+1}\} \in M^{\Delta+1}$ where, for any $j \in \{1, \ldots, \Delta+1\}$, $q_j = (q_{j_1}, \ldots, q_{j_k}) \in M$. 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_{i_j} 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}, $i = 1, \ldots, \delta$. Indeed, it su ces that one looks-up the several alternatives, distributing the q_{i_j} vertices (of probability p_i colored with color j) over the δ children of v (q_{i_j} may be $q_{i_j} - 1$ if $p(v) = p_i$ and $c = j$). More formally,

$$P_v(c, Q) = \bigvee_{(c_1,\ldots,c_\delta)} \bigvee_{\left(Q^1,\ldots,Q^\delta\right)} \left(P_{v_1}\left(c_1, Q^1\right) \wedge \ldots \wedge P_{v_\delta}\left(c_\delta, Q^\delta\right)\right) \tag{28}$$

where in the clauses of (28):

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

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

Observe now that $|M| \leqslant (n+1)^k$ and, consequently, $|M^{\Delta+1}| \leqslant (n+1)^{k(\Delta+1)}$. 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 su ces 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 3. *can be optimally solved in trees with complexity bounded above by* $(n+1)^{\Delta(k\Delta+k+1)+1}$ *where k denotes the number of distinct vertex-probabilities.*

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

Proposition 6. *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 6, 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 Property 6. 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 su ces to prove the proposition.

Consider now two particular class 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 been considered at level 0). Obviously trees in both classes can be polynomially checked. We are going to prove that, under identical vertex-probabilities, 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 1. *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 remark 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^* verifies $|S^*| > n_o$. Then the following two cases can occur.

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

- $\boldsymbol{S^* \subseteq N_O \cup N_E}$. 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 o 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 7. *Under identical vertex-probabilities, is polynomial in $\mathcal{T}_O$ and $\mathcal{T}_E$.*

Proof. By Lemma 1, trees in $\mathcal{T}_O$ and $\mathcal{T}_E$ fit Property 6. So, for these trees, 2-coloring is optimal. ∎

To conclude this paragraph, we deal with stars and show that is polynomial there, under any probability system.

Proposition 8. *Under any vertex-probability system 2-coloring is optimal for stars.*

Proof. Remark first that the center of the star constitutes a color per se in any feasible coloring. Then, Property 2 applied on star's leaves su ces to conclude the proof. ∎

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, is polynomial.

Proposition 9. *is polynomial in even cycles with identical vertex-probabilities.*

Proof. Remark that in even cycles, Property 6 applies immediately; therefore, the natural 2-coloring is optimal. ∎

Proposition 10. *is polynomial in odd cycles with identical vertex-probabilities.*

Proof. Consider an odd cycle C_{2k+1}, denote by $1, 2, \ldots, 2k+1$ its vertices and fix an optimal solution C^* for it. By Property 3, $|C^*| \leqslant 3$. 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 $S^* = \{2i : i = 1, \ldots, k\}$, ie., $|S^*| = k$. By Property 2,

$$f(C^*) \geqslant f(S^*) + f_r^* = 1 - (1-p)^k + f_r \tag{29}$$

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 Property 2, 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 (29), we get: $f(C^*) \geqslant 2 - 2(1-p)^k + p$. But the coloring $\hat{C} = (S^*, \{2i-1 : i = 1, \ldots, k\}, \{2k+1\})$ attains this value; therefore it is optimal for C_{2k+1}, qed. ∎

5 Split graphs

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 is **NP**-hard in split graphs even under identical vertex-probabilities. For this, we prove that the decision counterpart of in split graphs is **NP**-complete. This counterpart, denoted by (K) is defined as follows: "given a split graph $G(V,E)$ a system of identical vertex-probabilities for G and a constant $K \leqslant |V|$, does there exist a coloring the functional of which is at most K?".

