

# Decision-Making Models in the Problem of Building a Minimum Cost Road System

Anatoly A. Zabelin <sup>1\*</sup>[ORCID 0000-0003-0501-4068],  
Evgenia S. Kogan <sup>1</sup>[ORCID 0000-0001-5681-6189]

<sup>1</sup> Transbaikal State University, Chita, Russia  
anatanza@mail.ru

## ABSTRACT

The article proposes methods for solving the problem of building a system of roads of minimum cost in the case when the costs of constructing roads are random variables. Earlier in the scientific literature, the issues of estimating the probabilistic characteristics (of such problems as, for example, the average weight of the edge of the minimal spanning graph), or the asymptotics of numerical characteristics were discussed. The authors of this article have set a goal that is to find the decision rules that can be used to implement the construction of a road system in practice since the totality of the results of previous studies does not allow us to eliminate the uncertainty in choosing a specific solution that is applicable in practice. The present article examines the cases when the distribution of the cost of constructing roads is discrete (in general) or continuous (using the example of three towns and independent and uniform distribution of the costs of building roads). In the discrete case, the authors propose decision-making methods in the form of the most probable minimum spanning tree, or a tree that maximizes the expected utility or minimizes the expected risk in the game with nature, associated with the original problem. In the continuous case, the authors consider an analytical approach (in situations with a small number of towns), which can be used to find the probability distribution on the set of minimum spanning trees, as well as a solution method based on the transition to the average values of the construction costs for each road. If there are a large number of towns, it is advantageous to use numerical approximations that is replacing a continuous distribution with a discrete one, or computational experiments in which the values of the cost of constructing roads are generated according to the given distributions, and then the resulting array of values is statistically processed.

**Keywords:** *minimal spanning tree, weighted complete graph, graph with random weights, decisive rules, the game with nature*

## 1. INTRODUCTION

There is a practical task: there are several towns; it is necessary to connect these settlements by a system of roads so that the total cost of constructing this system is minimal. The indicated practical problem is equivalent to the formal mathematical problem of extracting a spanning tree with minimum weight from a weighted complete undirected graph. For the case when the weights of the edges of the graph are fixed numbers, the problem is solved. A number of algorithms for constructing a minimum spanning tree were developed by R.C. Prim [1], J.B. Kruskal [2], O. Boruvka [3], and others.

In practice, the cost of constructing roads cannot

be estimated with a fixed number, its value fluctuates in a certain finite interval, or on an infinite ray  $(0; +\infty)$ . There are two mathematical models that can be applied to this situation: graphs with interval weights and graphs with random weights.

Graphs with interval weights are considered in [4-7]. In [4], the author addresses the optimization problem under conditions of uncertainty. Also in this work, various approaches to solving the problem of conditional optimization have been investigated. In [5], optimization problems on graphs with interval parameters have been studied. The definition of the problem of discrete optimization with interval parameters is provided. The method of reducing the interval problem of discrete optimization to the

problem of discrete multicriteria optimization is also considered there. In [6], a solution to the transport problem under conditions of uncertainty is given. In [7], one can get acquainted with the problem of spanning trees with topological criteria and interval weights. A valid solution is a spanning tree with interval weights. A number of specific implementations of graph weights are proposed for finding the optimal solution.

Turning to the sources that consider issues related to graphs whose weights are random variables, it is evident that the studies are devoted either to estimates of the average indicators (such as the expected weight of a spanning tree), or the asymptotics of the average indicators (when the number of vertices of the graph tends to infinity). Thus, in [8], the issues related to obtaining a minimum spanning tree in a graph whose weights are realizations of independent random variables are considered. In [9, 10], the asymptotics of the average weight of a minimal spanning tree extracted from a complete graph with weights distributed uniformly over a unit segment is studied. In [11], an approximate polynomial algorithm is proposed for solving the problem of finding a  $d$ -homogeneous connected spanning subgraph of maximum weight in a complete undirected weighted  $n$ -vertex graph. In [12], there is an overview of the accumulated knowledge about graphs with random edge weights: asymptotic estimates of the parameters of minimal spanning trees are described, the problem of a minimal spanning tree constructed on vertices randomly taken in Euclidean space is considered.

