CAHIER DU LAMSADE

Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision
(Université Paris-Dauphine)
Unité de Recherche Associée au CNRS ESA 7024

A PRIORI OPTIMIZATION FOR THE PROBABILISTIC MAXIMUM INDEPENDENT SET PROBLEM

CAHIER N° 166
janvier 2000

Cécile MURAT [1]
Vangelis PASCHOS [1]

received: October 1999.

[1] *LAMSADE, Université Paris-Dauphine, Place du Maréchal De Lattre de Tassigny, 75775 Paris Cedex 16, France {murat,paschos@lamsade.dauphine.fr).*

Table of Contents

Optimisation a priori pour le problème du stable maximum probabiliste

Résumé

Nous commençons par donner une définition formelle du concept relativement nouveau des problèmes d'optimisation combinatoire probabilistes (sous la méthodologie dite a priori développée par Bertsimas, Jaillet et Odoni). Ensuite, nous étudions la complexité de résoudre optimalement le problème du stable maximum probabiliste sous différentes stratégies de modification, ainsi que la complexité de solutions approchées. Pour les diverses stratégies étudiées nous présentons aussi des résultats sur la restriction du stable maximum probabiliste sur les graphes biparties.

Mots-clé : problèmes combinatoires, complexité, stable, graphe

A priori optimization for the probabilistic maximum independent set problem

Abstract

We first propose a formal definition for the concept of probabilistic combinatorial optimization problem (under the a priori method). Next, we study the complexity of optimally solving probabilistic maximum independent set problem under several a priori optimization strategies as well as the complexity of approximating optimal solutions. For the different strategies studied, we present results about the restriction of probabilistic independent set on bipartite graphs.

Keywords: combinatorial problems, computational complexity, independent set, graph

1 Introduction

In probabilistic combinatorial optimization, the probabilities are associated with the data describing an optimization problem (and not with the relations between data as, for example, in random graph theory ([4])). For a particular datum, we can see the probability associated with it as a measure of how this datum is likely to be present in the instance to be optimized, and in this sense, probabilistic elements are explicitly included in problem's formulation. In such a formulation, the objective function is a kind of carefully defined mathematical expectation over all the possible sub-instances induced by the initial instance.

The fact that in the framework of probabilistic combinatorial optimization problems (PCOP), the randomness lies in the presence of the data, makes that the underlying models are very adequate for the modeling of natural problems, where randomness models uncertainty, or fuzzy information, or hardness to forecast phenomena, etc.

For instance, in a transportation network whenever situations of several types of crises have to be modeled, we meet PCOPs like probabilistic shortest (or longest) path, probabilistic minimum spanning tree, probabilistic location, probabilistic traveling salesman problem (here we model the fact that perhaps some cities will not be visited), etc. In industrial automation, the systems for foreseeing workshops' production give rise to probabilistic scheduling or probabilistic set coverings or packings. In computer science, mainly in what concerns parallelism or distributed computer networks, PCOPs have to be solved; for example, modeling load balancing with non-uniform processors and failures possibility becomes a probabilistic graph partitioning problem; also in network reliability theory many probabilistic routing problems are met ([3]).

Definition 1. An **NP** optimization (**NPO**) problem Π is commonly defined as a fourtuple $(\mathcal{P}, \mathrm{S}, v_P, \mathrm{opt})$ such that:

- $\mathcal{P}$ is the set of instances of Π and it can be recognized in polynomial time;
- given $P \in \mathcal{P}$ (let n be the size of P), $\mathrm{S}(P)$ denotes the set of feasible solutions of P; moreover, for every $S \in \mathrm{S}(P)$ (let $|S|$ be the size of S), $|S|$ is polynomial in n; furthermore, for any P and any S (with $|S|$ a polynomial of n), one can decide in polynomial time if $S \in \mathrm{S}(P)$;
- given $P \in \mathcal{P}$ and $S \in \mathrm{S}(P)$, $v_P(S)$ denotes the value of S; v_P *is polynomially computable* and is commonly called objective function;
- $\mathrm{opt} \in \{\max, \min\}$. ∎

Based upon definition 1, we will try to give in what follows a formal definition for PCOPs (under the a priori thought process).

In [2, 3, 11], a thought process called in the sequel *a priori methodology* is adopted. Consider an instance P (where all the data are present) of an **NPO** problem Π, a sub-instance I of P and an algorithm $\mathtt{U}$ (called *modification strategy*), receiving $S \in \mathrm{S}(P)$ (called *a priori solution*) as input. Roughly speaking, the a priori methodology consists *in running $\mathtt{U}$ in order to modify S and to produce a new solution, dealing with the (present) sub-instance I of P.*

The a priori methodology together with the re-optimization one (an exhaustive computational technique) are the only ones used until now in the study of PCOPs. In [2, 3, 11], particular PCOPs are studied, but no formal definition of what a PCOP is (in the a priori framework) is given. In what follows in this section, we draw a formal framework for the a priori probabilistic combinatorial optimization.

Definition 2. Consider an **NPO** problem $\Pi = (\mathcal{P}, \mathsf{S}, v_P, \mathrm{opt})$ and a modification strategy $\mathtt{U}$ modifying S and producing a feasible Π-solution for I. The probabilistic version PΠ of Π is a quintuple $(\mathcal{I}, \mathsf{S}, \mathtt{U}, E_{\mathrm{P}\Pi}(P_S^{\mathtt{U}}), \mathrm{opt})$ such that:

- $\mathcal{I}$ is the set of instances of PΠ defined as $\mathcal{I} = \{I\!P = (P, \vec{\Pr})\}$, where $P \in \mathcal{P}$ and $\vec{\Pr}$ is an n-vector of occurrence-probabilities; $\mathcal{I}$ can be recognized in polynomial time;
- given $I\!P = (P, \vec{\Pr}) \in \mathcal{I}$, $\mathsf{S}(I\!P) = \mathsf{S}(P)$
- $\mathtt{U}$ is a modification strategy, i.e., an algorithm which, given a solution $S \in \mathsf{S}(I\!P)$ and a sub-instance I of P, receives S as input and computes a feasible Π-solution for I;
- given $I\!P \in \mathcal{I}$ and $S \in \mathsf{S}(I\!P)$, $E_{\mathrm{P}\Pi}(P_S^{\mathtt{U}})$ denotes the value of S; if $p_i = \Pr[i]$, $i \in P$, is the occurrence probability of datum[1] i, $\mathcal{F}(I_S^{\mathtt{U}})$ and $F(I_S^{\mathtt{U}})$ are the solution obtained by running $\mathtt{U}(\mathsf{S})$ on I and its value, respectively, $\Pr[I] = \prod_{i \in I} p_i \prod_{i \in P \setminus I}(1 - p_i)$ is the "occurrence" probability of I, then $E_{\mathrm{P}\Pi}(P_S^{\mathtt{U}}) = \sum_{I \subseteq P} \Pr[I] F(I_S^{\mathtt{U}})$; $E_{\mathrm{P}\Pi}(P_S^{\mathtt{U}})$ is commonly called functional;
- opt is the same as for Π. ∎

The complexity of PΠ is the complexity of computing $\hat{S} = \mathrm{argopt}\{E_{\mathrm{P}\Pi}(P_S^{\mathtt{U}}) : S \in \mathsf{S}(I\!P)\}$. Dealing with definition 2, the following remark must be underlined.

Remark 1. In definition 2, $\mathtt{U}$ is part of the instance and this seems somewhat unusual with respect to standard complexity theory where no algorithm intervenes in the definition of a problem. But note that *$\mathtt{U}$ is absolutely not an algorithm for* PΠ, in the sense that it does not compute $S \in \mathsf{S}(I\!P)$. It simply fits S (no matter how S has been computed) to I. ∎

Moreover, let us note that changing $\mathtt{U}$ one changes the definition of the problem itself. In other words, given a deterministic problem Π, the probabilistic problems induced by the quintuples $Q1 = (\mathcal{I}, \mathsf{S}, \mathtt{U}1, E_{\mathrm{P}\Pi}(P_S^{\mathtt{U}1}), \mathrm{opt})$ and $Q2 = (\mathcal{I}, \mathsf{S}, \mathtt{U}2, E_{\mathrm{P}\Pi}(P_S^{\mathtt{U}2}), \mathrm{opt})$ are two *distinct* PCOPs. Moreover, as we will see in the sequel, the choice of the modification strategy plays a crucial role in the complexity of the problem given that a strategy may or may not allow inclusion in **NP**.

A common thought process dealing with a PCOP PΠ is first to express $E_{\mathrm{P}\Pi}(P_S^{\mathtt{U}})$ in an explicit way. Such expression of the functional allows to precisely characterize the a priori solution $\hat{S}$ optimizing it and, consequently, to decide if $\hat{S}$ can or cannot be computed in polynomial time. Of course, if $\hat{S}$ cannot be computed in polynomial time, it is very interesting to decide if at least the functional itself can be computed in polynomial time (in other words, if the problem at hand is or is not in **NP**). This remark introduces a (perhaps the most) meaningful difference between probabilistic and deterministic problems. The general formula of the functional as it is given in definition 2 (as well as in PCOP-literature) suggests that, in general, exhaustive computation of $E_{\mathrm{P}\Pi}(P_S^{\mathtt{U}})$ requires $O(2^n)$ distinct computations. Consequently, a positive statement about inclusion of PΠ in **NP** is not immediate, at least for a great number of PCOPs. Of course, in many cases, functional's computation can be simplified and performed in polynomial time, but as we will see in the sequel, this is not always the case.

Given a graph $G = (V, E)$ of order n, an independent set is a subset $V' \subseteq V$ such that not any two vertices in V' are linked by an edge in G, and the maximum independent set problem (IS) is to find an independent set of maximum size. A natural extension of IS which will be also mentioned and discussed in the sequel is the maximum-weight independent set (WIS). Here, the

[1] This datum can be a vertex if we deal with graph-problems, or a set if we deal with optimization problem in set-systems, etc.

vertices of the input-graph are provided with positive weights, and the objective becomes to maximize the sum of the weights of the vertices in an independent set.

An instance of probabilistic independent set (PIS) is a pair $(G, \vec{\Pr})$ and is obtained by associating with each $v_i \in V$ an "occurrence" probability p_i and by considering a modification strategy U transforming feasible IS-solution S of G into an independent set for the sub-graph of G induced by a set $I \subseteq V$. As mentioned above, the objective for PIS is to determine the a priori solution $\hat{S}$ maximizing the functional $E_{\text{PIS}}(G_S^{\mathtt{U}})$.

Let us note that PIS is quite different from PCOPs already studied ([2, 3, 11, 12]). There, Euclidean versions of minimization routing-problems (such as the traveling salesman, or the shortest path, or the minimum spanning tree), all of them defined in complete graphs, are considered. For all these problems only one modification strategy, consisting, given an a priori solution, in removing absent vertices (as our strategy U1 introduced in section 2) is considered. Since for this kind of problems any feasible solution is a connected subgraph of the input-graph and since this latter graph is supposed complete, there always exist a proper (and easy) way to connect the vertices of the surviving in order to construct a feasible solution for the "present sub-instance". Such strategies seem efficient for minimization problems. On the other hand, for PIS, since it is a maximization problem, strategies as U1 are, as we will see later, rather inefficient and lead to low-quality solutions. Finally, no particular restrictions are imposed here on the input-graphs.

