Proceedings of the International Multiconference on
Computer Science and Information Technology pp. 893–900
ISBN 978-83-60810-27-9
ISSN 1896-7094

On the PROBABILISTIC MIN SPANNING TREE problem (Extended abstract)

Nicolas Boria, Cécile Murat, Vangelis Th. Paschos
LAMSADE, CNRS and Université Paris-Dauphine, France
Email: {boria,murat,paschos}@lamsade.dauphine.fr

***Abstract*—We study a probabilistic optimization model for MIN SPANNING TREE, where any vertex v_i of the input-graph $G(V, E)$ has some presence probability p_i in the final instance $G' \subset G$ that will effectively be optimized. Supposing that when this "real" instance G' becomes known, a decision maker might have no time to perform computation from scratch, we assume that a spanning tree T, called *anticipatory* or *a priori* spanning tree, has already been computed in G and, also, that a decision maker can run a quick algorithm, called *modification strategy*, that modifies the anticipatory tree T in order to fit G'. The goal is to compute an anticipatory spanning tree of G such that, its modification for any $G' \subseteq G$ is optimal for G'. This is what we call PROBABILISTIC MIN SPANNING TREE problem. In this paper we study complexity and approximation of PROBABILISTIC MIN SPANNING TREE in complete graphs as well as of two natural subproblems of it, namely, the PROBABILISTIC METRIC MIN SPANNING TREE and the PROBABILISTIC MIN SPANNING TREE 1,2 that deal with metric complete graphs and complete graphs with edge-weights either 1, or 2, respectively.**

I. INTRODUCTION

The basic problematic of probabilistic combinatorial optimization (in graphs) is the following. We are given a graph $G(V, E)$ on which we have to solve some optimization problem Π. But, for some reasons depending on the reality modeled by G, Π is only going to be solved for some subgraph G' of G (determined by the vertices that will finally be present) rather than for the whole of G. The measure of how likely it is that a vertex $v_i \in V$ will belong to G' (i.e., will be present for the final optimization) is expressed by a probability p_i associated with v_i. How we can proceed in order to solve Π under this kind of uncertainty?

A first very natural idea that comes to mind is that one waits until G' is specified (i.e., it is present and ready for optimization) and, at this time, one solves Π in G'. This is what is called *reoptimization*.

But what if there remains very little time for such a computation? In this case, another way to proceed is the following. One solves Π in the whole of G in order to get a feasible solution (denoted by S), called *a priori* or *anticipatory solution*, which will serve her/him as a kind of benchmark for the solution on the effectively present subgraph G'. One has also to be provided with an algorithm that modifies S in order to fit G'. This algorithm is called *modification strategy* (let us denote it by M). The objective now becomes to compute an anticipatory solution that, when modified by M, remains "good" for any subgraph of G (if this subgraph is the one where Π will be finally solved).

This amounts to computing a solution that optimizes a kind of expectation of the value of the modification of S over all the possible subgraphs of G, i.e., the sum of the products of the probability that G' is the finally present graph multiplied by the value of the modification of S in order to fit G' over any subgraph G' of G. This expectation, depending on both the instance of the deterministic problem Π, the vertex-probabilities, and the modification strategy adopted, will be called the *functional*. Obviously, the presence-probability of G' is the probability that all of its vertices are present and the other vertices outside G' are absent.

Seen in this way, the probabilistic version PROBABILISTIC Π of a (deterministic) combinatorial optimization problem Π becomes another equally deterministic problem Π', the solutions of which have the same feasibility constraints as those of Π but with a different objective function where vertex-probabilities intervene.

In this sense, probabilistic combinatorial optimization is very close to what in the last couple of years has been called "one stage optimisation under independent decision models", an area very popular in the stochastic optimization community.

What are the main mathematical problems dealing with probabilistic consideration of a problem Π in the sense discussed above? One can identify at least five interesting mathematical and computational problems dealing with probabilistic combinatorial optimization:

1) write the functional down in an analytical closed form;
2) if such an expression of the functional is possible, prove that its value is polynomially computable (this amounts to proving that the modified problem Π' belongs to **NP**);
3) determine the complexity of the computation of the optimal *a priori* solution, i.e., of the solution optimizing the functional (in other words, determine the computational complexity of Π');
4) if Π' is **NP**-hard, study polynomial approximation issues;
5) always, under the hypothesis of the **NP**-hardness of Π', determine its complexity in the special cases where Π is polynomial, and in the case of **NP**-hardness, study approximation issues.

Let us note that, although curious, point 2 in the above list is neither trivial nor senseless. Simply consider that the summation for the functional includes, in a graph of order n, 2^n terms (one for each subgraph of G). So, polynomiality of the computation of the functional is, in general, not immediate.

Several optimization frameworks have been introduced by the operations research community for handling data uncertainty, the most well developed being *Stochastic programming* (see [1], [2] for basics) and *Robust discrete optimization* (see, for example, [3]).

The framework of *Probabilistic combinatorial optimization* where our work lies at, was introduced by [4], [5]. In [6], [5], [7], [8], [9], [4], [10], [11], [12], restricted versions of routing and network-design probabilistic minimization problems (in complete graphs) have been studied under the robustness model dealt here (called *a priori optimization*). In [13], the analysis of the probabilistic minimum travelling salesman problem, originally performed in [5], [4], has been revisited and refined.

Several other combinatorial problems have been recently treated in the probabilistic combinatorial optimization framework, including minimum coloring ([14], [15]), maximum independent set and minimum vertex cover ([16], [17]), longest path ([18]), Steiner tree problems ([19], [20]). Note also that probabilistic minimum spanning tree has also studied by [8] but under a very different probabilistic model.

We apply in this paper the probabilistic combinatorial optimization setting just described in the minimum spanning tree problem.