Proposition 11. (K) *is **NP**-complete in split graphs, even assuming identical vertex-probabilities.*

Proof. Inclusion of (K) in **NP** is immediate. In order to prove completeness, we will reduce 3- ([7]) to our problem. Given a family $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ of subsets of a ground set $\Gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_n\}$ (we assume that $\cup_{S_i \in \mathcal{S}} S_i = \Gamma$) such that $|S_i| = 3$, $i = 1, \ldots, m$, 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- and set $q = n/3$. The split graph $G(V,E)$ for - 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 $s_1, \ldots, s_m$ its vertices;
- ground set Γ is replaced by an independent set $X = \{v_1, \ldots, v_n\}$;
- $(s_i, v_j) \in E$ i $\gamma_j \notin S_i$;
- $p > 1 - (1/q)$;
- $K = mp + q(1-p) - q(1-p)^4$.

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

$$\begin{array}{rcl} \Gamma & = & \{\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6\} \\ \mathcal{S} & = & \{S_1, S_2, S_3, S_4, S_5\} \\ S_1 & = & \{\gamma_1, \gamma_2, \gamma_3\} \\ S_2 & = & \{\gamma_1, \gamma_2, \gamma_4\} \\ S_3 & = & \{\gamma_3, \gamma_4, \gamma_5\} \\ S_4 & = & \{\gamma_4, \gamma_5, \gamma_6\} \\ S_5 & = & \{\gamma_3, \gamma_5, \gamma_6\} \end{array} \tag{30}$$

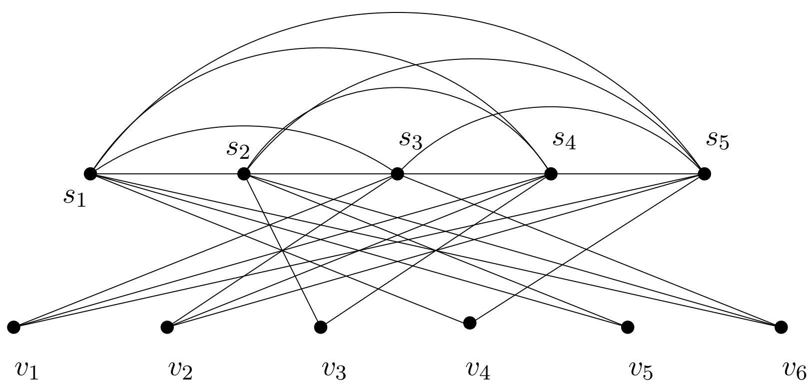

Figure 5: The split graph obtained from 3- -instance described in (30).

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 $f(C) \leqslant K$. There exist, in fact, two types of feasible coloring in G:

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

2. up to reordering of colors, C is of the form:

$$\begin{array}{rcl} C & = & (S_1, \dots, S_{q_4}, S_{q_4+1}, \dots, S_{q_4+q_3}, S_{q_4+q_3+1}, \dots, S_{q_4+q_3+q_2}, \\ & & \{v_{q_4+q_3+q_2+1}\}, \dots, \{v_m\}, X') \end{array} \tag{31}$$

   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;

   remark that coloring $C' = (\{s_1\}, \dots, \{s_m\}, X)$ is a particular case of (31) with $q_1 = q_2 = q_3 = 0$.

If C is of Type 1, then for any color $\{s_i, v_{i_1}, v_{i_2}, v_{i_3}\}$, $i = 1, \ldots, q$, 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 (31)). Note first that, for coloring C' mentioned at the end of Item 2 above, and for $p > 1 - (1/q)$:

$$f\left(C'\right) = mp + (1 - (1-p)^n) > mp + q(1-p) - q(1-p)^4 = K \tag{32}$$

Remark first that color X' (see Item 2) can never satisfy $|X'| \geqslant 4$; a contrario, using the unbalancing argument of Property 4, 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 (32)). Therefore, we can assume $|X'| \leqslant 3$. In this case, one can, by keeping the q_4 colors of size 4 unchanged, progressively unbalance the rest of the colors 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 been 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 4. *is **NP**-hard in chordal graphs even under identical vertex-probabilities.*

For the rest of this section we deal with approximation of 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 -solution $C^* = (S_1^*, S_2^*, \ldots, S_k^*)$ in $G(K, S, E)$.