It should be noted that both in the case of interval weights and in the case of weights that are random variables, the problem of uncertainty of the solution persists. In practice, it is required to obtain some unambiguous mathematical solution (plan) and implement it. However, in the works on interval graphs, such a definite and unambiguous solution is not proposed, and in the case of considering random weights, an unambiguous solution is not discussed at all. This implies the practical significance and relevance of the present work, the purpose of which is to search for a decision-making algorithm in the case of uncertainty of graph weights.

## 2. MATERIALS AND METHODS

The purpose of this study is to develop a mathematical model of the problem of making a decision on the construction of a minimum cost road system connecting several towns, with the additional condition that the cost of constructing each road is

not a known number, but is a random variable. This model should give a very definite answer to the question which spanning tree should be adopted as a solution to be implemented in practice in the future.

To achieve the goal, it is necessary to solve the following tasks: (a) to select a mathematical object that is an abstraction of a real situation; (b) to determine in mathematical language the criteria for the optimality of solving the problem; (c) to develop a decisive rule or algorithm of obtaining an optimal solution for each proposed optimality criterion.

In the search for a solution, we use the methods of probability theory and mathematical statistics, combining them with the methods of graph theory and decision theory. In addition, as part of the pre- and post-analysis of the problem, we rely on computational experiments with randomization.

The main difficulty is that the research is carried out on the border of three branches of mathematics – probability theory, decision theory, and graph theory, each of which has its own mathematical apparatus.

Let us describe a mathematical object that will be used to simulate a real situation. Consider an undirected weighted complete graph  $G = \langle V, E \rangle$  ( $V$  is a set of vertices and  $E$  is a set of edges) with  $n$  vertices ( $|V| = n$ ), whose edge weights are random variables. The vertices of the graph correspond to the towns, and the edges of the graph correspond to the roads connecting these settlements. We call such a graph a graph with random weights. Suppose these random variables are either discrete or absolutely continuous. We limit the values of random variables to a set of positive real numbers. The construction of a road system is equivalent to choosing a minimum spanning tree in the specified graph.

To begin with, let us consider the case when the distributions of the weights of the edges of the graph are discrete random variables with finite sets of values. Let each edge  $e_{ij}$  ( $i, j = \overline{1, n}$ ,  $i < j$ ) of the complete graph have weight  $\omega_{ij}$  distributed according to law (1):

$$\mathbf{P}(\omega_{ij} = w_{ij}^k) = p_{ij}^k, \quad k = 1 \dots n_{ij}. \quad (1)$$

Suppose random variables are independent in pairs and in aggregate. Then the probability of realization of some combination of weights on all edges of the graph is equal to the product of the probabilities of these weights. For each combination of weights, one can find the minimum spanning tree and obtain a probability distribution on the set of minimum spanning trees.

We propose some rules for deciding on the choice of a spanning tree.

**The first decisive rule** can be formulated as follows: from all minimal spanning trees, it is necessary to choose the one whose probability is maximal.

**Example.** Consider a complete graph on three vertices:  $V = \{v_1, v_2, v_3\}$ . This graph has three edges:  $E = \{e_{12}, e_{23}, e_{13}\}$ . Let the weight of edge  $e_{12}$  take values 1 and 6 with probabilities 0.6 and 0.4, respectively, the weight of edge  $e_{23}$  take values 2 and 5 with probabilities 0.45 and 0.55, respectively, and the weight of edge  $e_{13}$  take values 3 and 4 with probabilities 0.6 and 0.4, respectively. In total, we have eight combinations of edge weights, information about which we summarize in *Table 1*, in the first three columns of which, there are the weights of the edges; in the fourth column, there is the probability of a combination of weights; in the fifth column, there is a set of edges that make up the minimum spanning tree (MST); in the last column, there is the total weight of the edges (the cost of constructing a road system).