Except for its theoretical interest, PIS has also concrete applications. In [7], we have studied some aspects of the satellite shots planning problem. We have proposed a graph-theoretic modeling for this problem and we have proved that, via this modeling, the solution of the problem studied became exactly the computation of a maximum independent set in a kind of graph called "conflict graph". However, we have not taken into account that shots realized under strong cloud-covering are not operational. Consequently, it would be natural to also model weather forecasting. This can be done by associating probability p_i with vertex v_i of the conflict graph; the higher the vertex-probability, the more operational the shot taken. Such a model for the satellite shots planning problem allows, given an a priori IS-solution, computation of the expected number of operational shots.

There exist two interpretations of such an approach, each one characterized by its proper modification strategy:

- plan is firstly executed and one can know only after plan's execution if a shot is operational; in this case, one retains only the operational ones among the shots realized; this, in terms of PIS, amounts to application of strategy U1 introduced in section 2;
- weather forecasting becomes a certitude just before plan's execution; in this case, starting from an a priori IS-solution, one knows the vertices of this solution corresponding to non-operational shots, one discards them from the a priori solution and, finally, one renders the survived solution maximal by completing it by new vertices corresponding to operational shots; this amounts to application of other strategies, for example the ones denoted by U2, U3, U4 in the sequel and introduced in section 2.

Let us note that the probabilistic extension of the model of [7] can also be used to represent another concept, modeled in terms of PIS, where randomness on vertices represents this time probabilities that the corresponding shots are requested. Shot-probability equal to 1 means that this shot has already been requested, while shot-probability in [0,1] means that the corresponding shot will eventually be requested just before its realization. The corresponding PIS can be effectively solved by applying strategies U2 ([13]), or U3, or U4.

In what follows we consider maximal[2] (although not necessarily maximum) a priori independent sets and use five modification strategies, Ui, i = 1,...,5. For U1 and U5 we express their functionals in a closed form, we prove that they are computed in polynomial time, and we determine the a priori solutions that maximizing them. For U2 and U3, the expressions for the functionals are more complicated and it seems that they cannot be computed in polynomial time. Due to the complicated expressions for these functionals, we have not been able to characterize the a priori solutions maximizing them. Finally, for U4, we prove that the functional associated can be computed in polynomial time, but we are not able to precisely characterize the optimal a priori solution maximizing it. For all the strategies studied we also study the complexity of approximating optimal a priori solutions. Let us, once more, recall here that the strategies studied introduce in fact *five distinct* PCOPs denoted in the sequel by PIS1, PIS2, PIS3, PIS4 and PIS5, respectively. Finally, we study the probabilistic version of a natural restriction of IS, the one where the input graph is bipartite.

In the sequel, given a graph $G = (V, E)$ of order n, we sometimes denote by $V(G)$ the vertex-set of G. We denote by S a maximal solution of IS of G, by $\alpha(G)$ its cardinality, by S^* a maximum independent set of G, by $\alpha^*(G)$ its cardinality, by $\hat{S}$ an optimal PIS-solution (a priori solution) and by $\hat{\alpha}(G)$ its cardinality. Moreover, by $\Gamma(v_i)$, $i = 1,\ldots,n$, we denote the set of neighbors of the vertex v_i and the quantity $|\Gamma(v_i)|$ is called degree of v_i; by $\Gamma(V')$, $V' \subseteq V$, we denote the set $\cup_{v_i \in V'}\Gamma(v_i)$; also, $\delta_G = \min_{v_i \in V}\{|\Gamma(v_i)|\}$, $\Delta_G = \max_{v_i \in V}\{|\Gamma(v_i)|\}$ and μ_G is the average degree of G; finally, by $\Pr[v_i] = p_i$, we denote the fact that the presence probability of a vertex $v_i \in V$ equals p_i. Given a set $I \subseteq V$, we denote by $G[I] = (I, E_I)$ the subgraph of G induced by I (obviously, there are 2^n such graphs). Given a maximal solution S of IS (the a priori solution) in G, we denote by $S[I]$ the set $S \cap I$. For reasons of simplicity, the functional associated with PIS on graph G is denoted by $E_S^{\mathtt{U}}$ $(= \sum_{I \subseteq V} \Pr[I] F(I_S^{\mathtt{U}}))$.

2 The modification strategies and a preliminary result

In what follows we denote by GREEDY the classical greedy IS-algorithm. It works as follows: it orders the vertices of V in increasing degree-order, it includes the minimum-degree vertex in the solution, it deletes it together with its neighbors (as well as all edges incident to these vertices) from V (E), it reorders the vertices of the surviving graph and so on, until all vertices are removed. Moreover, we denote by SIMGREEDY, a simplified version of GREEDY where after removing a vertex and its neighbors, the algorithm does not reorder the vertices of the surviving graph.

2.1 Strategy U1

Given an a priori IS-solution S and a present subset $I \subseteq V$, modification strategy U1 consists in simply moving the absent vertices out of S.

```
BEGIN (*U1(G,S)*)
      OUTPUT F(I_S^U1) ← S[I] ← S ∩ I;
END. (*U1*)
```

2.2 Strategies U2 and U3

Modification strategy U2 is a two-step method: it first applies U1 to obtain $S[I]$, it next applies GREEDY on the graph $G[\tilde{I}] = G[I \setminus \{S[I] \cup \Gamma(S[I])\}]$ and, finally, it retains the union of the two independent sets obtained as final IS-solution for $G[I]$.

[2] If we add a vertex, the result is not an independent set.

```
BEGIN (*U2(G,S) /U3(G,S)/*)
      S[I] ← U1(G,S);
      G[Ĩ] ← G[I \ {S[I] ∪ Γ(S[I])}];
      S[Ĩ] ← GREEDY(G[Ĩ]);            /S[Ĩ] ← SIMGREEDY(G[Ĩ]);/
      OUTPUT F(I_S^U2) ← S[I] ∪ S[Ĩ];  /OUTPUT F(I_S^U3) ← S[I] ∪ S[Ĩ];/
END. (*U2 /U3/*)
```

Finally, strategy U3 is identical to U2 modulo the fact that, instead of GREEDY, algorithm SIMGREEDY is executed (instructions between slashes above refer to modification strategy U3).

2.3 Strategy U4

Strategy U4 starts from $S[I]$ and completes it with the isolated vertices (vertices with no neighbors) of the graph $G[\tilde{I}]$.

```
BEGIN (*U4(G,S)*)
      S[I] ← U1(G,S);
      G[Ĩ] ← G[I \ {S[I] ∪ Γ(S[I])}];
      OUTPUT F(I_S^U4) ← S[I] ∪ {v_i ∈ Ĩ : Γ(v_i) = ∅};
END. (*U4*)
```

2.4 Strategy U5

Strategy U5 applies the natural relation between a minimal vertex cover[3] and a maximal independent set in a graph, i.e., a minimal vertex cover (resp., maximal independent set) is the complement, with respect to the vertex set of the graph, of a maximal independent set (resp., a minimal vertex cover).

```
BEGIN (*U5(G,S)*)
(1)   C ← V \ S;
(2)   C[I] ← C ∩ I;
(3)   R ← {v_i ∈ C[I] : Γ(v_i) = ∅};
(4)   C[I] ← C[I] \ R;
(5)   OUTPUT F(I_S^U5) ← I \ C[I];
END. (*U5*)
```

2.5 A general mathematical formulation for the five functionals

Theorem 1. *Consider an a priori solution S of cardinality $\alpha(G)$ for G; consider strategies Uk, k = 1, ..., 5. With each vertex $v_i \in V$ we associate a probability p_i and a random variable $X_i^{Uk,S}$, k = 1, ..., 5, defined, for every $I \subseteq V$, by*

$$X_i^{Uk,S} = \begin{cases} 1 & v_i \in \mathcal{F}\left(I_S^{Uk}\right) \\ 0 & \textit{otherwise} \end{cases} \tag{1}$$

Then

$$E_S^{Uk} = \sum_{v_i \in S} p_i + \sum_{v_i \in (V \setminus S)} \Pr\left[X_i^{Uk,S} = 1\right]. \tag{2}$$

In particular, if, for each vertex $v_i \in V$, $p_i = p$, then

$$E_S^{Uk} = p\alpha(G) + \sum_{v_i \in (V \setminus S)} \Pr\left[X_i^{Uk,S} = 1\right]. \tag{3}$$

[3]Given a graph $G = (V, E)$, a vertex cover of G is a set $V' \subseteq V$ such that for any $v_i v_j \in E$ either v_i, or v_j belongs to V'.

Proof. By expression (1), $F(I_S^{\mathtt{Uk}}) = \sum_{i=1}^{n} X_i^{\mathtt{Uk},S}$. So,

$$\begin{aligned} E_S^{\mathtt{Uk}} &= \sum_{I \subseteq V} \Pr[I] F\left(I_S^{\mathtt{Uk}}\right) = \sum_{I \subseteq V} \Pr[I] \sum_{i=1}^{n} X_i^{\mathtt{Uk},S} = \sum_{i=1}^{n} \sum_{I \subseteq V} \Pr[I] X_i^{\mathtt{Uk},S} \\ &= \sum_{i=1}^{n} E\left(X_i^{\mathtt{Uk},S}\right) = \sum_{i=1}^{n} \Pr\left[X_i^{\mathtt{Uk},S} = 1\right] = \sum_{i=1}^{n} \Pr\left[X_i^{\mathtt{Uk},S} = 1\right] \left(1_{\{v_i \in S\}} + 1_{\{v_i \notin S\}}\right) \\ &= \sum_{i=1}^{n} \Pr\left[X_i^{\mathtt{Uk},S} = 1\right] 1_{\{v_i \in S\}} + \sum_{i=1}^{n} \Pr\left[X_i^{\mathtt{Uk},S} = 1\right] 1_{\{v_i \notin S\}}. \end{aligned}$$

But, if $v_i \in S$, then necessarily $X_i^{\mathtt{Uk},S} = 1$, $\forall I$ such that $v_i \in I$; so, $\Pr[X_i^{\mathtt{Uk},S} = 1] = p_i$, $v_i \in S$, and consequently,

$$E_S^{\mathtt{Uk}} = \sum_{v_i \in S} p_i + \sum_{i=1}^{n} \Pr\left[X_i^{\mathtt{Uk},S} = 1\right] 1_{\{v_i \notin S\}} = \sum_{v_i \in S} p_i + \sum_{v_i \in (V \setminus S)} \Pr\left[X_i^{\mathtt{Uk},S} = 1\right].$$

If $p_i = p$, $v_i \in V$, then the result of expression (3) is immediately obtained from expression (2). ∎

Let us note that, as it can be easily deduced from the proof of theorem 1, the above result holds for any strategy which it first determines $S[I]$, it next computes an independent set $S[\tilde{I}]$ on $G[\tilde{I}]$, and it finally considers as solution for $G[I]$ the set $S[I] \cup S[\tilde{I}]$.

3 The complexity of PIS1

3.1 Computing optimal a priori solutions

From expression (1), we have $X_i^{\mathtt{U1},S} = 0$, $\forall v_i \notin S$, and consequently, $\Pr[X_i^{\mathtt{U1},S} = 1] = 0$, for $v_i \in V \setminus S$; so, the following theorem is immediately derived for strategy U1.

Theorem 2. *Given a graph $G = (V, E)$, an a priori solution S and the modification strategy* U1, *then $E_S^{\mathtt{U1}} = \sum_{v_i \in S} p_i$, and is computed in $O(n)$. Optimal PIS1-solution $\hat{S}$ is a maximum-weight independent set in a weighted version of G where vertices are weighted by the corresponding probabilities. If $p_i = p$, $\forall v_i \in V$, then $E_S^{\mathtt{U1}} = p\alpha(G)$; in this case, $\hat{S} = S^*$ and $E_{\hat{S}}^{\mathtt{U1}} = p\alpha^*(G)$.*

The characterization of $\hat{S}$ given in theorem 2 immediately introduces the following complexity result for PIS1.