Given an edge-weighted graph $G(V, E)$, with positive edge weights $d : E \to \mathbb{Q}^+$, the minimum spanning tree problem (MIN SPANNING TREE) consists of determining a minimum total edge-weight tree spanning V.

MIN SPANNING TREE is a celebrated problem, very frequently modeling several kinds of networks in transports, communications, energy, logistics, etc. MIN SPANNING TREE has been actively studied under several optimization models like on-line computation, dynamic optimization, etc. Its study always motivates numerous researchers in theoretical computer science and in operational research.

In what follows, we first design a modification strategy and derive an analytic expression of the expected value (called *functional* in what follows) of a spanning tree of G, under this modification strategy.

We next show that the problem of *a priori optimization*, i.e., the problem of determining an anticipatory solution minimizing the functional, is **NP**-hard in general complete graphs (Section II).

Subsequently, we study complexity of the PROBABILISTIC MIN SPANNING TREE problem when dealing with particular cases of vertex-probabilities values and/or edge weights and particular cases of anticipatory solutions (Section IV).

We next derive polynomial-time approximation results for metric graphs and for graphs where edge-weights are either 1 or 2 (Section V). Because of the limits to the paper's size, some of the results are given without their proofs.

II. THE MODIFICATION STRATEGY, THE FUNCTIONAL ASSOCIATED WITH, AND THE COMPLEXITY OF PROBABILISTIC MIN SPANNING TREE

Consider a complete[1] weighted graph $G(V, E)$ on n vertices, with edge weights given by a function $d : E \to \mathbb{Q}^+$. Set $V = \{v_1, v_2, \ldots, v_n\}$. Each vertex $v_i \in V$, is associated with a presence probability $p_i \in \mathbb{Q}^+$ measuring, as already mentioned, how likely is that v_i will be present in the instance where PROBABILISTIC MIN SPANNING TREE will really be solved. We assume that subgraph $G'(V', E')$ of G materializes as the outcome of n independent Bernoulli trials, one per vertex $v_i \in V$: $v_i \in V'$ with probability p_i. Then, $E' = \{(u, v) \in E : u \in V' \text{ and } v \in V'\}$. Let us note that it seems to be natural that, for a fixed modification strategy M, given an anticipatory spanning tree T, some basic properties of its structure must be preserved in any tree T' built when M adjusts T to $G'(V', E')$, for any $V' \subseteq V$. Such a basic property is, for instance, the relation "predecessor-successor" in T. In order that this relation is preserved in any T', we assume that there exists a vertex, denoted by v_1 with $p_1 = 1$.

To motivate this assumption, consider a pacifist version of an application in [5]. Let G be a graph of order n and let its vertices represent researchers of a research network. The weight of an edge linking researcher i to researcher j quantifies the "inefficiency" risk incurred when i and j have to accomplish a common task. We wish to determine an organizational structure of this network where all the researchers implied accomplish some tasks and where the total "inefficiency" risk is minimized.

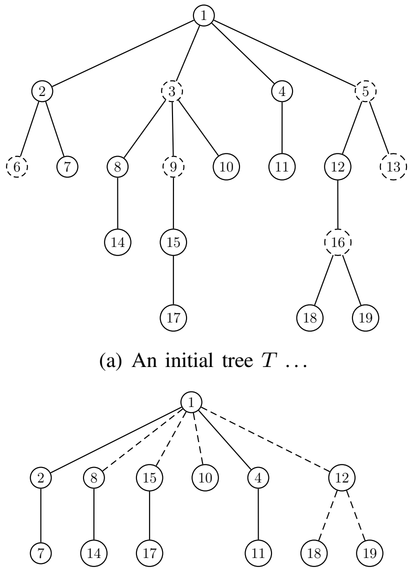
This can be modeled as a minimum spanning tree T^* of G. Let us now suppose that the project at hand involves several tasks where only a subset of researchers are implied, the project manager (the omnipresent vertex v_1) being involved to all of them. The probability associated with a researcher $v_i \neq v_1$, is the probability that v_i participates to an arbitrary task undertaken by the research network. Each such task will be represented by some subgraph G' of G and its "inefficiency" risk is the cost of a minimum spanning tree of G'. The modification strategy that is to be adopted for such a model must allow us to keep the same structure from a task to another one in order that the "inefficiency" risk remains as low as possible. Other applications from distributed systems also justify similar assumptions.

Formally, let $G(V, E)$ be a complete graph with $|V| = n$ and $(p_i)_{i=1,\ldots,n}$ a vertex-probability system with $p_1 = 1$ (in other words, vertex v_1 is assumed to be always present). Consider a tree T spanning V and number vertices in T in a left-to-right breadth-first-search (bfs) way. Consider a subgraph $G'(V', E') = G[V']$ of G induced by a set $V' \subseteq V$. The modification strategy (adjusting T to a spanning tree T' of G' and denoted by LEV in what follows) that will be analyzed in the sequel works as follows:

[1] The assumption that the input-graph is complete is made in order to ensure connectivity of the tree T' for any subgraph $G' \subseteq G$; in any case if the real-world problem at hand implies non-complete graphs, one can complete them by "appropriately heavy" edges.

1) remove the vertices of $V \setminus V'$ and the edges of E incident to these vertices; let $F(V') = \{T_1, T_2, \ldots, T_k\}$ be the so-obtained forest and assume that $v_1 \in V(T_1)$ and that, for $i, j = 2 \ldots, k$, $i < j$ if the index of the root of T_i is smaller than that of the root of T_j;
2) for $i = 2, \ldots, k$ add as father of the root of T_i its largest-index ancestor that is still present in $V(T)$.

Note that, given two vertices v_j and v_l with $j < l$ (in the bfs numbering of T), if v_j is not an ancestor of v_l in T, then edge (v_j, v_l) will never belong to any T' modification of T for any $V' \subseteq V$.