Lemma 2. $m \leqslant k \leqslant m + 1$.

Proof. 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 unbalancing argument of Property 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 12. *Coloring C is a 2-approximation for split graphs under any system of vertex-probabilities.*

Proof. Denote by $C^* = (S_1^*, S_2^*, \ldots, S_k^*)$, an optimal solution in G and assume that colors are ranged in decreasing-value order, i.e., $f(S_i^*) \geqslant f(S_{i+1}^*)$, $i = 1, \ldots, k-1$. From Lemma 2, $m \leqslant k \leqslant m+1$. If $k = m + 1$ and S_1^* is the color that is a subset of S, then unbalancing arguments of Property 2 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,

$$f(C) = \sum_{i=1}^{m} p_i + \left(1 - \prod_{i=1}^{n}(1 - q_i)\right) \tag{33}$$

$$f\left(C^*\right) \geqslant \sum_{i=2}^{m} p_i + \left(1 - (1 - p_1)\prod_{i=1}^{n}(1 - q_i)\right) \tag{34}$$

where (34) holds thanks to unbalancing arguments leading to Property 2, when we charge color S_1^* with all vertices of S. Observe also that:

$$1-\prod_{i=1}^{n}(1-q_i) \leqslant 1-(1-p_1)\prod_{i=1}^{n}(1-q_i) \tag{35}$$

$$1-(1-p_1)\prod_{i=1}^{n}(1-q_i) \geqslant p_1 \tag{36}$$

Combining (33) and (34), and using also (35) and (36), we get:

$$\begin{aligned}
\frac{f(C)}{f(C^*)} &\leqslant \frac{p_1+\sum_{i=2}^{m}p_i+\left(1-\prod_{i=1}^{n}(1-q_i)\right)}{\sum_{i=2}^{m}p_i+\left(1-(1-p_1)\prod_{i=1}^{n}(1-q_i)\right)} \overset{(35)}{\leqslant} \frac{p_1+\sum_{i=2}^{m}p_i+\left(1-(1-p_1)\prod_{i=1}^{n}(1-q_i)\right)}{\sum_{i=2}^{m}p_i+\left(1-(1-p_1)\prod_{i=1}^{n}(1-q_i)\right)} \\
&= 1+\frac{p_1}{\sum_{i=2}^{m}p_i+\left(1-(1-p_1)\prod_{i=1}^{n}(1-q_i)\right)} \overset{(36)}{\leqslant} 1+\frac{p_1}{p_1+\sum_{i=2}^{m}p_i} \leqslant 2
\end{aligned}$$

and the proof of the proposition is complete. ▮

We now restrict ourselves in the case of identical graph probabilities. We will devise a polynomial time approximation schema for in split graphs. For this we first need the following lemma.

Lemma 3. *Given a split graph $G(K,S,E)$, if there exists a vertex in S with degree m, then coloring C using $m+1$ colors, one color per vertex of K and one color for the whole of vertices of S is optimal.*

Proof. Obviously, if the condition of the lemma is verified, any feasible coloring of G will have no less than $m+1$ colors. Then, using either Property 4, either Property 5, one can immediately prove that any coloring of at least $m+1$ colors has value at least $f(C)$, qed. ▮

Assume now that we deal with split graphs that do not verify condition of Lemma 3, i.e., that any vertex in S has degree strictly smaller than m. Then the following lemma holds (recall that S is an independent set).