Theorem 3. *PIS1 is* **NP**-*hard.*

We now show that for $p_i = p$, $v_i \in V$, a mathematical expression for $E_S^{\mathtt{U1}}$ can be built directly without applying theorem 1 (used in next sections for the analysis of other strategies). Given that $0 \leqslant |\mathcal{F}(I_S^{\mathtt{U1}})| \stackrel{\text{def}}{=} F(I_S^{\mathtt{U1}}) = |S[I]| \leqslant \alpha(G)$, we get:

$$F\left(I_S^{\mathtt{U1}}\right) = |S[I]| \sum_{i=1}^{\alpha(G)} 1_{\{|S[I]|=i\}}.$$

So, the functional for U1 can be written as

$$\begin{aligned} E_S^{\mathtt{U1}} &= \sum_{I \subseteq V} \Pr[I] |S[I]| \sum_{i=1}^{\alpha(G)} 1_{\{|S[I]|=i\}} = \sum_{i=1}^{\alpha(G)} i \sum_{I \subseteq V} \Pr[I] 1_{\{|S[I]|=i\}} \\ &= \sum_{i=1}^{\alpha(G)} i \binom{\alpha(G)}{i} p^i (1-p)^{\alpha(G)-i} \sum_{j=0}^{n-\alpha(G)} \binom{n-\alpha(G)}{j} p^j (1-p)^{n-\alpha(G)-j} = p\alpha(G) \end{aligned}$$

where in the last summation we count all the sub-graphs $G[I]$ such that $|S[I]| = i$ and we add their probabilities; also, $\sum_{j=0}^{n-\alpha(G)} C_j^{n-\alpha(G)} p^j (1-p)^{n-\alpha(G)-j} = 1$. The above proof for $E_S^{\mathtt{U1}}$ can be generalized in order to compute every moment of any order for U1. For instance,

$$E\left[\left(G_S^{\mathtt{U1}}\right)^2\right] = \sum_{I \subseteq V} \Pr[I] F^2\left(I_S^{\mathtt{U1}}\right) = \sum_{i=1}^{\alpha(G)} i^2 \binom{\alpha(G)}{i} p^i (1-p)^{\alpha(G)-i} = p\alpha(G)(p\alpha(G) + 1 - p)$$

and, consequently,

$$\mathrm{Var}_S^{\mathtt{U1}} = E\left[\left(G_S^{\mathtt{U1}}\right)^2\right] - \left(E_S^{\mathtt{U1}}\right)^2 = \alpha(G)p(1-p).$$

So, for U1, the random variable representing the size $\alpha(G)$ of the a priori solution follows a binomial law with parameters $\alpha(G)$ and p.

3.2 Approximating optimal solutions for PIS1

In this section we show how, even if one cannot compute the optimal a priori solution in polynomial time, one can compute a sub-optimal solution, the value (expectation) of which is always greater than a factor times the value (expectation) of the optimal one. For this, we shall propose in what follows well-known (in the theory of polynomial approximation of **NP**-complete problems) polynomial algorithms computing "good" sub-optimal solutions, and will show that, also in probabilistic case, these algorithms work well.

Let us first recall that given an instance P of an **NP**-complete problem Π, a common way to estimate the capacity of a polynomial approximation algorithm A in finding good sub-optimal solutions for Π, is by evaluating its approximation ratio ([8]), i.e., the ratio $\mathtt{A}(P)/\mathrm{OPT}(P)$, where $\mathtt{A}(P)$ denotes the value of the solution computed by A on P and $\mathrm{OPT}(P)$ denotes the value of the optimal solution of P. Then the approximation ratio of A for a maximization problem Π is the quantity $\inf\{\mathtt{A}(P)/\mathrm{OPT}(P) : P \text{ instance of } \Pi\}$; the closer this ratio to 1, the better the approximation algorithm.

Recall that as we have already seen in section 3, PIS1 is equivalent to a weighted IS-problem, where each vertex is weighted by the corresponding probability. Consequently, the following theorem holds immediately.

Theorem 4. *If there exists a polynomial time approximation algorithm A solving WIS within approximation ratio ρ, then A polynomially solves PIS1 within the same approximation ratio ρ.*

In [6], an algorithm is developed for WIS achieving approximation ratio the minimum value between $\log n/(3(\Delta_G + 1)\log\log n)$ and $O(n^{-4/5})$. Using this algorithm in theorem 4, one gets the following corollary.

Corollary 1. *PIS1 can be approximated within*

$$\min\left\{\frac{\log n}{3(\Delta_G + 1)\log\log n}, O\left(n^{-4/5}\right)\right\}$$

The characterization of PIS1 in terms of a weighted IS-problem draws not only issues for finding reasonable a priori sub-optimal solutions but, unfortunately, limits the capacity of the problem to be "well-approximated" since, via this characterization, all the negative results applying to IS are immediately transferred to PIS1 also. So, PIS1 is hard to approximate within $n^{1-\epsilon}$, for any $\epsilon > 0$ ([10]).

3.3 PIS1 in bipartite graphs

Theorem 5. *Consider a bipartite graph $B = (V_1, V_2, E_B)$. Then, $E_S^{U1} = \sum_{v_i \in S} p_i$ and is computed in polynomial time. The optimal a priori solution $\hat{S}$ is a maximum-weight independent set in B considering that its vertices are weighted by the corresponding presence probabilities, and can be found in $O(n^{1/2}|E_B|)$. Consequently, in bipartite graphs,* PIS1 $\in \mathbf{P}$.

Proof. Concerning the expression for E_S^{U1} and the complexity of its computation, the proof is the same as the one of theorem 2.

Determining an optimal IS-solution in a bipartite graph is of polynomial complexity in both weighted and unweighted cases (see [9] for the unweighted case; for the weighted one, we quote here the result of [5] where the polynomiality of weighted IS in a class of graphs including the bipartite ones is proved). ∎

4 The complexities of PIS2 and PIS3

4.1 Expressions for E_S^{U2} and E_S^{U3}

Let $A_i = \sum_{I \subseteq V, |S \cap I| = i} \Pr[I] F(I_S^{U2})$. Then, E_S^{U2} can be nicely written as follows:

$$E_S^{U2} = \sum_{I \subseteq V} \Pr[I] F\left(I_S^{U2}\right) = \sum_{I \subseteq V} \left(\sum_{i=0}^{\alpha(G)} 1_{\{|S \cap I| = i\}} \right) \Pr[I] F(I_S^{U2}) = \sum_{i=0}^{\alpha(G)} \left(\sum_{\substack{I \subseteq V \\ |S \cap I| = i}} \Pr[I] F\left(I_S^{U2}\right) \right) = \sum_{i=0}^{\alpha(G)} A_i.$$

Quantities A_i, $i = 1, \ldots, \alpha(G)$ ($A_0 = 0$) are very natural and interesting from both theoretical and practical points of view. For instance, formula for E_S^{U2} given by expression above holds for every probability law; also, computing analytical expressions for A_i seems to be an interesting problem in combinatorial counting of graphs; moreover, thanks to the simple relation between E_S^{U2} and A_i, $i = 1, \ldots, \alpha(G)$, analytical expressions for the latter would produce explicit expressions for the former. Unfortunately, expression above for E_S^{U2}, even intuitive and smart, does not give any hint allowing precise characterization of $\hat{S}$.

In proposition 1, the proof of which is given in appendix A, $A_{\alpha^*(G)}$ and $A_{\alpha^*(G)-1}$ are explicitly computed, for the case of identical vertex-probabilities. However, the explicit computation for A_is of lower index produces very long and non-intuitive expressions.

Proposition 1. *Let $\Gamma'(v) = \Gamma(v) \setminus \{\Gamma(v) \cap \Gamma(S^* \setminus \{v\})\}$, $\ell_1 = |\{v \in S^* : \Gamma'(v) = \emptyset\}|$, $\ell_1' = \alpha^*(G) - \ell_1$, and $p_i = p$, $\forall v_i \in V$. Then,*

$$\begin{aligned} A_{\alpha^*(G)} &= \alpha^*(G) p^{\alpha^*(G)} \\ A_{\alpha^*(G)-1} &= p^{\alpha^*(G)-1}(1-p) \left(\alpha^*(G)\ell_1 - \ell_1 + \alpha^*(G)\ell_1' - \sum_{\substack{v \in S^* \\ \Gamma'(v) \neq \emptyset}} (1-p)^{|\Gamma'(v)|} \right). \end{aligned}$$

We shall now give an upper bound for the complexity of computing E_S^{U2}. For this, we will analyze (as an intermediate step) strategy U3 introduced in section 2.

Let $G' = G[V \setminus S] = (V', E')$, and let $V' = \{v_1, \ldots, v_{n-\alpha(G)}\}$ be the list of vertices of G' sorted in increasing-degree order; let us denote by V_i the set of the i first vertices of V' and let $G_i' = G'[V_i]$ (of course, for G_i' the vertices of V_i are not sorted in increasing-degree order). Let us denote by S_i the independent set found by U3 on (the present sub-instance of) G_i', and by α_i its cardinality, $i = 1, \ldots, n - \alpha(G)$. Finally, let $\alpha(G[I'])$ be the cardinality of the solution provided by U3 when applied in graph $G[I']$, $I' \subseteq V$.

Theorem 6.

$$E\left(\alpha_{n-\alpha(G)}\right) = \sum_{i=1}^{n-\alpha(G)} p_i \Pr\left[v_i \notin \Gamma(S_{i-1})\right]$$
$$E_S^{\mathtt{U3}} = E_S^{\mathtt{U1}} + E\left(\alpha_{n-\alpha(G)}\right).$$

If we denote by $T(E(\alpha_{n-\alpha(G)}))$ and $T(E_S^{\mathtt{U3}})$ the computation times of $E(\alpha_{n-\alpha(G)})$ and $E_S^{\mathtt{U3}}$, respectively, then $T(E(\alpha_{n-\alpha(G)})) = O(2^{n-\alpha(G)})$ and $T(E_S^{\mathtt{U3}}) = O(2^{n-\alpha(G)})$.