(a) An initial tree T …

(b) …and its modification by LEV

Fig. 1. An anticipatory spanning tree T and its adjustment.

Figure 1(b) gives an example of how strategy LEV works starting from an initial tree T shown in Figure 1(a) and assuming that vertices 3, 5, 6, 9, 13 and 16 are absent from V'.

The functional $E(G, T)$ associated with LEV is defined by:

$$E(G,T) = \sum_{V' \subseteq V} \Pr[V'] \, m(G', T')$$

where $\Pr[V'] = \prod_{v_i \in V'} p_i \prod_{v_i \in V \setminus V'} (1 - p_i)$ is the distribution describing probability of occurrence of a specific subset $V' \subseteq V$, i.e., of the graph $G[V']$ and $m(G', T')$ is the value of the tree T' spanning V' produced by application of LEV on the anticipatory tree T.

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

Our goal is to study the following problem: *find an algorithm for taking a priori decisions, i.e., that determines a spanning tree T^*, that optimizes $E(G, T)$*"; this is PROBABILISTIC MIN SPANNING TREE.

In what follows, we show that this problem is **NP**-hard in general complete graphs with $p_1 = 1$. We then study approximation of this problem in metric graphs as well as in a particular subclass of them where edge-weights are either 1 or 2. The approximation ratio is defined as $E(G, T)/E(G, T^*)$.

The following result holds for the functional $E(G, T)$ associated with an anticipatory spanning tree T and the modification strategy LEV.

Proposition 1: Consider a complete graph $G(V, E)$, provided with a vertex-probability system $(p_i)_{i=1,\ldots,n}$ with $p_1 = 1$, any edge (v_i, v_j) of which has weight d_{ij} and a spanning tree T of G. Then, the expectation associated with LEV can be expressed by:

$$E(G,T) = \sum_{(v_i,v_j) \in T} p_i p_j d_{ij} + \sum_{v_i \in V} \sum_{\substack{v_j \in D(v_i) \\ (v_i,v_j) \notin T}} p_i p_j \times \prod_{v_k \in \mu[v_i,v_j]} (1 - p_k) \, d_{ij} \quad (1)$$

where $D(v_i)$ denotes the set of successors of v_i in T and $\mu[v_i, v_j]$ denotes the set of vertices in the (unique) path of T from v_i to v_j not including neither of them. Expression (1) can be computed in polynomial time. Consequently, PROBABILISTIC MIN SPANNING TREE $\in$ **NPO**, the class of the optimization problems the decision versions of which are in **NP**.

Proof: Following LEV, if $(v_i, v_j) \in T$, then this edge will be in T' iff $v_i, v_j \in V'$; on the other hand, if $(v_i, v_j) \notin T$, then edge (v_i, v_j) will be added in T' iff $v_i, v_j \in V'$ and $v_j \in D(v_i)$ and every vertex in $\mu[v_i, v_j]$ is not in V'. From these observations we derive:

$$\begin{aligned} E(G,T) &= \sum_{V' \subseteq V} \Pr[V'] \, m(G', T') \\ &= \sum_{V' \subseteq V} \Pr[V'] \sum_{(v_i,v_j) \in T'} d_{ij} \\ &= \sum_{V' \subseteq V} \Pr[V'] \sum_{(v_i,v_j) \in T' \cap T} d_{ij} + \sum_{V' \subseteq V} \Pr[V'] \sum_{(v_i,v_j) \in T' \cap (E \setminus T)} d_{ij} \\ &= \sum_{(v_i,v_j) \in T} \sum_{V' \subseteq V} \Pr[V'] \, \mathbf{1}_{\{(v_i,v_j) \in T'\}} d_{ij} + \sum_{(v_i,v_j) \in (E \setminus T)} \sum_{V' \subseteq V} \Pr[V'] \, \mathbf{1}_{\{(v_i,v_j) \in T'\}} d_{ij} \\ &= \sum_{(v_i,v_j) \in T} p_i p_j d_{ij} + \sum_{v_i \in V} \sum_{\substack{v_j \in D(v_i) \\ (v_i,v_j) \notin T}} p_i p_j \prod_{v_k \in \mu[v_i,v_j]} (1 - p_k) \, d_{ij} \end{aligned}$$

Clearly, (1) can be computed in polynomial time, since the ranges of the indices implied are polynomial. ■

As already mentioned, PROBABILISTIC MIN SPANNING TREE consists of determining an anticipatory spanning tree T^* of G minimizing $E(G,T)$.

Unfortunately, Proposition 1 does not derive a compact combinatorial characterization for the optimal anticipatory solution of PROBABILISTIC MIN SPANNING TREE.

In particular, the form of the functional does not imply solution, for instance, of some well-defined weighted version of the (deterministic) MIN SPANNING TREE. This is due to the second term of the expression for $E(G,T)$ in (1). There, the "costs" assigned to the edges depend on the structure of the anticipatory solution chosen and of the present subgraph of G.

The decision version of PROBABILISTIC MIN SPANNING TREE, denoted by PROBABILISTIC MIN SPANNING TREE(K) can be stated as follows: "given an edge-weighted complete graph $G(V,E)$, provided with a vertex-probability system $(p_i)_{i=1,\dots,n}$ with $p_1 = 1$ and a constant K, does there exist a tree T such that $E(G,T) \leqslant K$?", where $E(G,T)$ is given by (1).

Dealing with PROBABILISTIC MIN SPANNING TREE(K), by a technical reduction from 3 EXACT COVER, the following proposition holds.

Proposition 2: PROBABILISTIC MIN SPANNING TREE(K) is **NP**-complete.

