



## A Heuristic Algorithm for Optimal Hamiltonian Cycles in Weighted Graphs

Tadeusz Ostrowski

The Jacob of Paradyż University of Applied Sciences in Gorzów Wlkp, Poland ul. Teatralna 25, 66-400 Gorzów Wlkp.

Petroula M. Mavrikiou

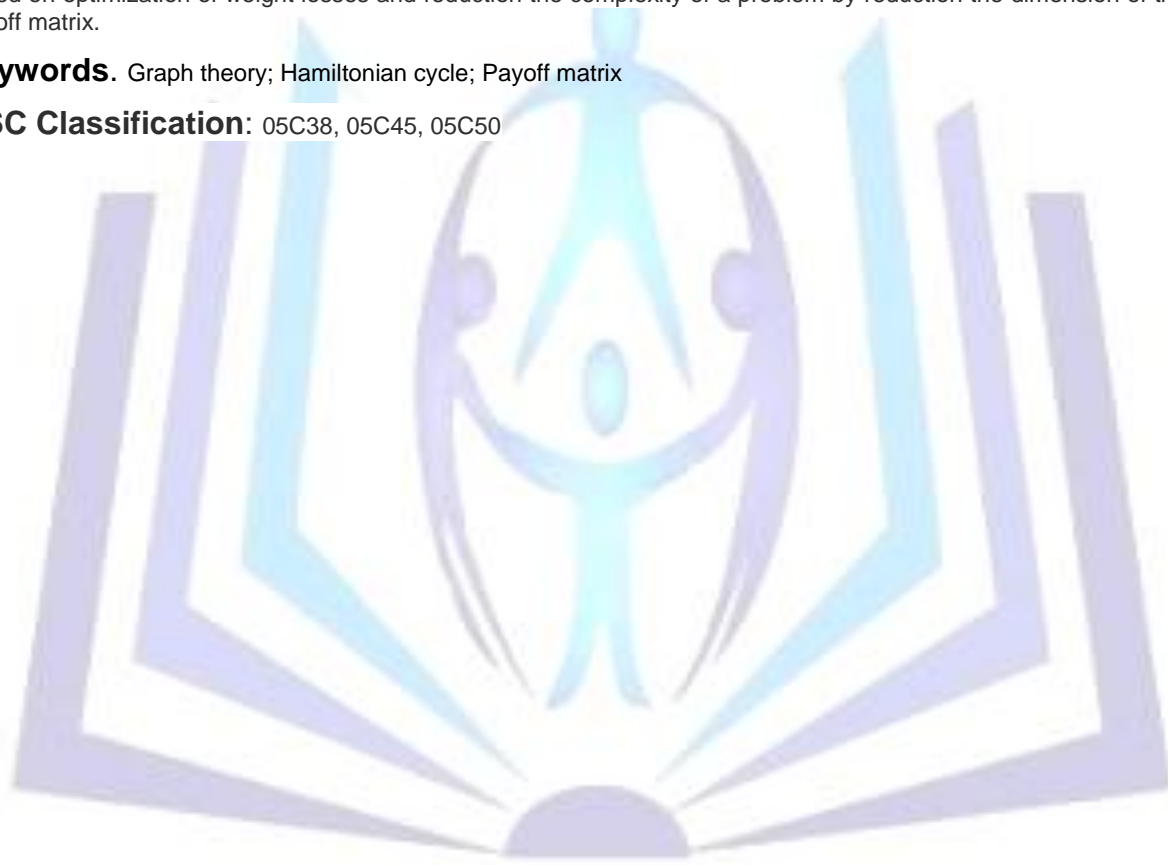
School of Economic Sciences and Administration, Frederick University P.O. Box 24 729, 1303 Nicosia, Cyprus

### AFSTRACT

The paper focuses on finding of the optimal Hamiltonian cycle, when it is regarded with respect to cost, time, distance or difficulty level of the route. The problem is strictly related to the traveling salesman problem proved to be NP-complete for general graphs. The paper gives a heuristic algorithm for finding the optimal spanning cycle in a weighted graph. Its idea is based on optimization of weight losses and reduction the complexity of a problem by reduction the dimension of the graph payoff matrix.