Lemma 4. *Any subset of S the vertices of which have all the same neighbors in K, will be colored with the same color in any optimal coloring of G.*

Proof. Suppose a contrario that the statement of the lemma is false. Let $X=\{x_1,x_2,\ldots,x_j\}$ be a subset of S the vertices of which have the same neighbors but are colored with di erent colors. Let S_i be the largest color containing one of the vertices of X. Then, it is feasible to add the rest of the vertices of X in S_i by "improving" (by Properties 4, or 5) the value of the optimal solution. ▮

We are ready now to prove the following proposition that is the central part for the devising of our approximation schema. It a rms that if the size of the clique in G is fixed, then can be solved in polynomial time.

Proposition 13. *If m, the size of K in $G(K,S,E)$, is fixed, then can be solved in linear time.*

Proof. Recall that we deal with the case where vertices of S have degree at most $m-1$. We will count the number of all the distinct-value colorings of G. For this, we will construct a bipartite graph $B(U,D,E')$ with:

- $U=K$;

- D is in bijection with all the subsets of S, each such subset consisting of vertices of S having the same neighbors ; in other words, we contract any set of Lemma 4 into a single vertex; note that $|D| \leqslant \sum_{i=1}^{m-1} C_m^i < 2^m$;
- for any subset S' of S for which the neighbors of its vertices are $\{v_{i_1}, \dots, v_{i_k}\}$, the vertex of D corresponding to S' is linked to vertices $v_{i_1}, \dots, v_{i_k}$ in U.

The graph B just built has at most $m + 2^m$ vertices. The number of all the possible m-colorings of its vertices is then bounded by m^{m+2^m} which bounds also the number of the possible m-colorings of D, and this bound is a constant if m is so.

So, one can choose the best among the $m+1$-coloring of Lemma 3 and the m-colorings discussed just above, in order to produce an optimal solution for in linear time, since for any such coloring, its storing can be performed in linear time. The proof of the proposition is now complete. ∎

Consider now the following algorithm for , denoted by SCHEMA:

1. fix an $\epsilon > 0$;
2. if $m \leqslant 1/\epsilon$, then optimally solve by exhaustive look-up of all the feasible m-colorings as well as of coloring C of Proposition 13;
3. if $m \geqslant 1/\epsilon$, then output coloring C of Proposition 12.

Proposition 14. *Algorithm SCHEMA is a polynomial time approximation schema for in split graphs, under identical vertex-probabilities.*

Proof. By Proposition 13, if Step 2 is executed, the solution computed, in polynomial time since ϵ is a fixed constant, is optimal for . We deal now with Step 3 and the coloring C produced at this step. Denote by C^* an optimal coloring of G. Taking into account Property 4 (for (38) below), the following expressions hold:

$$f(C) = m \times p + (1 - (1-p)^n) \tag{37}$$
$$f(C^*) \geqslant (m-1)p + \left(1 - (1-p)^{n+1}\right) \tag{38}$$

Combination of (37) and (38), we get:

$$\frac{f(C)}{f(C^*)} \leqslant \frac{m \times p + (1 - (1-p)^n)}{(m-1)p + (1 - (1-p)^n)} \leqslant \frac{m+1}{m} \leqslant 1 + \epsilon$$

So, one can fix any arbitrarily small ϵ and then SCHEMA can solve in polynomial time within ratio $1+\epsilon$; hence, this algorithm is a polynomial time approximation schema for , qed. ∎

6 Concluding remarks and open problems

The problem dealt in this paper is quite di erent from the ones studied in [15, 16]. There, when strategies consisted of dropping absent vertices out of the a priori solution, the optimal a priori solutions were a maximum weight independent set, or a minimum weight vertex-covering, of the input graph considering that vertices are weighted by their probabilities. Here, as we have seen, the weight of an independent set is not an additive function and this makes that becomes very particular with respect to the probabilistic problems mentioned just above.

There exists a list of interesting open problems dealing with the results of this paper. For example, the complexity of remains open, notably for natural graph-families as: bipartite graphs with identical vertex-probabilities, paths and cycles with distinct vertex-probabilities, trees, etc.