Sketch of proof: PROBABILISTIC MIN SPANNING TREE $\in$ **NP**. In order to show completeness, we reduce 3 EXACT COVER to PROBABILISTIC MIN SPANNING TREE. 3 EXACT COVER that is defined as follows: "given a ground set X of size $3q$ and a collection $\mathcal{E}$ of $3q$ subsets of X each of size 3, does there exist a subcollection $\mathcal{E}' = \{S_1,\dots,S_q\}$ of $\mathcal{E}$ such that $\bigcup_{i=1}^{q} S_i = X$?" (obviously, $\mathcal{E}'$ is a partition of X). 3 EXACT COVER is **NP**-complete ([21]).

Consider an arbitrary instance $I(X,\mathcal{E})$ of 3 EXACT COVER; we construct the following instance for PROBABILISTIC MIN SPANNING TREE:

- the vertex-set V is a set of $6q+2$ vertices built by associating a vertex x_i with an element $x_i \in X$, a vertex y_j with a set $S_j \in \mathcal{E}$ and by adding a vertex r (playing the role of the omnipresent root) and a vertex s (representing the solution); for some positive fixed constant $p < 1/2$, vertices x_i are provided with probability p, vertices y_j with probability $1-p$ and vertices r and s with probability 1;
- edge-weights are defined as follows:
  - for every $S_j = \{x_{i_1}, x_{i_2}, x_{i_3}\}$, $j = 1,\dots,3q$, $d_{i_1 j} = d_{i_2 j} = d_{i_3 j} = 1$;
  - edges linking s to vertices y_j have weight $M > 0$ and those linking s to vertices of x_i have weight $M/p + 2$;
  - edges linking r to vertices y_j as well as edge (r,s) have all weight 0, while edges linking r to vertices x_i have weight $M/p^2 + 1$;
  - all the other edges have arbitrarily large weight $B \gg M/p^2 + 1$;
- $K = q(M(1+2p) + 3p(p+1))$.

It is easy to see that this reduction is polynomial. It is illustrated in Figure 2 where, for readability, some edges, in particular those of weight B are omitted.

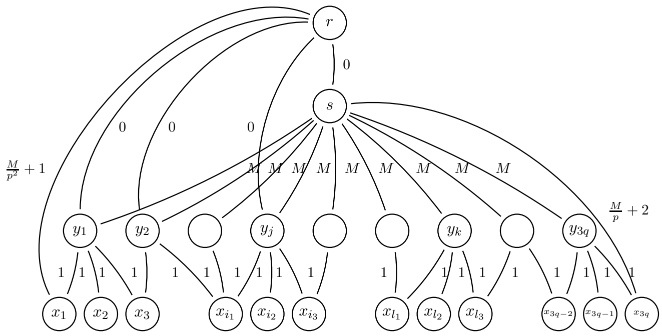


Fig. 2. An example for the reduction from 3 EXACT COVER to PROBABILISTIC MIN SPANNING TREE.

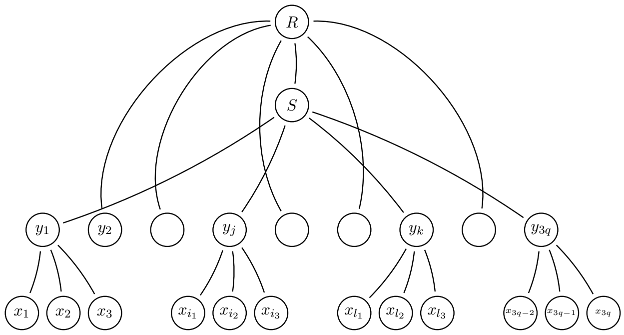


Fig. 3. The shape of T^*.

One can prove that if 3 EXACT COVER admits a solution $\mathcal{E}^*$, then G has a minimum spanning tree T^* the shape of which is as in Figure 3 and whose value is:

$$\begin{aligned} E(G,T^*) &= \\ &\quad 0 + 2q \times 0 + q[(1-p)M + 3p(1-p) \times 1 + \\ &\quad 3p(1-(1-p)) \times (M/p + 2)] \\ &= q[M - pM + 3p - 3p^2 + 3pM + 6p^2] \\ &= q[M(1+2p) + 3p(p+1)] \;=\; K \end{aligned}$$

This can be done by inspection of all the possible solutions for MIN SPANNING TREE in G. ■

III. REOPTIMIZATION AND MIN SPANNING TREE

As mentioned in Section I, a complementary framework to the one of the a priori optimization, is the *reoptimization* consisting of solving ex nihilo and optimally the portion of the instance presented for optimization. Reoptimization is introduced in [4]. Let $\mathrm{opt}(G')$ refer to the weight of the optimum spanning tree on G' for every subgraph $G'(V',E')$ of G. The expected minimum weight over the distribution of subgraphs of G, i.e., the functional of reoptimization is defined by $E^*(G) = \sum_{V' \subseteq V} \Pr[V']\,\mathrm{opt}(G')$.

Obviously, denoting by T^* the optimal anticipatory solution of PROBABILISTIC MIN SPANNING TREE: $E^*(G) \leqslant E(G,T^*)$. Denote also by $\mathrm{opt}(G)$ the value of an optimal solution T^* for (deterministic) MIN SPANNING TREE. By elementary but technical combinatorial arguments, the following result holds.

Proposition 3: Consider a complete edge-weighted graph G defined on a set V of n vertices $V = \{v_1, \ldots, v_n\}$ associated with a system of vertex probabilities $p_1 = 1$, $p_i = p$, $i = 2, \ldots, n$. Then, $E^*(G) \geqslant p\,\mathrm{opt}(G)$.