Proof. $E(\alpha_i) = E(\alpha_i|v_i \text{ present})p_i + E(\alpha_i|v_i \text{ absent})(1-p_i)$. Moreover, for strategy U3 we have the following relation, setting $S_0 = \emptyset$, $\Gamma(S_0) = \emptyset$ and $\alpha_0 = 0$:

$$\alpha_i = \begin{cases} \alpha_{i-1} & \Gamma(v_i) \cap S_{i-1} \neq \emptyset \\ \alpha_{i-1} + 1 & \text{otherwise} \end{cases}$$

So, $E(\alpha_i|v_i \text{ present}) = E(\alpha_{i-1}) + \Pr[\Gamma(v_i) \cap S_{i-1} = \emptyset]$; consequently,

$$E\left(\alpha_i\right) = p_i E\left(\alpha_{i-1}\right) + p_i \Pr\left[v_i \notin \Gamma(S_{i-1})\right] + (1-p_i)\, E\left(\alpha_{i-1}\right).$$

Since $E(\alpha_0) = 0$,

$$E\left(\alpha_i\right) = \sum_{j=1}^{i} p_j \Pr\left[v_j \notin \Gamma(S_{j-1})\right]$$
$$E\left(\alpha_{n-\alpha(G)}\right) = \sum_{i=1}^{n-\alpha(G)} p_i \Pr\left[v_i \notin \Gamma(S_{i-1})\right].$$

Now, let $I' = f(I) = I \setminus \{S[I] \cup \Gamma(S[I])\}$. This set represents the subset of vertices of I which are not contained neither in S, nor in the neighbor-set of $S[I]$. Consequently, U3 will be applied on $G[I']$; so, $F(I_S^{\mathtt{U3}}) = |S[I]| + \alpha(G[I'])$ and consequently

$$E_S^{\mathtt{U3}} = \sum_{I \subseteq V} \Pr[I]|S[I]| + \sum_{I \subseteq V} \Pr[I]\alpha\left(G[I']\right) = E_S^{\mathtt{U1}} + \sum_{I \subseteq V} \Pr[I]\alpha\left(G[I']\right)$$
$$= E_S^{\mathtt{U1}} + \sum_{I \subseteq V} \left(\sum_{I' \subseteq V} 1_{\{f(I)=I'\}} \right) \Pr[I]\alpha\left(G[I']\right) = E_S^{\mathtt{U1}} + \sum_{I' \subseteq V} \sum_{\substack{I \subseteq V \\ f(I)=I'}} \Pr[I]\alpha\left(G[I']\right).$$

Since $\Pr[I'] = \sum_{I \subseteq V, f(I)=I'} \Pr[I]$, we get

$$E_S^{\mathtt{U3}} = E_S^{\mathtt{U1}} + \sum_{I' \subseteq V} \Pr\left[I'\right] \alpha\left(G[I']\right) = E_S^{\mathtt{U1}} + E\left(\alpha_{n-\alpha(G)}\right).$$

Let us now introduce the random variable C_i representing the solution of PIS3 whenever we consider only the present vertices of V_i, i.e., when we apply the greedy algorithm implied by strategy U3 in the graph induced by the present vertices of V_i.

$$C_i = \begin{cases} C_{i-1} & v_i \text{ is absent} \\ C_{i-1} & v_i \text{ is present and } \Gamma(v_i) \cap C_{i-1} \neq \emptyset \\ C_{i-1} \cup \{v_i\} & v_i \text{ is present and } \Gamma(v_i) \cap C_{i-1} = \emptyset \end{cases}$$

We then have $E(\alpha_{n-\alpha(G)}) = \sum_{i=1}^{n-\alpha(G)} p_i \Pr[\Gamma(v_i) \cap C_{i-1} = \emptyset]$.

In order to prove the result of the theorem, we prove that the quantity $\Pr[\Gamma(v_i) \cap C_{i-1} = \emptyset]$ is not computable in polynomial time.

$$\begin{aligned}
\Pr[\Gamma(v_i) \cap C_{i-1} = \emptyset] &= (1 - p_{i-1}) \Pr[\Gamma(v_i) \cap C_{i-2} = \emptyset] \\
&\quad + p_{i-1} \Pr[\Gamma(v_i) \cap C_{i-2} = \emptyset] \Pr[\Gamma(v_{i-1}) \cap C_{i-2} \neq \emptyset] \\
&\quad + p_{i-1} \Pr[\Gamma(v_i) \cap (C_{i-2} \cup \{v_{i-1}\}) = \emptyset] \Pr[\Gamma(v_{i-1}) \cap C_{i-2} = \emptyset] \\
&= (1 - p_{i-1}) \Pr[\Gamma(v_i) \cap C_{i-2} = \emptyset] \\
&\quad + p_{i-1} \Pr[\Gamma(v_i) \cap C_{i-2} = \emptyset] (1 - \Pr[\Gamma(v_{i-1}) \cap C_{i-2} = \emptyset]) \\
&\quad + p_{i-1} \Pr[\Gamma(v_i) \cap (C_{i-2} \cup \{v_{i-1}\}) = \emptyset] \Pr[\Gamma(v_{i-1}) \cap C_{i-2} = \emptyset] \\
&= \Pr[\Gamma(v_i) \cap C_{i-2} = \emptyset] \\
&\quad - p_{i-1} \Pr[\Gamma(v_i) \cap C_{i-2} = \emptyset] \Pr[\Gamma(v_{i-1}) \cap C_{i-2} = \emptyset] \\
&\quad + p_{i-1} \Pr[\Gamma(v_i) \cap (C_{i-2} \cup \{v_{i-1}\}) = \emptyset] \Pr[\Gamma(v_{i-1}) \cap C_{i-2} = \emptyset] .
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\Pr[\Gamma(v_i) \cap (C_{i-2} \cup \{v_{i-1}\}) = \emptyset] &= \Pr[(\Gamma(v_i) \cap C_{i-2}) \cup (\Gamma(v_i) \cap \{v_{i-1}\}) = \emptyset] \\
&= \Pr[(\Gamma(v_i) \cap C_{i-2} = \emptyset) \cap (\Gamma(v_i) \cap \{v_{i-1}\} = \emptyset)] \\
&= \Pr[\Gamma(v_i) \cap C_{i-2} = \emptyset] \, 1_{\{\Gamma(v_i) \cap \{v_{i-1}\} = \emptyset\}} .
\end{aligned}$$

Consequently,

$$\begin{aligned}
\Pr[\Gamma(v_i) \cap C_{i-1} = \emptyset] &= \Pr[\Gamma(v_i) \cap C_{i-2} = \emptyset] \\
&\quad - p_{i-1} \Pr[\Gamma(v_i) \cap C_{i-2} = \emptyset] \Pr[\Gamma(v_{i-1}) \cap C_{i-2} = \emptyset] \\
&\quad + p_{i-1} \Pr[\Gamma(v_i) \cap C_{i-2} = \emptyset] \Pr[\Gamma(v_{i-1}) \cap C_{i-2} = \emptyset] \, 1_{\{\Gamma(v_i) \cap \{v_{i-1}\} = \emptyset\}} .
\end{aligned}$$

Let $t_{i-1}(v_i)$ be the computational time of $\Pr[\Gamma(v_i) \cap C_{i-1} = \emptyset]$. By the above equalities we easily deduce that

$$t_{i-1}(v_i) = t_{i-2}(v_i) + t_{i-2}(v_{i-1}) .$$

In order to compute this recurrence relation, at each step we need to know two terms of the precedent step; so for the computation of $\Pr[\Gamma(v_i) \cap C_{i-1} = \emptyset]$, we need 2^{i-1} computations and expression for $T(E_S^{\mathtt{U3}})$ immediately follows. ∎

Algorithm U3 is, as it has been already noted, a simplified version of algorithm U2. Moreover, there exist graphs where the two algorithms give the same results by performing identical choices and deletions of vertices (for example, consider a graph on n isolated vertices). Consequently, computation time of U3 is a (worst-case) lower bound for the one of U2 and the following theorem holds.

Theorem 7. *Let $T(E_S^{\mathtt{U2}})$ be the computational time of $E_S^{\mathtt{U2}}$. Then, $T(E_S^{\mathtt{U2}}) = \Omega(T(E_S^{\mathtt{U3}}))$.*

The result of theorem 6 simply gives an upper bound on the complexity of computing $E_S^{\mathtt{U3}}$ and does not prove that $E_S^{\mathtt{U3}}$ is not computable in polynomial time (if this was true, it would be a very interesting result since, in this case, PIS2 and PIS3 would not belong to **NP**). In fact, the result of theorem 6 is based upon a particular recursion-formula and a particular way for computing it. In any case, one can easily prove that PIS2 is intractable (following the notation in the appendix of [8], PIS2 is a kind of starry problem).

Indeed, if one can polynomially determine an optimal a priori solution $\hat{S}$ for PIS2, then one can simply consider an instance of IS as a PIS2-instance with $p_i = 1$, $\forall v_i \in V$. It is easy to see that, in this case, $\hat{S} = S^*$ and the following theorem immediately holds.

Theorem 8. *Unless **P=NP**, PIS2 is computationally intractable.*

4.2 Bounds for E_S^{U2}

For lack of characterizing the complexity of computing E_S^{U2}, we build in this paragraph upper and lower bounds for it.

Theorem 9. *Let $\tilde{\Delta}$ be the maximum degree of $G[\tilde{I}]$. Then, on the hypothesis of distinct vertex-probabilities:*

$$\sum_{v_i \in S} p_i + \sum_{I \subseteq V} \prod_{v_i \in I} p_i \prod_{v_i \notin I} (1-p_i) \frac{\left|\tilde{I}\right|}{\tilde{\Delta}+1} \leqslant E_S^{U2} \leqslant \sum_{v_i \in S} p_i + \sum_{I \subseteq V} \prod_{v_i \in I} p_i \prod_{v_i \notin I} (1-p_i) \left|\tilde{I}\right| \quad (4)$$

while, on the hypothesis of identical vertex-probabilities:

$$p\alpha(G) + \sum_{I \subseteq V} p^{|I|}(1-p)^{n-|I|} \frac{\left|\tilde{I}\right|}{\tilde{\Delta}+1} \leqslant E_S^{U2} \leqslant p\alpha(G) + \sum_{I \subseteq V} p^{|I|}(1-p)^{n-|I|} \left|\tilde{I}\right| \quad (5)$$

Always under the latter hypothesis: $pn/(\Delta_G+1) \leqslant E_S^{U2} \leqslant n(\Delta_G+p)/(\Delta_G+1)$.

Proof. Remark first that $|S[I]| \leqslant F(I_S^{U2}) \leqslant |S[I]| + |I \setminus (S[I]) \cup (\Gamma(S[I]) \cap I)| = |S[I]| + |\tilde{I}|$; consequently,

$$E_S^{U2} \leqslant \sum_{I \subseteq V} \Pr[I] \left(|S[I]| + \left|\tilde{I}\right|\right) \leqslant E_S^{U1} + \sum_{I \subseteq V} \Pr[I] \left|\tilde{I}\right| = \sum_{v_i \in S} p_i + \sum_{I \subseteq V} \Pr[I] \left|\tilde{I}\right|. \quad (6)$$

On the other hand, $\Pr[I] = \prod_{v_i \in I} p_i \prod_{v_i \notin I} (1-p_i)$. The upper bound results from the combination of expression (6) and the one for $\Pr[I]$.

We now prove the lower bound for E_S^{U2}. Let $\alpha(G[\tilde{I}])$ be the cardinality of the solution provided by U2 when applied to $G[\tilde{I}]$; we then have

$$E_S^{U2} = \sum_{v_i \in S} p_i + \sum_{I \subseteq V} \Pr[I] \alpha\left(G\left[\tilde{I}\right]\right) = \sum_{v_i \in S} p_i + \sum_{I \subseteq V} \prod_{v_i \in I} p_i \prod_{v_i \notin I} (1-p_i) \alpha\left(G\left[\tilde{I}\right]\right) \quad (7)$$

For $\alpha(G[\tilde{I}])$, since the greedy algorithm implied by U2 provides a maximal independent set, the following holds ([1]): $\alpha(G[\tilde{I}]) \geqslant |\tilde{I}|/(\tilde{\Delta}+1)$. By substituting the expression for $\alpha(G[\tilde{I}])$ in expression (7), we obtain the lower bound claimed.

In the case where all the vertices have the same presence probability p, $\sum_{v_i \in S} p_i = p\alpha(G)$ and $\Pr[I] = p^{|I|}(1-p)^{n-|I|}$, and expression (5) follows immediately.

In order to obtain bounds implied by the last expression of the theorem (always assuming identical occurrence probabilities), we use inequality $\alpha(G) \geqslant n/(\Delta_G+1)$. Moreover, $\sum_{I \subseteq V} p^{|I|}(1-p)^{n-|I|} = 1$ (so, $\sum_{I \subseteq V} p^{|I|}(1-p)^{n-|I|} \geqslant 0$), and

$$\left|\tilde{I}\right| = |I \setminus \{S[I] \cup \Gamma(S[I])\}| \leqslant |I \setminus S[I]| = |I \setminus (S \cap I)| \leqslant |V \setminus (S \cap V)| = |V \setminus S| = n - \alpha(G).$$

So, $\sum_{I \subseteq V} p^{|I|}(1-p)^{n-|I|} |\tilde{I}| \leqslant (n-\alpha(G)) \sum_{I \subseteq V} p^{|I|}(1-p)^{n-|I|} = n - \alpha(G)$ and combining the above inequalities, we obtain the claimed bounds. ∎

4.3 Approximating optimal solutions for PIS2

4.3.1 Using argmax$\{\sum_{v_i \in S} p_i\}$ as a priori solution

Set $\tilde{S} = \text{argmax}\{\sum_{v_i \in S} p_i : S \text{ independent set of } G\}$ and suppose that it is used as a priori solution for PIS2. Then, the following holds.