In another order of ideas, an interesting approximation strategy for solving hard minimization problem is the so-called "master-slave" approximation. It consists of solving a minimization problem (the master one) by iteratively solving a maximization one (the slave problem) (for more details on this technique, cf., [1, 12, 19]). This kind of technique has a very natural application in the case of minimum coloring where the slave problem is the maximum independent set. It consists of iteratively computing an independent set in the graph, of coloring its vertices with the same unused color, of removing it from the graph and of repeating these stages in the subsequent surviving subgraphs until all vertices are colored. The slave independent set problem for is the one of determining the independent set S^* maximizing quantity $|S|/(1 - \prod_{v_i \in S}(1 - p_i))$ over any independent set of the input graph. Obviously, this problem is **NP**-hard in general graphs since for $p_i = 1$ for any vertex of the input graph we recover the classical maximum independent set problem. However, approximation of it in general graphs and complexity and, eventually, approximation results in graph-families as the ones dealt in this paper seem us interesting to be studied.

References

[1] L. Alfandari and V. Th. Paschos. Master-slave strategy and polynomial approximation. *Comput. Opti. Appl.*, 16:231–245, 2000.

[2] I. Averbakh, O. Berman, and D. Simchi-Levi. Probabilistic a priori routing-location problems. *Naval Res. Logistics*, 41:973–989, 1994.

[3] C. Berge. *Graphs and hypergraphs*. North Holland, Amsterdam, 1973.

[4] D. J. Bertsimas. On probabilistic traveling salesman facility location problems. *Transportation Sci.*, 3:184–191, 1989.

[5] D. J. Bertsimas. The probabilistic minimum spanning tree problem. *Networks*, 20:245–275, 1990.

[6] D. J. Bertsimas, P. Jaillet, and A. Odoni. A priori optimization. *Oper. Res.*, 38(6):1019–1033, 1990.

[7] M. R. Garey and D. S. Johnson. *Computers and intractability. A guide to the theory of NP-completeness*. W. H. Freeman, San Francisco, 1979.

[8] P. Jaillet. Probabilistic traveling salesman problem. Technical Report 185, Operations Research Center, MIT, Cambridge Mass., USA, 1985.

[9] P. Jaillet. A priori solution of a traveling salesman problem in which a random subset of the customers are visited. *Oper. Res.*, 36:929–936, 1988.

[10] P. Jaillet. Shortest path problems with node failures. *Networks*, 22:589–605, 1992.

[11] 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.

[12] D. S. Johnson. Approximation algorithms for combinatorial problems. *J. Comput. System Sci.*, 9:256–278, 1974.

[13] 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.

[14] C. Murat and V. Th. Paschos. The probabilistic longest path problem. *Networks*, 33:207–219, 1999.

[15] C. Murat and V. Th. Paschos. A priori optimization for the probabilistic maximum independent set problem. *Theoret. Comput. Sci.*, 270:561–590, 2002.

[16] C. Murat and V. Th. Paschos. The probabilistic minimum vertex-covering problem. *Int. Trans. Opl Res.*, 9(1):19–32, 2002.

[17] C. Murat and V. Th. Paschos. The probabilistic minimum coloring problem. In H. L. Bodlaender, editor, *Proc. 29th International Workshop on Graph Theoretical Concepts in Computer Science, WG'03*, volume 2880 of *Lecture Notes in Computer Science*, pages 346–357. Springer-Verlag, 2003.

[18] C. Murat and V. Th. Paschos. The probabilistic minimum coloring problem. Annales du LAMSADE 1, LAMSADE, Université Paris-Dauphine, 2003. Available on `http://www.lamsade.dauphine.fr/cahdoc.html#cahiers`, submitted.

[19] H. U. Simon. On approximate solutions for combinatorial optimization problems. *SIAM J. Disc. Math.*, 3(2):294–310, 1990.