Sketch of proof: Since vertex v_1 (assumed to be the root of every tree solution of MIN SPANNING TREE in every subgraph of G) is always present, setting $G' = G[V']$, $E^*(G)$ can be written as:

$$E^*(G) = \sum_{k=2}^{n} p^{k-1}(1-p)^{n-k} \times \sum_{\substack{V' \subseteq V \\ |V'|=k}} \mathrm{opt}(G')$$

Set $D_k = \sum_{V' \subseteq V, |V'|=k} \mathrm{opt}(G')$. Then:

$$E^*(G) = \sum_{k=2}^{n} p^{k-1}(1-p)^{n-k} D_k$$

By somewhat technical combinatorial arguments it can be proved that

$$D_k \geqslant \binom{n-2}{k-2} \mathrm{opt}(G)$$

and putting it together with the expression for $E^*(G)$ derives the claim. ∎

Combining inequality $E^*(G) \leqslant E(G,T^*)$ and Proposition 3, the following holds: $E(G,T^*) \geqslant p\,\mathrm{opt}(G)$. Equality is attained for $p = 1$.

IV. Particular cases

We study in this section some particular but natural cases carrying over assumptions either on the values of vertex-probabilities and/or edge-weights, or on the form of the anticipatory solution.

Revisit functional's expression (1). For a vertex v_i, denote by $f(v_i)$ its father in T, by $p_{f(i)}$ the presence probability of $f(v_i)$ and by $A(v_i)$ the set of its ancestors in T. Then, (1) can be rewritten as:

$$\begin{aligned} E(G,T) = & \\ & \sum_{v_i \in V\setminus\{v_1\}} p_i p_{f(i)} d_{if(v_i)} + \\ & \sum_{v_i \in V\setminus\{v_1\}} \sum_{v_j \in A(v_i)\setminus\{f(v_i)\}} p_i p_j \prod_{v_k \in \mu[v_i,v_j]} (1-p_k)\, d_{ij} \\ = & \sum_{v_i \in V\setminus\{v_1\}} C_i \end{aligned} \quad (2)$$

where:

$$\begin{aligned} C_i = & \; p_i p_{f(i)} d_{if(v_i)} + \\ & \sum_{v_j \in A(v_i)\setminus\{f(v_i)\}} p_i p_j \prod_{v_k \in \mu[v_i,v_j]} (1-p_k) d_{ij} \end{aligned}$$

and can be seen as the contribution of vertex v_i in $E(G,T)$.

Based upon (2) and the expression for C_i, the following result holds.

Proposition 4: If edge-weights are all identical, then:

$$E(G,T) = d \times \sum_{j=2}^{n} p_j$$

In this case all the anticipatory solutions have the same value.

Let us give an illustration of Proposition 4. It can easily be shown that, if $d_{ij} = d$, $(v_i, v_j) \in E$, then $C_i = d \times p_i$, $v_i \in V$. In order to give some intuition about it let us consider the anticipatory tree of Figure 4, assume that edge weights in the input-graph are identical and equal to d and take, say, vertex 7. The contribution of it in (1) is:

$$\begin{aligned} C_7 = & \\ & d \times p_7 \left[p_5 + p_4(1-p_5) + p_2(1-p_5)(1-p_4) + \right. \\ & \left. (1-p_5)(1-p_4)(1-p_2)\right] \\ = & \; d \times p_7 \end{aligned} \quad (3)$$

The same holds for the contribution of any other vertex in the tree.

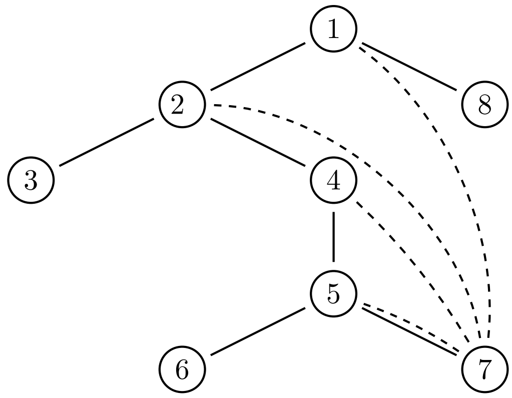


Fig. 4. About a reinterpretation of functional's expression (1).

Indeed, consider some vertex $v_i \in V$ and assume, for simplicity, that vertices in the path of T from v_1 to v_i are numbered from 1 to i. By writing down C_i and by some algebra as previously in (3) we derive $C_i = d \times p_i$. Hence, (2) becomes:

$$E(G,T) = d \times \sum_{j=2}^{n} p_j$$

Corollary 1: If $p_1 = 1$, $p_i = p$, $i = 2, \ldots, n$ and $d_{ij} = d$, $(v_i, v_j) \in E$, $i \neq j$, then, for any tree T spanning V, $E(G,T) = dp(n-1)$.

The above can be directly generalized for deriving a general upper bound for $E(G,T)$ and for every anticipatory solution T. Set $D = \max\{d_{ij} : (v_i, v_j) \in E\}$.

Corollary 2: If $D = \max\{d_{ij} : (v_i, v_j) \in E\}$ then, for any anticipatory solution T of PROBABILISTIC MIN SPANNING TREE, $E(G,T) \leqslant D \times \sum_{j=2}^{n} p_j$.

Let us now address the following question: “can an optimal solution for MIN SPANNING TREE remain an optimal solution for PROBABILISTIC MIN SPANNING TREE and if yes under which conditions?”. In what follows we deal with two particular structures of trees, the star and the path.

Consider the star rooted at (the omnipresent) vertex v_1. The following result holds.

Proposition 5: Let T be a star rooted at v_1. If T is an optimal solution for MIN SPANNING TREE then it is also an optimal anticipatory solution for PROBABILISTIC MIN SPANNING TREE.