**Keywords.** Graph theory; Hamiltonian cycle; Payoff matrix

**MSC Classification:** 05C38, 05C45, 05C50



## Council for Innovative Research

Peer Review Research Publishing System

Journal: JOURNAL OF ADVANCES IN MATHEMATICS

Vol .11, No.6

[www.cirjam.com](http://www.cirjam.com) , [editorjam@gmail.com](mailto:editorjam@gmail.com)



## INTRODUCTION

Investigating cycles (closed paths) running through every vertex of a weighted undirected graph we will restrict out to simple graphs, i.e. the graphs have no self loops and no parallel arcs (multi-edges). Otherwise, the loops can be removed and the arc with the optimal weight - the smallest or the biggest - out of the parallel arcs can be chosen. The study of Hamiltonian cycles has been motivated by practical applications (see, for example, [4] or [9]).

Let be given a set of  $n$  objects (vertices of a graph). Between any of those two vertices there is a path but it can also happen that two vertices are not directly connected and from a vertex to another vertex you can get through one or even more different vertices. Every path is related to exactly one number (weight) which expresses cost, time, distance or difficulty of the given path. Hamilton cycle does not exist for all graphs. Unfortunately, there is no specific way to find whether Hamilton cycle exists or not.

## RESULTS

### Losses Minimization (Maximization) Algorithm

In the paper is assumed that the graph is not trivial and all paths connected with a vertex are not related to the same number. For the given graph we can construct a payoff matrix  $\mathbf{C}^{n \times n}$ , called often a distance matrix, where for all  $k, l = 1, 2, \dots, n$  the elements of the payoff matrix are as follow:

$$c_{kl} \text{ is a positive number, if there is a path from } k \text{ to } l, \quad (1)$$

It is also assumed that if there are no loops then the weight of "the path" from any vertex to itself is equal to zero. It can be said zero indicates the forbidden path or there is no path. Some authors adopt infinity for  $k = l$ .

$$c_{kl} = 0, \text{ if } k = l \text{ or there is no path from } k \text{ to } l. \quad (2)$$

If  $c_{kl} = c_{lk}$  ( $k, l = 1, 2, \dots, n$ ), then the payoff matrix  $\mathbf{C}$  is symmetric, i.e.  $\mathbf{C} = \mathbf{C}^T$ , where  $\mathbf{C}^T$  denotes transposed matrix. In general, it is not necessary for the matrix  $\mathbf{C}$  to be symmetric. But if  $\mathbf{C}$  is symmetric then only rows or only columns can be investigated. Suppose that the given payoff matrix  $\mathbf{C}$  represents a graph containing Hamiltonian cycles. Losses Minimization (Maximization) Algorithm (LMA to be short), as adopted to obtain minimum total weight, goes as follows:

- for all rows and columns of the payoff matrix  $\mathbf{C}$  differences between the smallest nonzero elements are determined;
- the bigger difference gives the loss of higher weights so the smallest element in the row or column with the biggest difference indicates one path of the optimal Hamiltonian cycle;
- the element, containing the obtained path, is replaced with zero and it leads to the new reduced payoff matrix;
- differences between the smallest nonzero elements in the reduced payoff matrix are investigated, and so on.

Formally, denote for any fixed  $k = 1, \dots, n$

$$d_k = |c'_{kk} - c'_{ik}|, \quad (3)$$

$$d_{kj} = |c'_{kj} - c'_{kj}|, \quad (4)$$

where  $c'_{kk}, c'_{kk}$  ( $c'_{kj}, c'_{kj}$ ) are two smallest elements in the  $k$ -th column ( $k$ -th row, respectively) of the payoff matrix  $\mathbf{C}$ . Let denote

$$m = \max(d_j) \quad (5)$$

The smallest element in the row (or column), containing the difference  $m$ , determines the path which must be chosen. It implies that the smallest element of the row (or column) can be replaced by zero, and as the result new payoff matrix is created. The procedure has to be continued  $(n - 1)$  times, and the last path of the optimal Hamiltonian cycle will be determined uniquely.

If LMA is adopted to maximum total weight, then for all rows and columns of the payoff matrix  $\mathbf{C}$  differences of the biggest nonzero elements should be determined because bigger difference implies losing smaller weights. Next in the row (column) with the biggest difference the biggest element in that row (column, respectively) must be chosen. As the result, one path of the optimal Hamiltonian cycle is obtained. Formally, let denote for any fixed  $k = 1, \dots, n$

$$D_k = |c''_{kk} - c''_{ik}| \quad (6)$$

$$D_{kj} = |c''_{kj} - c''_{kj}| \quad (7)$$

where now  $c''_{kk}, c''_{kk}$  ( $c''_{kj}, c''_{kj}$ ) mean two biggest elements of the  $k$ -th column ( $k$ -th row, respectively) of the payoff matrix  $\mathbf{C}$ . Denote

$$M = \max(D_j) \quad (8)$$

The biggest element in the row (or column), containing the difference  $M$ , indicates the path which must be chosen. As the result we obtain a reduced payoff where the biggest element is replaced by zero. The procedure above has to be continued  $(n - 1)$  times, and the last path is determined.