Theorem 10. *Approximation of $\hat{S}$ by $\bar{S} = \text{argmax}\{\sum_{v_i \in S} p_i : S \text{ independent set of } B\}$ guarantees for PIS2 approximation ratio*

$$\max\left\{\frac{p_{\min}}{1+p_{\max}}, \frac{1}{\Delta_G + 1}\right\}.$$

Proof. By expression (7) (proof of theorem 9), and since $\alpha(G[\tilde{I}]) \leqslant \alpha^*(G)$, we get

$$\sum_{v_i \in S} p_i \leqslant E_S^{U2} = \sum_{v_i \in S} p_i + \sum_{I \subseteq V} \Pr[I]\alpha\left(G[\tilde{I}]\right) \leqslant \alpha^*(G)\left(p_{\max} + \sum_{I \subseteq V} \Pr[I]\right) = \alpha^*(G)\left(1 + p_{\max}\right) \quad (8)$$

Since, $\bar{S} = \text{argmax}\{\sum_{v_i \in S} p_i : S \text{ independent set of } G\}$, then

$$E_{\bar{S}}^{U2} \geqslant \sum_{v_i \in \bar{S}} p_i \geqslant \sum_{v_i \in S^*} p_i \geqslant p_{\min}\alpha^*(G) \quad (9)$$

Remark now that expression (8) holds also for $E_{\hat{S}}^{U2}$; consequently, combining expressions (8) and (9) we obtain

$$\frac{E_{\bar{S}}^{U2}}{E_{\hat{S}}^{U2}} \geqslant \frac{p_{\min}\alpha^*(G)}{\left(1+p_{\max}\right)\alpha^*(G)} = \frac{p_{\min}}{1+p_{\max}} \quad (10)$$

On the other hand, remark that from expression (2)

$$E_{\hat{S}}^{U2} \leqslant \sum_{v_i \in V} p_i \quad (11)$$

and from the lefthand-side of expression (4)

$$E_{\bar{S}}^{U2} \geqslant \sum_{v_i \in \bar{S}} p_i \quad (12)$$

Combination of expressions (11) and (12) gives

$$\frac{E_{\bar{S}}^{U2}}{E_{\hat{S}}^{U2}} \geqslant \frac{\sum_{v_i \in \bar{S}} p_i}{\sum_{v_i \in V} p_i} \geqslant \frac{1}{\Delta_G + 1} \quad (13)$$

where the last inequality (remark that $\bar{S}$ is maximal) is the weighted version of Turàn's theorem ([16]).

Expressions (10) and (13) conclude the theorem. ∎

4.3.2 Polynomial time approximations for PIS2

The set $\bar{S}$ considered in the previous paragraph cannot be computed in polynomial time. Instead, suppose that one uses a polynomial time approximation algorithm A (achieving approximation ratio ρ) for (unweighted) IS in order to compute a solution SOL (obviously, we can suppose that SOL is maximal) on G where vertex-probabilities are omitted. Then, expression (9) in the proof of theorem 10 becomes

$$E_{\text{SOL}}^{U2} \geqslant p_{\min}|\text{SOL}| \geqslant p_{\min}\rho\alpha^*(G)$$

and with exactly the same arguments as in theorem 10, the following theorem can be proved.

Theorem 11. *If there exists a polynomial time approximation algorithm $\mathcal{A}$ solving IS within approximation ratio ρ, then algorithm $\mathcal{A}$ polynomially solving PIS2 within approximation ratio*

$$\max\left\{\frac{1}{\Delta_G+1}, \left(\frac{p_{\min}}{1+p_{\max}}\right)\rho\right\}.$$

If $\mathcal{A}$ is the algorithm of [6], then:

- *in the case of fixed vertex-probabilities, PIS2 can be approximately solved in polynomial time within ratio*

$$\min\left\{O\left(\frac{\log n}{3(\Delta_G+1)\log\log n}\right), O\left(n^{-4/5}\right)\right\}$$

- *in the case where probabilities depend on n, PIS2 is polynomially approximable within ratio*

$$\max\left\{\frac{1}{\Delta_G+1}, \left(\frac{p_{\min}}{1+p_{\max}}\right)\min\left\{O\left(\frac{\log n}{3(\Delta_G+1)\log\log n}\right), O\left(n^{-4/5}\right)\right\}\right\}.$$

4.4 PIS2 in bipartite graphs

Theorem 12. *Consider a bipartite graph $B=(V_1,V_2,E_B)$. Then,*

$$E^{\mathtt{U2}}_{V_1} = \sum_{v_i\in V_1} p_i + \sum_{v_i\in V_2} p_i \prod_{v_j\in\Gamma(v_i)}(1-p_j)$$

and can be computed in polynomial time. Consequently, in bipartite graphs, whenever color-class V_1 (or V_2) is considered as a priori solution, PIS2 $\in$ **NP**.

Proof. Let us first note the following:

- if all the vertices of V_1 are absent, then the solution provided by U2 is exactly the present vertices of the color-class V_2;
- if all the vertices of V_1 are present, then, despite the state of the set V_2, the solution of the present sub-instance of B is exactly the color-class V_1;
- in the case where a part of the vertices of V_1 is present, the final solution for $B[I]$ will eventually include some vertices of V_2.

Applying the result of theorem 1, we get:

$$E^{\mathtt{U2}}_{V_1} = \sum_{v_i\in V_1} p_i + \sum_{v_i\in V_2} \Pr\left[X_i^{\mathtt{U2},V_1}=1\right] \tag{14}$$

(recall that $\Pr[X_i^{\mathtt{U2},S}=1]$ represents the probability that vertex $v_i\notin S$ will be chosen when applying U2).

Note also that $\Pr[X_i^{\mathtt{U2},V_1}=1]$, $v_i\in V_2$, depends only on the present vertices of $\Gamma(v_i)$; consequently, it does not depend on the other elements of V_2. Henceforth, insertion of the elements of V_2 is performed independently the ones from the others and $v_i\in V_2$ will be introduced in the solution for $B[I]$ only if $\Gamma(v_i)\cap V_1[I]=\emptyset$. So, $\Pr[X_i^{\mathtt{U2},V_1}=1]=p_i\prod_{v_j\in\Gamma(v_i)}(1-p_j)$.

Replacing this expression for $\Pr[X_i^{\mathtt{U2},V_1}=1]$ in expression (14), we obtain the result claimed for $E^{\mathtt{U2}}_{V_1}$. One can see that this expression implies the computation of $E^{\mathtt{U2}}_{V_1}$ in at most $O(n^2)$ steps. ∎

From the proof of theorem 12, one can see how the particular structure of the bipartite graph intervenes in a significant way to simplify the expression for the functional and, consequently, its computation. Expression (14) holds thanks to the fact that the vertex set of B can be partitioned into two independent sets.

Corollary 2. *Suppose* $\Pr[v_i] = p$, $v_i \in V_1 \cup V_2$, *denote by* n_1 *and* n_2 *the sizes of* V_1 *and* V_2, *respectively, and suppose that* $n_1 \geqslant n_2$. *Then,* $E^{\mathtt{U2}}_{V_1} = pn_1 + p\sum_{v_i \in V_2}(1-p)^{|\Gamma(v_i)|}$. *Naturally,* $E^{\mathtt{U2}}_{V_1}$ *is computed in* $O(n^2)$.

From corollary 2 we can obtain the following framing of $E^{\mathtt{U2}}_{V_1}$ by $E^{\mathtt{U1}}_{V_1}$ for the case of identical vertex-probabilities: $E^{\mathtt{U1}}_{V_1} + n_2 p(1-p)^{\Delta_B} \leqslant E^{\mathtt{U2}}_{V_1} \leqslant E^{\mathtt{U1}}_{V_1} + n_2 p(1-p)^{\delta_B}$. So, in regular bipartite graphs (i.e., the ones where $\Delta_B = \delta_B = \Delta$): $E^{\mathtt{U2}}_{V_1} = E^{\mathtt{U1}}_{V_1} + n_2 p(1-p)^{\Delta}$.

Consider set $\bar{S} = \operatorname{argmax}\{\sum_{v_i \in S} p_i : S \text{ independent set of } B\}$ as a priori solution. Then, considering vertex-probabilities as vertex-weights and using the result of [5], $\bar{S}$ can be computed in polynomial time. Moreover,

$$E^{\mathtt{U2}}_{\bar{S}} \geqslant \sum_{v_i \in \bar{S}} p_i \geqslant \frac{\sum_{v_i \in V} p_i}{2}.$$

Using the expression above together with expression (11), approximation ratio 1/2 is immediately yielded.

Proposition 2. $\bar{S} = \operatorname{argmax}\{\sum_{v_i \in S} p_i : S \text{ independent set of } B\}$ *is a polynomial time approximation of* $\hat{S}$ *guaranteeing approximation ratio 1/2 for PIS2 in bipartite graphs.*

Obviously, from theorem 12 and the discussion just above, the same approximation ratio can be yielded if one uses $\operatorname{argmax}\{\sum_{v_i \in V_1} p_i, \sum_{v_i \in V_2} p_i\}$ as a priori solution.

5 The complexity of PIS4

5.1 An expression for $E^{\mathtt{U4}}_S$

Recall that strategy U4 starts from $S[I]$ and completes it with the isolated vertices of the graph $G[\tilde{I}]$.