Proof: Recall that by (2), $E(G,T) = \sum_{v_i \in V\setminus\{v_1\}} C_i$ where C_i is given by:

$$C_i = p_i p_{f(i)} d_{if(v_i)} + \sum_{v_j \in A(v_i)\setminus\{f(v_i)\}} p_i p_j \prod_{v_k \in \mu[v_i,v_j]} (1-p_k)\, d_{ij}$$

Observe now that, if the vertices of T are numbered in a dfs order (starting from the root) and if the set $A(v_i)$ of the ancestors of a vertex v_i in T is exactly the set $A(v_i) = \{v_1, v_2, \ldots, v_{i-1}\}$, then C_i can be written as:

$$C_i = \sum_{j=1}^{i-1} p_i p_j d_{ij} \prod_{l=j+1}^{i-1} (1-p_l) \qquad (4)$$

Since the star T is a minimum spanning tree, it holds that, for any vertex i, $d_{1i} \leqslant d_{ij}$, for every $j \neq 1, i$. Hence:

$$C_i \geqslant p_i \times d_{1i} \times \sum_{j=1}^{i-1} p_j \prod_{l=j+1}^{i-1} (1-p_l)$$

If we denote by C_i^T the contribution of vertex v_i in the functional $E(T)$ of the star T, then, for every i, $C_i^T = p_i \times d_{1i}$. So, in order to complete the proof of the proposition, we have to show that, for any $v_i \in V$, $C_i^T \leqslant C_i$, where C_i refers to every other spanning tree of G. For this, it suffices to prove that:

$$\sum_{j=1}^{i-1} p_j \prod_{l=j+1}^{i-1} (1-p_l) \geqslant 1 \qquad (5)$$

We show (5) by induction on i. For $i = 2$, the lefthand side of (5) is equal to $p_1 = 1$, so the inequality claimed is true. Suppose it true for $i = n$, i.e.:

$$\sum_{j=1}^{n-1} p_j \prod_{l=j+1}^{n-1} (1-p_l) \geqslant 1 \qquad (6)$$

Then, at range $n+1$ it holds:

$$\begin{aligned}
&\sum_{j=1}^{n} p_j \prod_{l=j+1}^{n} (1-p_l) = \\
&\quad \sum_{j=1}^{n-1} p_j \prod_{l=j+1}^{n} (1-p_l) + p_n \\
&= \sum_{j=1}^{n-1} p_j \prod_{l=j+1}^{n-1} (1-p_l) \times (1-p_n) + p_n \\
&= (1-p_n) \times \sum_{j=1}^{n-1} p_j \prod_{l=j+1}^{n-1} (1-p_l) + p_n \\
&\overset{(6)}{\geqslant} (1-p_n) + p_n = 1
\end{aligned}$$

as claimed. ■

Unfortunately, in the case where optimal solution for MIN SPANNING TREE is a path, optimality of such a solution for PROBABILISTIC MIN SPANNING TREE cannot be derived as previously in the case of stars.

Indeed, consider a complete graph G, the adjacency matrix of which is given in Table I and its vertex-probability system is $(1, p \ldots, p)$, with $p < (K-2)/(K-1)$ and $K \geqslant n$.

TABLE I
THE ADJACENCY MATRIX OF A GRAPH G WHERE OPTIMAL SOLUTIONS FOR MIN SPANNING TREE AND PROBABILISTIC MIN SPANNING TREE DO NOT COINCIDE.

	v_1	v_2	v_3	$\cdots$	v_{n-2}	v_{n-1}	v_n
v_1	0	1	2	$\cdots$	2	2	2
v_2	1	0	1	$\cdots$	2	2	2
v_3	2	1	0	$\cdots$	2	2	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$	$\vdots$
v_{n-2}	2	2	2	$\cdots$	2	1	K
v_{n-1}	2	2	2	$\cdots$	1	0	1
v_n	2	2	2	$\cdots$	K	1	0

Optimal MIN SPANNING TREE-solution in G is unique and is the path $P = (1, 2, \ldots, n-1)$ with value $n-1$. The functional $E(G,P)$ of path P is:

$$E(G,P) = (2n-3)p + (K-n)p^2 - (K-2)p^3$$

On the other hand, the unique optimal anticipatory solution for PROBABILISTIC MIN SPANNING TREE is the tree T^* of Figure 5 with functional's value:

$$E(G,T^*) = (2n-3)p + (2-n)p^2 + p^3 < E(G,P)$$

when $p < (K-2)/(K-1)$.

V. APPROXIMATION OF PROBABILISTIC METRIC MIN SPANNING TREE

In this section, we study PROBABILISTIC METRIC MIN SPANNING TREE problem, that is PROBABILISTIC MIN SPANNING TREE in metric complete graphs, i.e., in complete graphs whose edge-weights satisfy the triangular property that can

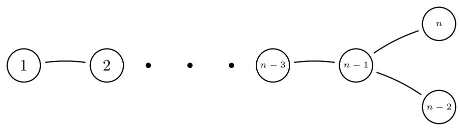


Fig. 5. The optimal solution of the graph of Table I.

be expressed as follows: if (v_i, v_j, v_k) is a K_3 in G, then $d_{ab} \leqslant d_{ac} + d_{bc}$, for any permutation (a, b, c) of $\{i, j, k\}$.

Note that, as it can be seen from the proof of Proposition 2, it does not apply in the case of metric graphs. Indeed, the complexity status of PROBABILISTIC METRIC MIN SPANNING TREE remains open, even if we feel that this variant is also **NP**-complete. In what follows we give approximation results for the general case of PROBABILISTIC METRIC MIN SPANNING TREE as well as for a natural subcase, namely PROBABILISTIC MIN SPANNING TREE 1,2, where edge weights are either 1, or 2.

We denote by $\hat{T}$ and T^* a tree computed by Kruskal's algorithm (i.e., a minimum spanning tree of G) and an optimal anticipatory solution for PROBABILISTIC METRIC MIN SPANNING TREE, respectively, and we assume that they are represented as sets of edges.