### Example (symmetric payoff matrix)



Given a graph of weighted edges of a tourist route with nodes A, B, C, D, E, which state for some tourist objects. Assume that weights express time, cost or level of difficulty of the edges (the less, the better) for minimum, and beauty of view or attractiveness (the more, the better) for maximum. The objective functions (minimum and maximum) should be find for the following payoff matrix of the route graph

$$\begin{matrix}
 \mathbf{A} \\
 \mathbf{B} \\
 \mathbf{C} \\
 \mathbf{D} \\
 \mathbf{E}
 \end{matrix}
 \begin{bmatrix}
 0 & 7 & 4 & 5 & 4 \\
 7 & 0 & 2 & 1 & 5 \\
 4 & 2 & 0 & 3 & 2 \\
 5 & 1 & 3 & 0 & 6 \\
 4 & 5 & 2 & 6 & 0
 \end{bmatrix}
 \tag{9}$$

Let start from the minimum. The payoff matrix is symmetric (the graph is undirected) and according to (3) we have

$$d_{1j} = |c'_{13} - c''_{15}| = d_{3j} = |c'_{32} - c''_{35}| = 0,$$

$$d_{2j} = |c'_{23} - c''_{24}| = 1,$$

$$d_{4j} = |c'_{42} - c''_{43}| = d_{5j} = |c'_{53} - c''_{51}| = 2$$

Therefore, by (5),  $m = \max(d_{ij}) = d_{4j} = d_{5j}$ .

Since then we obtain  $\min(c_{4j}) = c_{42} = 1$ ,  $\min(c_{5j}) = c_{53} = 2$ .

and two paths of the route are determined. The two paths are DB (or BD) and EC (or CE). Putting  $c_{23} = c_{32} = c_{45} = c_{54} = 0$ , the payoff matrix (9) will have the form

$$\begin{matrix}
 \mathbf{A} \\
 \mathbf{B} \\
 \mathbf{C} \\
 \mathbf{D} \\
 \mathbf{E}
 \end{matrix}
 \begin{bmatrix}
 0 & 7 & 4 & 5 & 4 \\
 7 & 0 & 2 & 0 & 5 \\
 4 & 2 & 0 & 3 & 0 \\
 5 & 0 & 3 & 0 & 6 \\
 4 & 5 & 0 & 6 & 0
 \end{bmatrix}
 \tag{10}$$

Since  $d_{2j} = |c'_{23} - c''_{25}| = 3$  is the maximal difference, it implies that the path of the minimal route must be BC (or CB). Because the nodes B, C have been taken twice, therefore all elements of the second and third row, column in (10) are equal to zero and can be omitted.

$$\begin{matrix}
 \mathbf{A} \\
 \mathbf{D} \\
 \mathbf{E}
 \end{matrix}
 \begin{bmatrix}
 0 & 5 & 4 \\
 5 & 0 & 6 \\
 4 & 6 & 0
 \end{bmatrix}
 \tag{11}$$

It is easy to see in (11) that  $d_{5j} = |c'_{51} - c''_{54}| = 2$  is the maximal difference and next path should be AE (or EA). At last the path AD (or DA) is obtained and the minimal route (if A is the beginning of the cycle and D is its next node) has the form

$$\min_C = A \rightarrow D \rightarrow B \rightarrow C \rightarrow E \rightarrow A, f(\min_C) = 5 + 1 + 2 + 2 + 4 = 14.$$

Applying LMM for maximum, according to (6) we have now

$$D_{1j} = |c'_{12} - c''_{14}| = D_{2j} = |c'_{21} - c''_{25}| = 2,$$

$$D_{3j} = |c'_{31} - c''_{34}| = D_{4j} = |c'_{45} - c''_{41}| = D_{5j} = |c'_{54} - c''_{52}| = 1.$$

By (8) we obtain  $\text{Max}(D_{ij}) = D_{1j} = D_{2j}$ .