Proposition 3. *Given a graph* $G = (V, E)$, *an a priori independent set* S *and the modification strategy U4, then*

$$E\left(G^{\mathtt{U4}}_S\right) = \sum_{v_i \in S} p_i + \sum_{v_i \in (V \setminus S)} p_i \prod_{v_j \in \Gamma_S(v_i)} (1-p_j) \times \prod_{v_k \in \Gamma_{V \setminus S}(v_i)} \left((1-p_k) + p_k \left(1 - \prod_{v_l \in \Gamma_S(v_k)} (1-p_l) \right) \right) \tag{15}$$

where $\Gamma_S(v_i) = \Gamma(v_i) \cap S$ *and* $\Gamma_{V \setminus S}(v_i) = \Gamma(v_i) \cap (V \setminus S)$. $E(G^{\mathtt{U4}}_S)$ *can be computed in polynomial time. If* $p_i = p$, *for all* $v_i \in V$, *then*

$$E\left(G^{\mathtt{U4}}_S\right) = p\alpha(G) + \sum_{v_i \in (V \setminus S)} p(1-p)^{|\Gamma_S(v_i)|} \prod_{v_k \in \Gamma_{V \setminus S}(v_i)} \left(1 - p(1-p)^{|\Gamma_S(v_k)|}\right) \tag{16}$$

Proof. Set $B_i = \Pr[X_i^{\mathcal{U}4,S} = 1]$. Then

$$B_i = \sum_{\substack{I \subseteq V \\ v_i \in I}} \Pr[I] 1_{\{v_i \in \mathcal{F}(I^{\mathcal{U}4}_S)\}}$$

Let v_i be any vertex of $V \setminus S$ and let I be any subset of V containing v_i. Obviously,

$$v_i \in \mathcal{F}(I^{\mathtt{U4}}_S) \iff v_i \text{ isolated in } G[\tilde{I}] \iff (v_i \in G[\tilde{I}]) \wedge (v_i \text{ has lost all its neighbors in } \tilde{I}).$$

Since $v_i \notin S$, $v_i \in \tilde{I}$ only if there does not belong to the neighborhood of any vertex in $S[I]$, i.e.,

$$v_i \in G[\tilde{I}] \iff \Gamma_S(v_i) \cap I = \emptyset \tag{17}$$

On the other hand, all the neighbors of v_i in $G[\tilde{I}]$ have been removed iff $\Gamma_{V\setminus S}(v_i) \cap \tilde{I} = \emptyset$. This last condition is satisfied only if every vertex of $\Gamma_{V\setminus S}(v_i)$ is either absent or (being present) has been removed from $\tilde{I}$ because it belonged to the neighborhood of a vertex in $S[I]$. In all,

$$\begin{aligned}\Gamma_{V\setminus S}(v_i) \cap \tilde{I} = \emptyset \iff\ & \forall v_j \in \Gamma_{V\setminus S}(v_i)\,((v_j \text{ is absent }) \vee \\ & ((v_j \text{ is present}) \wedge (\exists v_k \in \Gamma(v_j) \cap S \text{ such that } v_k \text{ is present})))\end{aligned} \tag{18}$$

From relations (17) and (18) and the discussion above we get ($v_i \in V \setminus S$):

$$\begin{aligned}
B_i &= \sum_{\substack{I\subseteq V\\ v_i\in I}} \Pr[I] 1_{\{v_i \in \mathcal{F}(I_S^{U4})\}} = \sum_{\substack{I\subseteq V\\ v_i\in I}} \Pr[I] 1_{\left\{\substack{v_i\in G[\tilde{I}]\\ \Gamma_{V\setminus S}(v_i)\cap \tilde{I}=\emptyset}\right\}} \\
&= \sum_{\substack{I\subseteq V\\ v_i\in I}} \Pr[I] 1_{\{v_i\in G[\tilde{I}]\}} 1_{\{\Gamma_{V\setminus S}(v_i)\cap\tilde{I}=\emptyset\}} = \sum_{I\subseteq V} \Pr[I] 1_{\left\{\substack{\{v_i\}\cap I=\{v_i\}\\ \Gamma_S(v_i)\cap I=\emptyset}\right\}} 1_{\{\Gamma_{V\setminus S}(v_i)\cap\tilde{I}=\emptyset\}} \\
&= \sum_{I\subseteq V} \Pr[I] 1_{\{\{v_i\}\cap I=\{v_i\}\}} 1_{\{\Gamma_S(v_i)\cap I=\emptyset\}} \prod_{v_j\in\Gamma_{V\setminus S}(v_i)} \left(1_{\{I\cap\{v_j\}=\emptyset\}} + 1_{\{I\cap\{v_j\}=\{v_j\}\}} 1_{\{I\cap\Gamma_S(v_j)\neq\emptyset\}}\right) \\
&= p_i \prod_{v_k\in\Gamma_S(v_i)} (1-p_k) \prod_{v_j\in\Gamma_{V\setminus S}(v_i)} \left((1-p_j) + p_j\left(1 - \prod_{v_l\in\Gamma_S(v_j)} (1-p_l)\right)\right)
\end{aligned} \tag{19}$$

Replacing the second term of expression (2) by expression (19) we easily obtain expression (15).

Moreover, one can see that computation of $E(G_S^{U4})$ in expression (15) takes at most n^2 multiplications. Also, setting $p_i = p$, $\forall v_i \in V$, we immediately obtain expression (16). ∎

5.2 Using S^* or argmax$\{\sum_{v_i\in S} p_i\}$ as a priori solutions

Expression (15) although polynomial does not allow precise characterization of the optimal a priori solution $\hat{S}$ associated with U4. In theorem 13 below, we restrict ourselves to the case of identical vertex probabilities and suppose that $\alpha^*(G)$, a maximum-size independent set of G is used as an a priori solution. Our objective is to estimate the ratio $E_{S^*}^{U4}/E_{\hat{S}}^{U4}$.

Theorem 13. *Under identical vertex-probabilities:*

$$\frac{E_{S^*}^{U4}}{E_{\hat{S}}^{U4}} \geqslant \frac{\alpha^*(G)}{n}.$$

The ratio $\alpha^*(G)/n$ is always bounded below by $1/(\Delta_G+1)$.

Proof. Set $|\Gamma_S(v_i)| = k_i$, $|\Gamma_{V\setminus S}(v_i)| = |\Gamma(v_i)| - k_i$. Then expression (16) becomes

$$E\left(G_S^{U4}\right) = p\alpha(G) + \sum_{v_i\in(V\setminus S)} p(1-p)^{k_i} \prod_{v_k\in\Gamma_{V\setminus S}(v_i)} \left(1-p(1-p)^{k_k}\right).$$

Also, note that, $\forall k$, $k_k \leqslant \Delta_G - 1$ (v_i and v_k are in $V\setminus S$ and $v_iv_k \in E$) and $E(G_S^{U4})$ is increasing in k_k. Then,

$$E\left(G_{\hat{S}}^{U4}\right) \leqslant p\hat{\alpha}(G) + \sum_{v_i\in(V\setminus\hat{S})} p(1-p)^{k_i} \prod_{v_k\in\Gamma_{V\setminus\hat{S}}(v_i)} \left(1-p(1-p)^{\Delta_G-1}\right)$$

$$
\begin{aligned}
&\leqslant && p\hat{\alpha}(G) + \sum_{v_i \in (V \setminus \hat{S})} p(1-p)^{k_i} \left(1 - p(1-p)^{\Delta_G - 1}\right)^{|\Gamma(v_i)| - k_i} \\
&\overset{1-p \leqslant 1-p(1-p)^{\Delta_G-1}}{\leqslant} && p\hat{\alpha}(G) + \sum_{v_i \in (V \setminus \hat{S})} p\left(1 - p(1-p)^{\Delta_G - 1}\right)^{|\Gamma(v_i)|} \\
&\overset{|\Gamma(v_i)| \geqslant 1}{\leqslant} && p\hat{\alpha}(G) + (n - \hat{\alpha}(G))\, p\left(1 - p(1-p)^{\Delta_G - 1}\right) \;\leqslant\; pn
\end{aligned} \tag{20}
$$

On the other hand, if we consider S^* as an a priori solution, we get

$$
\begin{aligned}
E\left(G^{\mathrm{U4}}_{S^*}\right) &= p\alpha^*(G) + \sum_{v_i \in (V \setminus S^*)} p(1-p)^{k_i} \prod_{v_k \in \Gamma_{V \setminus S^*}(v_i)} \left(1 - p(1-p)^{k_k}\right) \\
&\overset{1-p(1-p)^{k_k} \geqslant 1-p}{\geqslant} p\alpha^*(G) + \sum_{v_i \in (V \setminus S^*)} p(1-p)^{k_i} \prod_{v_k \in \Gamma_{V \setminus S^*}(v_i)} (1-p) \\
&= p\alpha^*(G) + \sum_{v_i \in (V \setminus S^*)} p(1-p)^{k_i} (1-p)^{|\Gamma(v_i)| - k_i} \\
&= p\alpha^*(G) + \sum_{v_i \in (V \setminus S^*)} p(1-p)^{|\Gamma(v_i)|} \\
&\geqslant p\alpha^*(G) + (n - \alpha^*(G))\, p(1-p)^{\Delta_G} \\
&= p\alpha^*(G)\left(1 - (1-p)^{\Delta_G}\right) + pn(1-p)^{\Delta_G}
\end{aligned} \tag{21}
$$

Combining expressions (20) and (21), we have

$$
\frac{E\left(G^{\mathrm{U4}}_{S^*}\right)}{E\left(G^{\mathrm{U4}}_{\hat{S}}\right)} \geqslant \frac{\alpha^*(G)}{n}\left(1 - (1-p)^{\Delta_G}\right) + (1-p)^{\Delta_G} \geqslant \frac{\alpha^*(G)}{n} \geqslant \frac{1}{\Delta_G + 1}. \quad \blacksquare
$$

For instance, if G is cubic, i.e., $\Delta_G = 3$, then $\alpha^*(G) \geqslant n/4$ then, $E(G^{\mathrm{U4}}_{S^*})/E(G^{\mathrm{U4}}_{\hat{S}}) \geqslant ((1/4)(1 - (1-p)^3)) + (1-p)^3 \geqslant 1/4$. If, in addition $p = 1/2$, then $E(G^{\mathrm{U4}}_{S^*})/E(G^{\mathrm{U4}}_{\hat{S}}) \geqslant 11/32$.

The result of theorem 13 can be easily extended in the case where vertex-probabilities are distinct and $\bar{S} = \mathrm{argmax}\{\sum_{v_i \in S} p_i : S \text{ independent set of } G\}$ is used as a priori solution. Moreover, without loss of generality, one can suppose that $\bar{S}$ is maximal. Then, from expression (15) one easily gets:

$$
E\left(G^{\mathrm{U4}}_{\hat{S}}\right) \leqslant \sum_{v_i \in \hat{S}} p_i + \sum_{v_i \in (V \setminus \hat{S})} p_i = \sum_{v_i \in V} p_i \tag{22}
$$

$$
E\left(G^{\mathrm{U4}}_{\bar{S}}\right) \geqslant \sum_{v_i \in \bar{S}} p_i \tag{23}
$$

Combining the above expressions we obtain

$$
\frac{E\left(G^{\mathrm{U4}}_{\bar{S}}\right)}{E\left(G^{\mathrm{U4}}_{\hat{S}}\right)} \geqslant \frac{\sum_{v_i \in \bar{S}} p_i}{\sum_{v_i \in V} p_i} \geqslant \frac{1}{\Delta_G + 1}
$$

where the last inequality is the weighted version of Turàn's theorem.

Finally, let us note that the same approximation ratio can be obtained if one treats vertex probabilities as weights and uses as a priori solution the one computed by the greedy IS-algorithm.

In the weighted case, this algorithm iteratively chooses the vertex maximizing the ratio "vertex-weight over vertex-degree" and eliminates its neighbors. In this case, if S' is the independent set computed we have ([15])

$$E\left(G_{S'}^{U4}\right) \geqslant \sum_{v_i \in S'} p_i \geqslant \frac{\sum_{v_i \in V} p_i}{\Delta_G + 1}$$

and using expression (22), approximation ratio bounded below by $1/(\Delta_G + 1)$ is immediately concluded.

5.3 PIS4 in bipartite graphs

The particular structure of a bipartite graph, denoted as previously by $B = (V_1, V_2, E)$, does not allow refinement of the result of proposition 3 in order to obtain a better characterization of the a priori solution maximizing $E(G_S^{U4})$ and of the complexity of its computation.