**Table 1.** Discrete distribution of edge weights from the example

$w_{12}$	$w_{23}$	$w_{13}$	Prob.	MST	Weight
1	2	3	0.162	$e_{12}, e_{23}$	3
1	2	4	0.132	$e_{12}, e_{23}$	3
1	5	3	0.162	$e_{12}, e_{13}$	4
1	5	4	0.132	$e_{12}, e_{13}$	5
6	2	3	0.108	$e_{23}, e_{13}$	5
6	2	4	0.088	$e_{23}, e_{13}$	6
6	5	3	0.108	$e_{23}, e_{13}$	8
6	5	4	0.088	$e_{23}, e_{13}$	9

Source: Calculated by the authors

According to *Table 1*, the distribution of minimal spanning trees can be constructed. It can be seen that

**Table 3.** The payoff matrix, constructed according to the data from the example

	(1,2,3)	(1,2,4)	(1,5,3)	(1,5,4)	(6,2,3)	(6,2,4)	(6,5,3)	(6,5,4)
$e_{12}, e_{13}$	-4	-5	-4	-5	-9	-10	-9	-10
$e_{23}, e_{13}$	-5	-6	-8	-9	-5	-6	-8	-9
$e_{12}, e_{23}$	-3	-3	-6	-6	-8	-8	-11	-11

Source: Calculated by the authors

According to the criterion of maximum expected utility, which seems appropriate to be applied in our case, since the distribution of the states of nature is known, we get that the expected winnings of the first

player when they choose the first, second, or third row, respectively, are  $-6.32$ ,  $-6.81$  and  $-6.37$ . The maximum of them, equal to  $-6.32$ , is achieved when choosing the spanning tree  $\{e_{12}, e_{13}\}$ .

the situation in which the minimum spanning tree consists of edges  $e_{23}$  and  $e_{13}$  is most likely realized, and this probability is 0.392 (*Table 2*).

At the same time, the total weight of the edges included in the minimum spanning tree takes values in the range from 5 to 9, depending on the combination realized (generally speaking, these are the heaviest spanning trees). It is possible to calculate the average weight of the spanning tree, which is 6.95. It follows that there is no relationship between the weight of the spanning tree and the shape of the tree.

**Table 2.** Probability distribution of minimal spanning trees from the example

MST	$e_{12}, e_{13}$	$e_{23}, e_{13}$	$e_{12}, e_{23}$
Probability	0.294	0.392	0.294

Source: Calculated by the authors

Note that the probability of a situation in which the minimum spanning tree differs from tree  $\{e_{23}, e_{13}\}$  is 0.588, and this is higher than the probability of a situation in which the minimum spanning tree is tree  $\{e_{23}, e_{13}\}$ . This suggests that the probability of error in the decision-making process can be very high.

If we take the game with nature as a model for our task, then we can get **the second decisive rule**. The first player is the decision-maker. The second player is nature. Nature's strategies are combinations of edge weights. Each combination of weights has its own probability which can be calculated. The strategies of the first player are spanning trees. As a win for the first player, we take the value opposite to the weight of the spanning tree (since the first player spends this amount on constructing a tree).

All the winnings of the first player can be incorporated into a matrix. Thus, within the framework of the above example, the winning matrix has the following form (*Table 3*).

Similarly, the minimum expected risk criterion can be applied. In our example, it gives the same result that is choosing tree  $\{e_{12}, e_{13}\}$ .

Let us now consider the case when the weights of the edges have absolutely continuous distributions. In this study, we limit ourselves to uniform distributions of weights on a unit segment, and also consider only the case of the simplest graph — with three vertices.

Let weights  $w_{12} = X$ ,  $w_{23} = Y$ ,  $w_{13} = Z$  be distributed absolutely continuously with the following distribution density functions  $f_X(x)$ ,  $f_Y(y)$ ,  $f_Z(z)$ .