Observe also that, by the metric property the weight of any edge of G is smaller than the weight of any spanning tree of G and a fortiori than the weight $m(G, \hat{T}) = \text{opt}(G)$ of $\hat{T}$. Indeed, let (v_i, v_j) be any edge of G, T be some spanning tree of G and $P(v_i, v_j)$ be the unique path from v_i to v_j in T.

According to the metric hypothesis:

$$d_{ij} \leqslant w\left(P(v_i, v_j)\right) \leqslant m(G, T)$$

where $w(P(v_i, v_j))$ is the weight of the path $P(v_i, v_j)$. In what follows, given a set Q of (weighted) edges, we denote by $w(Q)$ its total weight.

We first handle the general metric case. For any $V_i \subseteq V$, we set $G_i(V_i, E_i) = G[V_i]$. Let T_i^* and $\hat{T}_i$ be the spanning trees on G_i resulting from the application of strategy LEV on T^* and $\hat{T}$, respectively. Set:

$$\begin{aligned} r\left(G_i\right) &= \frac{m\left(G_i, \hat{T}_i\right)}{m\left(G_i, T_i^*\right)} \\ E(r) &= \sum_{V_i \subseteq V} \Pr\left[V_i\right] r\left(G_i\right) \end{aligned}$$

Quantity $E(r)$ is indeed the average approximation ratio of a minimum spanning tree for PROBABILISTIC METRIC MIN SPANNING TREE. In the following proposition an upper bound is given for $E(r)$.

Proposition 6: $E(r) \leqslant (n+2)/4$.

Proof: Fix an induced subgraph G_i of G and let $\hat{T} \cap \hat{T}_i = S$. Edges of S are part of an optimal spanning tree on G and thus, they are also part of an optimal spanning tree of G_i. Indeed, revisiting the proof of optimality of Kruskal's Algorithm, one can see that a tree T is a minimum spanning tree on G, iff all the edges of T are of minimum weight in at least one cut of G. Applying this to PROBABILISTIC MIN SPANNING TREE, any edge e belonging to S is of minimum weight in at least one cut of G; thus, e is also of minimum weight in one cut in any subgraph G_i (provided that e appears in G_i) and, therefore, it belongs to a minimum spanning tree in all the subgraphs of G where it is present.

Discussion just above leads to:

$$w(S) \leqslant m\left(G_i, T_i^*\right) \tag{7}$$

The edge-set $\hat{T}_i \setminus S$ is the set of the edges used by LEV to reconnect the S. As observed in the beginning of Section V, the weight of each edge of $\hat{T}_i \setminus S$ is smaller than, or equal to, $m(G_i, T_i^*)$, so:

$$w\left(\hat{T}_i \setminus S\right) \leqslant \left|\hat{T}_i \setminus S\right| m\left(G_i, T_i^*\right) \tag{8}$$

Combining (7) and (8), we get:

$$r_i = \frac{m\left(G_i, \hat{T}_i\right)}{m\left(G_i, T_i^*\right)} \leqslant 1 + \left|\hat{T}_i \setminus S\right| \tag{9}$$

The quantity $|\hat{T}_i \setminus S|$ is, as mentioned above, the number of edges inserted by strategy LEV to reconnect S, but it also represents the number of vertices present in G_i, but whose fathers in $\hat{T}$ (assumed rooted at v_1) are absent from G_i. For each vertex of $\hat{T}$ except for those directly connected to the root v_1, the probability to be present in G_i but not its father is $p(1-p)$. Obviously, for the vertices directly connected to the root, this probability is 0. In order to count the number of edges in $\hat{T}_i \setminus S$, one can consider a set of $n-1-X$ Bernoulli trials (where X is the number of vertices directly connected to v_1 in $\hat{T}$), with a probability of success $p(1-p)$, each success adding an edge to $\hat{T}_i \setminus S$. In this way, $|\hat{T}_i \setminus S|$ is a random variable following a binomial law, so one can directly compute its expectation:

$$\begin{aligned} \left|\hat{T}_i \setminus S\right| &\sim B\left(n-1-X, p(1-p)\right) \\ E\left(\left|\hat{T}_i \setminus S\right|\right) &= (n-1-X)p(1-p) \end{aligned} \tag{10}$$

Summing (9) for each G_i, we derive:

$$E\left(\frac{m\left(G_i, \hat{T}_i\right)}{m\left(G_i, T_i^*\right)}\right) \leqslant 1 + E\left(\left|\hat{T}_i \setminus S\right|\right)$$

and combining it with (10), we can easily get:

$$\begin{aligned} E\left(\frac{m\left(G_i, \hat{T}_i\right)}{m\left(G_i, T_i^*\right)}\right) &\leqslant 1 + (n-1-X)p(1-p) \\ &\overset{X \geqslant 1}{\leqslant} \frac{n}{4} + \frac{1}{2} = \frac{n+2}{4} \end{aligned}$$

as claimed. ■

We now study the approximation of PROBABILISTIC METRIC MIN SPANNING TREE. The following result can be proved.

Proposition 7: The PROBABILISTIC METRIC MIN SPANNING TREE problem is approximable in polynomial time within ratio bounded above by:

$$\min\left\{1 + (n-2)(1-p), \frac{2}{p}\right\}$$

To conclude the paper, let us focus on a particular but natural and well-studied class of metric complete graphs where edge weights are either 1 or 2. It is easy to see that any such graph is metric. The following result, that will be subsequently improved, holds for PROBABILISTIC MIN SPANNING TREE 1,2.

Proposition 8: A minimum spanning tree of G is a $(2-p)$-approximation for PROBABILISTIC MIN SPANNING TREE 1,2.

In what follows we refine the result above. For this, we consider an execution of Kruskal's algorithm that starts by introducing in the tree all the edges of weight 1 incident to the vertex v_1. Let us denote by $\tilde{T}$ the spanning tree so constructed; notice that $\tilde{T}$ is a minimum spanning tree for G.