However, if $\text{argmax}\{|V_1|, |V_2|\}$ is used as a priori solution, then expression (15) can be simplified. Plainly, let us revisit it and suppose, without loss of generality, that $V_1 = \text{argmax}\{|V_1|, |V_2|\}$. So, we have $S = V_1$ and $V \setminus S = V_2$. Consequently, for $v_i \in V_2$, $\Gamma(v_i) \cap V_2 = \Gamma_{V \setminus S}(v_i) = \emptyset$ and the term $\prod_{v_k \in \Gamma_{V \setminus S}(v_i)}((1-p_k)+p_k(1-\prod_{v_l \in \Gamma_S(v_k)}(1-p_l)))$ (the last product of expression (15)), computed on an empty set takes, by convention, value 1. So, dealing with bipartite graphs, expression (15) becomes

$$E\left(B_{V_1}^{U4}\right) = \sum_{v_i \in V_1} p_i + \sum_{v_i \in V_2} p_i \prod_{v_j \in \Gamma_{V_1}(v_i)} (1 - p_j) \tag{24}$$

In what follows, we will prove that if we use the color-class $\text{argmax}\{\sum_{v_i \in V_1} p_i, \sum_{v_i \in V_2} p_i\}$ as a priori solution for PIS4, then it is solved within approximation ratio 1/2. In fact, let $V_1 = \text{argmax}\{\sum_{v_i \in V_1} p_i, \sum_{v_i \in V_2} p_i\}$. Then, by expression (24), we get

$$E\left(B_{V_1}^{U4}\right) \geqslant \sum_{v_i \in V_1} p_i \geqslant \frac{\sum_{v_i \in V_1 \cup V_2} p_i}{2} \tag{25}$$

Combining expression (25) with expression (22) we obtain

$$\frac{E\left(B_{V_1}^{U4}\right)}{E\left(B_{\hat{S}}^{U4}\right)} \geqslant \frac{\frac{\sum_{v_i \in V_1 \cup V_2} p_i}{2}}{\sum_{v_i \in V_1 \cup V_2} p_i} = \frac{1}{2}.$$

Of course, the same worst case approximation ratio is also achieved if one sees probabilities as weights and considers the maximum-weight independent set (of total weight at least equal to $\sum_{v_i \in V_1} p_i$; this set can be found in polynomial time ([5])) as a priori solution and the following theorem concludes the discussion above.

Theorem 14. *Sets* $\text{argmax}\{\sum_{v_i \in V_1} p_i, \sum_{v_i \in V_2} p_i\}$, *and* $\text{argmax}_{S \text{ independent set of } B}\{\sum_{v_i \in S} p_i\}$ *are polynomial approximations of PIS4 in bipartite graphs achieving approximation ratio bounded below by 1/2.*

Finally remark that for the case of identical vertex-probabilities, the result of theorem 14 could be obtained by direct combination of expressions (22) and (23).

6 The complexity of PIS5

6.1 In general graphs

We recall that strategy U5 considers the restriction $C[I]$ of an a priori vertex cover in the present subgraph $G[I]$ of G, it removes the isolated vertices (if any) from $C[I]$ and it finally takes the complement, with respect to I, of the resulting set.

Theorem 15. *Given a graph $G = (V, E)$, an a priori independent set S and the modification strategy U5, then*

$$E\left(G_S^{\mathtt{U5}}\right) = \sum_{v_i \in S} p_i + \sum_{v_i \in (V \setminus S)} p_i \prod_{v_j \in \Gamma(v_i)} (1 - p_j). \tag{26}$$

$E(G_S^{U5})$ is computable in polynomial time.

Proof. Lines (1) to (4) of modification strategy U5 in section 2 constitute a modification strategy, denoted by U in what follows, for probabilistic vertex cover problem. For an a priori vertex cover C, the functional associated with U is (see [14] for a detailed computation):

$$E_C^{\mathtt{U}} = \sum_{v_i \in C} p_i - \sum_{v_i \in C} p_i \prod_{v_j \in \Gamma(v_i)} (1 - p_j). \tag{27}$$

Using expression (27), we have the following for $E(G_S^{\mathtt{U5}})$:

$$\begin{aligned}
E\left(G_S^{\mathtt{U5}}\right) &= \sum_{I \subseteq V} \Pr[I] F\left(G[I]_S^{\mathtt{U5}}\right) = \sum_{I \subseteq V} \Pr[I] \left(|I| - F\left(G[I]_{V \setminus S}^{\mathtt{U}}\right)\right) \\
&= \sum_{I \subseteq V} \Pr[I]|I| - \sum_{I \subseteq V} \Pr[I] F\left(G[I]_{V \setminus S}^{\mathtt{U}}\right) = \sum_{I \subseteq V} \Pr[I] \sum_{v_i \in V} 1_{\{v_i \in I\}} - E(G_{V \setminus S}^{\mathtt{U}}) \\
&= \sum_{v_i \in V} \sum_{I \subseteq V} \Pr[I] 1_{\{v_i \in I\}} - E\left(G_{V \setminus S}^{\mathtt{U3}}\right) = \sum_{v_i \in V} p_i - \sum_{v_i \in (V \setminus S)} p_i + \sum_{v_i \in (V \setminus S)} p_i \prod_{v_j \in \Gamma(v_i)} (1 - p_j) \\
&= \sum_{v_i \in S} p_i + \sum_{v_i \in (V \setminus S)} p_i \prod_{v_j \in \Gamma(v_i)} (1 - p_j). \blacksquare
\end{aligned}$$

Expression (26) can be rewritten as

$$\begin{aligned}
E\left(G_S^{\mathtt{U5}}\right) &= \sum_{v_i \in V} p_i - \sum_{v_i \in (V \setminus S)} p_i + \sum_{v_i \in (V \setminus S)} p_i \prod_{v_j \in \Gamma(v_i)} (1 - p_j) \\
&= \sum_{v_i \in V} p_i - \sum_{v_i \in (V \setminus S)} p_i \left(1 - \prod_{v_j \in \Gamma(v_i)} (1 - p_j)\right).
\end{aligned}$$

Since, given a graph G, the quantity $\sum_{v_i \in V} p_i$ is constant, maximization of $E(G_S^{\mathtt{U5}})$ becomes equivalent to the minimization of $\sum_{v_i \in (V \setminus S)} p_i(1 - \prod_{v_j \in \Gamma(v_i)}(1 - p_j))$. But S being a maximal independent set, $V \setminus S$ is a minimal vertex covering of G and in order to find the vertex covering C minimizing the quantity $\sum_{v_i \in C} p_i(1 - \prod_{v_j \in \Gamma(v_i)}(1 - p_j))$, one has simply to consider each vertex $v_i \in V$ as weighted by the weight $w_i = p_i(1 - \prod_{v_j \in \Gamma(v_i)}(1 - p_j))$ and to search for a minimum-weight vertex cover. Consequently the following theorem characterizes the a priori solution maximizing $E(G_S^{\mathtt{U5}})$.

Theorem 16. *The a priori solution $\hat{S}$ maximizing $E(G_S^{\mathtt{U5}})$ is the complement, with respect to V, of a minimum-weight vertex cover of G where every vertex v_i is weighted by a weight $p_i(1 - \prod_{v_j \in \Gamma(v_i)}(1 - p_j))$. Consequently, PIS5 is **NP**-hard.*

In other words, theorem 16 establishes that, as in the case of PIS1, PIS5 is equivalent to a WIS. Since weights do not intervene in the ratio obtained in [6], corollary 1 holds also for PIS5.

6.2 In bipartite graphs

Since maximum-weight independent is polynomial in bipartite graphs ([5]), so does minimum-weight vertex covering. So the following theorem immediately holds.

Theorem 17. *The a priori solution $\hat{S}$ maximizing $E(B_S^{U5})$ is the complement, with respect to $V_1 \cup V_2$, of a minimum-weight vertex cover of B where every vertex v_i is weighted by a weight $p_i(1 - \prod_{v_j \in \Gamma(v_i)}(1 - p_j))$. Consequently, PIS5 is polynomial for bipartite graphs.*

7 Conclusions

We have drawn a formal framework for the study of PCOPs and studied five variants of the probabilistic maximum independent set, defined with respect to natural modification strategies used to adapt an a priori solution to the "present sub-instance".

Table 1: Complexities of computing functionals and characterizations and complexities of computing a priori solutions for several variants of probabilistic independent set.

	PIS1	PIS2	PIS4	PIS5
General graphs				
$T(E)$	$O(n)$	$O\left(2^{n-\alpha(G)}\right)$	$O\left(n^2\right)$	$O\left(n^2\right)$
$\hat{S}$	$\mathrm{WIS}(G_p^w)$	hard	?	$\mathrm{WIS}(G_{p'}^w)$
Complexity	**NP**-hard	?	?	**NP**-hard
Bipartite graphs				
$T(E)$	$O(n)$	$?^{(\alpha)}$	$O\left(n^2\right)$	$O\left(n^2\right)$
$\hat{S}$	$\mathrm{WIS}(B_p^w)$	?	?	$\mathrm{WIS}(B_{p'}^w)$
Complexity	**P**	?	?	**P**

$^{(\alpha)}$ Polynomial if $\mathrm{argmax}\{|V_1|, |V_2|\}$ is used as a priori solution.

Dealing with the quality of the solutions obtained we have the following relation for the same a priori solution S (let us note that, until now, we have not been able to compare $E(G_S^{U2})$ with $E(G_S^{U3})$):

$$E(G_S^{U1}) \leqslant E(G_S^{U5}) \leqslant E(G_S^{U4}) \leqslant E(G_S^{U2}) \tag{28}$$

First inequality in expression (28) is obvious and follows from theorem 2 and expression (26) of theorem 15. In order to prove the second inequality, remark that (expression (15))

$$p_k \times \left(1 - \prod_{v_l \in \Gamma_S(v_k)} (1 - p_l)\right) \geq 0$$

and consequently,

$$\begin{aligned} E\left(G_S^{U4}\right) &\geqslant \sum_{v_i \in S} p_i + \sum_{v_i \in (V \setminus S)} p_i \times \prod_{v_j \in \Gamma_S(v_i)} (1 - p_j) \times \prod_{v_k \in \Gamma_{V \setminus S}(v_i)} (1 - p_k) \\ &\geqslant \sum_{v_i \in S} p_i + \sum_{v_i \in (V \setminus S)} p_i \times \prod_{v_j \in \left(\Gamma_S(v_i) \cup \Gamma_{V \setminus S}(v_i)\right)} (1 - p_j) \\ &= \sum_{v_i \in S} p_i + \sum_{v_i \in (V \setminus S)} p_i \times \prod_{v_j \in \Gamma(v_i)} (1 - p_j) \;\geqslant\; E\left(G_S^{U5}\right). \end{aligned}$$

Table 2: Approximating a priori solutions in general and bipartite graphs.

	S	Approximation ratio
		General graphs
PIS1	The one computed in [6]	$\mathbf{r}$
PIS2	The one computed in [6]	$\max\left\{\left(\frac{p_{\min}}{1+p_{\max}}\right)\mathbf{r}, \frac{1}{\Delta_G+1}\right\}$ [(a)]
	$\text{argmax}\{\sum_{v_i\in S} p_i\}$	$\max\left\{\frac{p_{\min}}{1+p_{\max}}, \frac{1}{\Delta_G+1}\right\}$
PIS4	$\text{argmax}\{\sum_{v_i\in S} p_i\}$	$\frac{1}{\Delta_G+1}$ [(b)]
	The output of the greedy algorithm	$\frac{1}{\Delta_G+1}$
PIS5	The one computed in [6]	$\mathbf{r}$
		Bipartite graphs
PIS2	$\text{argmax}\{\sum_{v_i\in S} p_i\}$	$\frac{1}{2}$ [(c)]
PIS4	$\text{argmax}\{\sum_{v_i\in S} p_i\}$	$\frac{1}{2}$ [(d)]

[(a)] $\min\{O(\log n/(3(\Delta_G+1)\log\log n)), O(n^{-4/5})\}$ if vertex-probabilities independent of n.

[(b)] $\alpha^*(G)/n$ whenever $S = S^*$ and vertex-probabilities are identical.

[(c)] The same ratio is achieved considering $\text{argmax}\{\sum_{v_i\in V_1} p_i, \sum_{v_i\in V_2} p_i\}$ as a priori solution.

[(d)] The same ratio is achieved considering $\text{argmax}\{\sum_{v_i\in V_1} p_i, \sum_{v_i\in V_2} p_i\}$ as a priori solution.