Find the probability that the minimum spanning tree is a tree in which edge  $e_{12}$  is excluded. This requires the simultaneous execution of two inequalities:  $X > Y$  and  $X > Z$ . Suppose random variables  $X, Y, Z$  are independent. Then the density function of their joint distribution is (2):

$$f(x, y, z) = f_X(x) \cdot f_Y(y) \cdot f_Z(z), \quad (2)$$

denote (3):

$$D =$$

$$= \{(x, y, z) / (x > y) \& (x > z), (x, y, z) \in \mathbb{R}\}, \quad (3)$$

then (4):

$$\begin{aligned} P(\{X > Y\} \& \{X > Z\}) &= \iiint_D f(x, y, z) dx dy dz = \\ &= \int_{-\infty}^{\infty} f_X(x) dx \left( \int_x^{\infty} f_Y(y) dy \cdot \int_x^{\infty} f_Z(z) dz \right) = \\ &= \int_{-\infty}^{\infty} (1 - F_Y(x))(1 - F_Z(x)) f_X(x) dx, \end{aligned} \quad (4)$$

where  $F_Y(\cdot)$ ,  $F_Z(\cdot)$  are integral distribution functions of random variables  $Y$  and  $Z$ .

Similarly, the probabilities that the minimum spanning tree will be a tree in which edge  $e_{23}$  or edge  $e_{13}$  is excluded are, respectively, (5), (6):

$$\int_{-\infty}^{\infty} (1 - F_X(y))(1 - F_Z(y)) f_Y(y) dy, \quad (5)$$

$$\text{and } \int_{-\infty}^{\infty} (1 - F_X(z))(1 - F_Y(z)) f_Z(z) dz, \quad (6)$$

In particular, if all the weights are distributed equally and uniformly on the unit interval with

density function  $f(\cdot)$  and distribution function  $F(\cdot)$ , then (7):

$$\begin{aligned} P(\{X > Y\} \& \{X > Z\}) &= \\ &= \int_{-\infty}^{\infty} (1 - F(x))^2 f(x) dx = \int_{-\infty}^{\infty} (1 - F(x))^2 dF(x) = \\ &= \int_0^1 (1 - t)^2 dt = \int_0^1 (1 - 2t + t^2) dt = \\ &= \left( t - t^2 + \frac{t^3}{3} \right) \Big|_0^1 = \frac{1}{3}, \end{aligned} \quad (7)$$

similarly, the remaining probabilities are  $\frac{1}{3}$ .

Consequently, it can be concluded that in the case when the edge weights of a complete three-vertex graph are distributed independently and uniformly on a unit segment, then the probability of any minimal spanning tree is the same, the distribution on minimal spanning trees is discrete and uniform. Therefore, the decision-maker can choose **any of the three options** for building a road system.

We propose **another decision-making procedure** in the case of continuous distributions of edge weights. The decision-maker can calculate the mathematical expectations of each weight  $Ew_{ij}$  and proceed to a solution with constant weights by associating weight  $Ew_{ij}$  with each edge  $e_{ij}$ . *Note that this approach can be applied not only to continuous weight distributions but also to discrete ones.*

Thus, in the conditions of Example 1, the mathematical expectations of the weights are  $Ew_{12} = 3$ ,  $Ew_{23} = 3.65$ , and  $Ew_{13} = 3.4$ , and the minimum spanning tree, in this case, is  $\{e_{12}, e_{13}\}$ .

### 3. RESULTS

Several decision-making procedures have been proposed in the case of a discrete distribution of edge weights:

- 1) choosing the most probable minimum spanning tree;
- 2) making a decision based on building a game with nature and using the method of maximum expected utility or minimum expected risk.

Having considered a special case of a continuous distribution of weights, we have obtained that the

decision-maker can with equal probability choose any of the possible spanning trees.

Also, a decision-making procedure by replacing the probabilistic distribution of edge weights with their mathematical expectations has been proposed. In this case, the decision-making is also unambiguous.

#### 4. DISCUSSION

Comparing the proposed methods of decision-making, we can draw the following conclusions:

- 1) The method of choosing the most probable spanning tree does not involve tree weights, therefore it can lead to large losses. The probability of an error (the decision-maker chose one spanning tree, and nature implemented a different situation with another MST) can also be quite high.
- 2) The decision-making method based on the consideration of the game with nature takes into account the weights of the edges and the probability distribution of the states of nature, therefore it is more acceptable than the method of choosing the most probable spanning tree.
- 3) If the weight distributions are absolutely continuous and independent, with a small number of edges, it is possible to obtain probability distributions on the set of minimal spanning trees and use this distribution to select the most probable MST. However, for sufficiently large graphs, the analytical method leads to complex calculations. In this case, to calculate the probabilities of the MST, one can proceed to the approximation of continuous distributions by discrete ones, or to the Monte Carlo method. The applicability of these methods requires additional research.
- 4) The method based on the transition from edge weight distributions to their mathematical expectations is perhaps the simplest method of practical problem-solving. Of course, with such a transition, there is a loss of information and accuracy of the solution.
- 5) It is necessary to develop an assessment of losses or the probability of error when making a decision. It may be worth considering the probabilities that a case that does not correspond to the decision taken will actually occur, or evaluating possible

deviations of the edge weights from their mathematical expectations.

#### 5. CONCLUSION

As a result of the conducted research, several methods have been proposed for solving the practical problem of constructing a road system with a discrete or continuous distribution of edge weights. Each method has its drawbacks. In the future, it is necessary to investigate the case of continuous edge weights more deeply, as well as to develop a methodology for estimating losses or the probability of error when making a decision.

#### AUTHORS' CONTRIBUTIONS

The authors have contributed to the work in the following percentage: 70% – Anatoly A. Zabelin, 30% – Evgenia S. Kogan.

#### REFERENCES

- [1] R.C. Prim, "Shortest connection networks and some generalizations", *Bell System Technical Journal*, 1957, vol. 36(6), pp. 1389-1401. DOI: 10.1002/j.1538-7305.1957.tb01515.x
- [2] J.B. Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem", in *Proceedings of the American Mathematical Society*, 1956, vol. 7(1), pp. 48-50. DOI: 10.1090/S0002-9939-1956-0078686-7
- [3] O. Borůvka, "About a certain minimal problem" [O jistém problému minimálním], *Práce Mor. Přírodověd. Spol. V Brně III*, 1929, vol. 3, pp. 37-58.
- [4] V.I. Levin, "Interval approach to optimization with uncertainty", *Systems of Control, Communication and Security*, 2015, vol. 4, pp. 123-141. (In Russ.).
- [5] V.A. Perepelitsa, I.V. Kozin, N.K. Maksishko, "Optimization Problems on Graphs with Interval Parameters", [Zadachi optimizatsii na graphah s interval'nymi parametrami], *Cybernetics and Systems Analysis*, 2009, vol. 2, pp. 3-14. (In Russ.).
- [6] T.A. Starostina, I.N. Rosenberg, "Transport problems solutions in uncertainty conditions", *Systems and Means of Informatics*, 2005, vol. 15(2), pp. 270-281. (In Russ.).
- [7] O.I. Shaposhnikova, L.G. Temirova, "Interval spanning tree problems on a topological

- criterion”, Polythematic Online Scientific Journal of Kuban State Agrarian University, 2016, vol. 115(1), pp. 369-378. (In Russ.).
- [8] G.S. Lueker, “Optimization problem on graphs with independent random edge weights”, *SIAM Journal of Computing*, 1981, vol. 10(2), pp. 338-351. DOI: 10.1137/0210024
- [9] A.M. Frieze, “On the value of a random minimum spanning tree problem”, *Discrete Applied Mathematics*, 1985, vol. 10(1), pp. 47-56. DOI: 10.1016/0166-218X(85)90058-7
- [10] C. Cooper, A. Frieze, N. Inze, S. Janson, J. Spencer, “On the length of a random minimum spanning tree”, *Combinatorics Probability and Computing*, 2016, vol. 25(1-2), pp. 89-107. DOI: 10.1017/S0963548315000024
- [11] A.E. Baburin, E.H. Gimadi, “An approximate algorithm for obtaining a d-uniform connected spanning subgraph of maximum weight in a complete graph with random edge weights”, [Priblizhennyj algoritm poiska d-odnorodnogo svyaznogo ostovnogo podgrafa maksimal'nogo vesa v polnom grafe so sluchajnymi vesami reber], *Discrete Analysis and Operations Research*, ser. 2, 2006, vol. 13(2), pp. 1-21. (In Russ.).
- [12] A. Riva, “The Random Minimum Spanning Tree Problem”, *Universita Degli Studi di Milano, Anno Academic*, 2019, 82 p.