Proposition 9: $\tilde{T}$ approximates T^* within ratio:

$$\frac{1 + (2-p)(n-2)}{n - 1 + \frac{(1-p)^2 - (1-p)^n}{p}}$$

One can see that when p is fixed (i.e., independent on n), the approximation ratio achieved is strictly better than 2. On the other hand, when $p \sim 1/n$ then, since:

$$\lim_{n\to+\infty} \frac{(1-p)^n}{p} = \frac{n}{e}$$

the ratio claimed in Proposition 9 tends to 1.225, for large values of n.

If $p \sim 1/n^k$, $k > 1$, then for large values of n, this ratio tends to 1. Finally, if $p \sim 1/n^k$, $k < 1$, then (always for large values of n) the ratio is asymptotically equal to 2.

VI. CONCLUSION

In this paper we have treated the PROBABILISTIC MIN SPANNING TREE problem under the framework of probabilistic combinatorial optimization. We have proposed a fast modification strategy (LEV) for reconstructing a second-stage tree and shown that problem of optimizing the expectation of the second-stage cost by selecting an appropriate first-stage (anticipatory) solution is in **NPO** under the proposed modification strategy and we have shown that the general case of PROBABILISTIC MIN SPANNING TREE is **NP**-hard.

We have also given approximation results for the probabilistic problem associated with the LEV strategy. We also have studied particular cases of anticipatory solutions.

Finally we have given approximation results for PROBABILISTIC METRIC MIN SPANNING TREE and PROBABILISTIC MIN SPANNING TREE 1,2.

There are several open questions subject for further research. To our opinion, the major among them are the complexities of PROBABILISTIC METRIC MIN SPANNING TREE and PROBABILISTIC MIN SPANNING TREE 1,2 (we conjecture that they are both **NP**-hard) and the improvement of their approximation ratios.

REFERENCES

[1] G. W. Dantzig, "Linear programming under uncertainty," *Management Sci.*, vol. 1, pp. 197–206, 1951.
[2] A. Prekopa, *Stochastic programming*. The Netherlands: Kluwer Academic Publishers, 1995.
[3] P. Kouvelis and G. Yu, *Robust discrete optimization and its applications*. Boston: Kluwer Academic Publishers, 1997.
[4] P. Jaillet, "Probabilistic traveling salesman problem," Operations Research Center, MIT, Cambridge Mass., USA, Technical Report 185, 1985.
[5] D. J. Bertsimas, "Probabilistic combinatorial optimization problems," PhD Thesis, Operations Research Center, MIT, Cambridge Mass., USA, 1988.
[6] I. Averbakh, O. Berman, and D. Simchi-Levi, "Probabilistic a priori routing-location problems," *Naval Res. Logistics*, vol. 41, pp. 973–989, 1994.
[7] D. J. Bertsimas, "On probabilistic traveling salesman facility location problems," *Transportation Sci.*, vol. 3, pp. 184–191, 1989.
[8] ——, "The probabilistic minimum spanning tree problem," *Networks*, vol. 20, pp. 245–275, 1990.
[9] D. J. Bertsimas, P. Jaillet, and A. Odoni, "A priori optimization," *Oper. Res.*, vol. 38, no. 6, pp. 1019–1033, 1990.
[10] P. Jaillet, "A priori solution of a traveling salesman problem in which a random subset of the customers are visited," *Oper. Res.*, vol. 36, no. 6, pp. 929–936, 1988.
[11] ——, "Shortest path problems with node failures," *Networks*, vol. 22, pp. 589–605, 1992.
[12] P. Jaillet and A. Odoni, "The probabilistic vehicle routing problem," in *Vehicle routing: methods and studies*, B. L. Golden and A. A. Assad, Eds. Amsterdam: North Holland, 1988.
[13] L. Bianchi, J. Knowles, and N. Bowler, "Local search for the probabilistic traveling salesman problem: correlation to the 2-p-opt and 1-shift algorithms," *European J. Oper. Res.*, vol. 161, no. 1, pp. 206–219, 2005.
[14] C. Murat and V. T. Paschos, "On the probabilistic minimum coloring and minimum k-coloring," *Discrete Appl. Math.*, vol. 154, pp. 564–586, 2006.
[15] N. Bourgeois, F. Della Croce, B. Escoffier, C. Murat, and V. T. Paschos, "Probabilistic coloring of bipartite and split graphs," *J. Comb. Optimization*, vol. 17, no. 3, pp. 274–311, 2009.
[16] C. Murat and V. T. Paschos, "A priori optimization for the probabilistic maximum independent set problem," *Theoret. Comput. Sci.*, vol. 270, pp. 561–590, 2002, preliminary version available at http://www.lamsade.dauphine.fr/~paschos/documents/c166.pdf.
[17] ——, "The probabilistic minimum vertex-covering problem," *Int. Trans. Opl Res.*, vol. 9, no. 1, pp. 19–32, 2002, preliminary version available at http://www.lamsade.dauphine.fr/~paschos/documents/c170.pdf.
[18] ——, "The probabilistic longest path problem," *Networks*, vol. 33, pp. 207–219, 1999.
[19] V. T. Paschos, O. A. Telelis, and V. Zissimopoulos, "Steiner forests on stochastic metric graphs," in *Proc. Conference on Combinatorial Optimization and Applications, COCOA'07*, ser. Lecture Notes in Computer Science, A. Dress, Y. Xu, and B. Zhu, Eds., vol. 4616. Springer-Verlag, 2007, pp. 112–123.
[20] ——, "Probabilistic models for the STEINER TREE problem," *Networks*, to appear.
[21] M. R. Garey and D. S. Johnson, *Computers and intractability. A guide to the theory of NP-completeness*. San Francisco: W. H. Freeman, 1979.