Last inequality is due to the fact that for every subgraph $G[I]$, the cardinality of the solution computed applying U2 will be greater than the one of the solution computed applying U4 since GREEDY (called by U2) will always add in the solution, at least the isolated vertices of $G[\tilde{I}]$.

Table 1 summarizes the main results of this paper about the complexities of computing the functionals and the ones of computing the a priori solutions maximizing them. In this table we denote by G_p^w, a graph G whose vertices are weighted by their corresponding probabilities, by $G_{p'}^w$ a graph whose vertex v_i is weighted by the quantity $p_i(1-\prod_{v_j\in\Gamma(v_i)}(1-p_j))$, $1 \leqslant i \leqslant n$, by $T(E)$ the time needed for the computation of the functional E and by $\text{WIS}(G_p^w)$ (resp., $\text{WIS}(G_{p'}^w)$), the fact that the a priori solution maximizing the functional is a maximum-weight independent set in G_p^w (resp., $G_{p'}^w$). Finally, G and B denote general and bipartite graphs, respectively.

In table 2, where we denote by $\mathbf{r}$ the quantity $\min\{O(\log n/(3(\Delta_G+1)\log\log n)), O(n^{-4/5})\}$, a summary of the main approximation results is presented. Let us note that the approximation ratio in the line for PIS5 in general graphs of table 2 is directly obtained with arguments exactly analogous to the ones of corollary 1 in section 3.2, considering $p_i(1-\prod_{v_j\in\Gamma(v_i)}(1-p_j))$ as vertex-weight for v_i, $i = 1,\ldots,n$.

References

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

[2] D. J. Bertsimas. *Probabilistic combinatorial optimization problems.* PhD thesis, Operations Research Center, MIT, Cambridge Mass., USA, 1988.

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

[4] B. Bollobás. *Random graphs.* Academic Press, London, 1985.

[5] J.-M. Bourjolly, P. L. Hammer, and B. Simeone. Node-weighted graphs having the König-Egervary property. *Math. Programming Stud.*, 22:44–63, 1984.

[6] M. Demange and V. T. Paschos. Maximum-weight independent set is as "well-approximated" as the unweighted one. Technical Report 163, LAMSADE, Université Paris-Dauphine, 1999.

[7] V. Gabrel, A. Moulet, C. Murat, and V. T. Paschos. A new single model and derived algorithms for the satellite shot planning problem using graph theory concepts. *Ann. Oper. Res.*, 69:115–134, 1997.

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

[9] F. Harary. *Graph theory.* Addison-Wesley, Reading, MA, 1969.

[10] J. Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. To appear in Acta Mathematica.

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

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

[13] C. Murat. *Probabilistic combinatorial optimization graph-problems.* PhD thesis, LAMSADE, Université Paris-Dauphine, 1997. In french.

[14] C. Murat and V. T. Paschos. The probabilistic minimum vertex covering problem. Manuscript, LAMSADE, Université Paris-Dauphine, Paris, France.

[15] V. T. Paschos. A survey about how optimal solutions to some covering and packing problems can be approximated. *ACM Comput. Surveys*, 29(2):171–209, 1997.

[16] P. Turán. On an extremal problem in graph theory (in Hungarian). *Mat. Fiz. Lapok*, 48:436–452, 1941.

A Proof of proposition 1

A.1 Determining $A_{\alpha^*(G)}$

Let us first prove a small preliminary lemma.

Lemma 1. *If $S \subseteq I$, then $F(I_S^{\mathtt{U2}}) = \alpha(G)$.*

Proof. In fact, by definition of U2, $F(I_S^{\mathtt{U2}}) = |S[I]| + |S[\tilde{I}]|$. But if $S \subseteq I$, then $S[I] = S$ and this implies $S[I] \cup \Gamma(S[I]) = S \cup \Gamma(S)$. Moreover, the maximality of S implies that $S \cup \Gamma(S) = V$; and consequently, $S[\tilde{I}] = \emptyset$. So, $F(I_S^{\mathtt{U2}}) = |S[I]| = \alpha(G)$. ∎

For $A_{\alpha^*(G)}$ we have $|S^* \cap I| = \alpha^*(G)$ and, by lemma 1, $F(I_{S^*}^{\mathtt{U2}}) = \alpha^*(G)$; therefore,

$$A_{\alpha^*(G)} = \alpha^*(G)\left(\sum_{\substack{I \subseteq V \\ |S^* \cap I| = \alpha^*(G)}} \Pr[I]\right) = \alpha^*(G) p^{\alpha^*(G)} \sum_{i=0}^{n-\alpha^*(G)} \binom{n-\alpha^*(G)}{i} p^i (1-p)^{n-\alpha^*(G)-i} = \alpha^*(G) p^{\alpha^*(G)}$$

where, in the above expression, the term $p^{\alpha^*(G)}$ represents the fact that the $\alpha^*(G)$ vertices of S^* are all present in I ($S^* \cap I = S^*$) and the term $\sum_{i=0}^{n-\alpha^*(G)} C_i^{n-\alpha^*(G)} p^i (1-p)^{n-\alpha^*(G)-i}$ stands for all possible choices for the rest of the elements of V (as we have already seen, this term equals 1).

A.2 Determining $A_{\alpha^*(G)-1}$

Consider an element v of S^* and note that, since S^* is a maximum independent set, $\Gamma'(v)$ is either empty, or a clique on at least one vertex (if not, S^* could be augmented to $(S^* \setminus \{v\}) \cup \Gamma'(v)$). Consequently, by considering that v is the element of S^* which is absent from I, we have:

- if $I \cap \Gamma'(v) = \emptyset$, then the independent set $S^* \setminus \{v\}$ remains a maximal one for $G[I]$, so $F(I_{S^*}^{\mathtt{U2}}) = \alpha^*(G) - 1$;
- if $I \cap \Gamma'(v) \neq \emptyset$, then the independent set $S^* \setminus \{v\}$ (included in $G[I]$) can be augmented by exactly one element of $\Gamma'(v)$, so $F(I_{S^*}^{\mathtt{U2}}) = \alpha^*(G)$.

We now study the two following cases with respect to $\Gamma'(v)$, namely $\Gamma'(v) = \emptyset$ and $\Gamma'(v) \neq \emptyset$.

$\Gamma'(v) = \emptyset$.
$F(I_{S^*}^{\mathtt{U2}}) = \alpha^*(G) - 1$ and

$$\sum_{\substack{I \subseteq V \\ (S^* \setminus \{v\}) \subseteq I}} \Pr[I] F\left(I_{S^*}^{\mathtt{U2}}\right) = (\alpha^*(G) - 1) \sum_{\substack{I \subseteq V \\ (S^* \setminus \{v\}) \subseteq I}} \Pr[I]$$

$$\sum_{\substack{I \subseteq V \\ (S^* \setminus \{v\}) \subseteq I}} \Pr[I] = p^{\alpha^*(G)-1}(1-p) \sum_{i=0}^{n-\alpha^*(G)} \binom{n-\alpha^*(G)}{i} p^i (1-p)^{n-\alpha^*(G)-i} = p^{\alpha^*(G)-1}(1-p)$$

Consequently,

$$\sum_{\substack{I \subseteq V \\ (S^* \setminus \{v\}) \subseteq I}} \Pr[I] F\left(I_{S^*}^{\mathtt{U2}}\right) = (\alpha^*(G) - 1)\, p^{\alpha^*(G)-1}(1-p).$$

$\Gamma'(v) \neq \emptyset$.
Here, we study the following two subcases, namely, $I \cap \Gamma'(v) = \emptyset$ and $I \cap \Gamma'(v) \neq \emptyset$.

Subcase $I \cap \Gamma'(v) = \emptyset$:
here $F(I^{\mathrm{U2}}_{S^*}) = \alpha^*(G) - 1$ and, moreover, no vertex of $\Gamma'(v)$ is contained in I; so, we get

$$\sum_{\substack{I \subseteq V \\ (S^* \setminus \{v\}) \subseteq I \\ I \cap \Gamma'(v) = \emptyset}} \Pr[I] = p^{\alpha^*(G)-1}(1-p)(1-p)^{|\Gamma'(v)|} = p^{\alpha^*(G)-1}(1-p)^{|\Gamma'(v)|+1}. \qquad \text{(A.1)}$$

Subcase $I \cap \Gamma'(v) \neq \emptyset$:
$F(I^{\mathrm{U2}}_{S^*}) = \alpha^*(G)$ and I contains at least a vertex of $\Gamma'(v)$; so, we get for this case:

$$\sum_{\substack{I \subseteq V \\ (S^* \setminus \{v\}) \subseteq I \\ I \cap \Gamma'(v) \neq \emptyset}} \Pr[I] = p^{\alpha^*(G)-1}(1-p)\left[1 - (1-p)^{|\Gamma'(v)|}\right]; \qquad \text{(A.2)}$$

Consequently, for case $\Gamma'(v) \neq \emptyset$ we have combining expressions (A.1) and (A.2) together with expressions for $F(I^{\mathrm{U2}}_{S^*})$ of the two subcases:

$$\sum_{\substack{I \subseteq V \\ (S^* \setminus \{v\}) \subseteq I}} \Pr[I] F\left(I^{\mathrm{U2}}_{S^*}\right) = p^{\alpha^*(G)-1}(1-p)\left[\alpha^*(G) - (1-p)^{|\Gamma'(v)|}\right].$$

This concludes the study of case $\Gamma'(v) \neq \emptyset$.

By summing the above expressions over all elements of S^*, we get:

$$\begin{aligned}
A_{\alpha^*(G)-1} &= \sum_{v \in S^*} \sum_{I \subseteq V} \Pr[I] F\left(I^{\mathrm{U2}}_{S^*}\right) 1_{\{(S^* \setminus \{v\}) \subseteq I\}} \\
&= \sum_{v \in S^*} \sum_{I \subseteq V} \Pr[I] F\left(I^{\mathrm{U2}}_{S^*}\right) 1_{\{(S^* \setminus \{v\}) \subseteq I\}} \left(1_{\{\Gamma'(v) = \emptyset\}} + 1_{\{\Gamma'(v) \neq \emptyset\}}\right) \\
&= \ell_1 \left(\alpha^*(G) - 1\right) p^{\alpha^*(G)-1}(1-p) + \sum_{\substack{v \in S^* \\ \Gamma'(v) \neq \emptyset}} p^{\alpha^*(G)-1}(1-p)\left(\alpha^*(G) - (1-p)^{|\Gamma'(v)|}\right) \\
&= \ell_1 \left(\alpha^*(G) - 1\right) p^{\alpha^*(G)-1}(1-p) + \left(\alpha^*(G) - \ell_1\right) p^{\alpha^*(G)-1}(1-p)\alpha^*(G) \\
&\quad - p^{\alpha^*(G)-1}(1-p) \sum_{\substack{v \in S^* \\ \Gamma'(v) \neq \emptyset}} (1-p)^{|\Gamma'(v)|} \\
&= p^{\alpha^*(G)-1}(1-p)\left(\ell_1 \alpha^*(G) - \ell_1 + \alpha^*(G)^2 - \ell_1 \alpha^*(G) - \sum_{\substack{v \in S^* \\ \Gamma'(v) \neq \emptyset}} (1-p)^{|\Gamma'(v)|}\right) \\
&= p^{\alpha^*(G)-1}(1-p)\left(-\ell_1 + \alpha^*(G)(\ell_1 + \ell_1') - \sum_{\substack{v \in S^* \\ \Gamma'(v) \neq \emptyset}} (1-p)^{|\Gamma'(v)|}\right).
\end{aligned}$$