Both first and second row has the same maximal element, but second row should be chosen because the loss sum of weights is less than for first row. Therefore one path of the route, AB (or BA) is determined. The reduced matrix has the form

$$\begin{matrix}
 \mathbf{A} \\
 \mathbf{B} \\
 \mathbf{C} \\
 \mathbf{D} \\
 \mathbf{E}
 \end{matrix}
 \begin{bmatrix}
 0 & 0 & 4 & 5 & 4 \\
 0 & 0 & 2 & 1 & 5 \\
 4 & 2 & 0 & 3 & 2 \\
 5 & 1 & 3 & 0 & 6 \\
 4 & 5 & 2 & 6 & 0
 \end{bmatrix}
 \tag{12}$$



It is easy to see that  $\text{Max}(D_{ij}) = D_{2j} = 3$  and  $c_{25} = 5$  is the maximal element in second row. Since then the next determined path is BE (or EB), and the node B appears twice (both as the beginning and as the end of a path). In (12) all elements of second row and columns, will be replaced by zero and the reduced matrix is now as follows

$$\begin{array}{l} \mathbf{A} \\ \mathbf{C} \\ \mathbf{D} \\ \mathbf{E} \end{array} \begin{bmatrix} 0 & 4 & 5 & 4 \\ 4 & 0 & 3 & 2 \\ 5 & 3 & 0 & 6 \\ 4 & 2 & 6 & 0 \end{bmatrix} \quad (13)$$

The next path must be ED (or DE). Reducing it we get the payoff matrix

$$\begin{array}{l} \mathbf{A} \\ \mathbf{C} \\ \mathbf{D} \\ \mathbf{E} \end{array} \begin{bmatrix} 0 & 4 & 5 & 4 \\ 4 & 0 & 3 & 2 \\ 5 & 3 & 0 & 0 \\ 4 & 2 & 0 & 0 \end{bmatrix} \quad (14)$$

The next chosen path must be DC (or CD). Note that it cannot be DA (or AD), because it leads to the closed path but not Hamiltonian cycle. At the end we get the path CA (or AC), and the maximal route (if A is the beginning of the cycle and B is its next node) is

$$\max_C = A \rightarrow B \rightarrow E \rightarrow D \rightarrow C \rightarrow A, f(\max_C) = 7 + 5 + 6 + 3 + 4 = 25$$

### Example (not symmetric payoff matrix)

Let be given a weighted graph of a route with nodes A, B, C, D, E, which state for some tourist objects. Assume that weights express time, cost or level of difficulty, beauty of view or attractiveness of the edges, and the weights depend on direction of the path between nodes. The objective function is the minimum (maximum, respectively) of weights, and the payoff matrix of the graph has the form

$$\begin{array}{l} \mathbf{A} \\ \mathbf{B} \\ \mathbf{C} \\ \mathbf{D} \end{array} \begin{bmatrix} 0 & 12 & 19 & 7 \\ 11 & 0 & 7 & 8 \\ 8 & 1 & 0 & 5 \\ 4 & 3 & 14 & 0 \end{bmatrix} \quad (15)$$

At first the minimum will be found. Since the payoff matrix is asymmetric now, therefore differences of nonzero minimal elements both rows and columns should be investigated. By formulas (3) and (4), we obtain for rows:

$$d_{1j} = |c'_{14} - c''_{12}| = 5, d_{2j} = |c'_{23} - c''_{24}| = 1,$$

$$d_{3j} = |c'_{32} - c''_{34}| = 4, d_{4j} = |c'_{41} - c''_{42}| = 1,$$

and for columns:

$$d_{i1} = |c'_{31} - c''_{41}| = 4, d_{i2} = |c'_{32} - c''_{42}| = 2,$$

$$d_{i3} = |c'_{23} - c''_{43}| = 7, d_{i4} = |c'_{34} - c''_{14}| = 2,$$

Since then, by (5), we have

$$m = \max(d_{ij}) = d_{i3} = 7.$$

Therefore

$$\min(c_{i3}) = c_{23} = 7.$$

And we obtain one path of the route; it is BC. Consequently, the payoff matrix (15) reduces now to the form given below

$$\begin{array}{l} \mathbf{A} \\ \mathbf{B} \\ \mathbf{C} \\ \mathbf{D} \end{array} \begin{bmatrix} 0 & 12 & 19 & 7 \\ 11 & 0 & 0 & 8 \\ 8 & 0 & 0 & 5 \\ 4 & 3 & 14 & 0 \end{bmatrix} \quad (16)$$

By formulas (3) and (4), we obtain for rows:



$$d_{1j} = |c'_{12} - c''_{14}| = 5, d_{2j} = |c'_{21} - c''_{24}| = 3, \\ d_{3j} = |c'_{34} - c''_{31}| = 3, d_{4j} = |c'_{42} - c''_{41}| = 1,$$

and for columns:

$$d_{i1} = |c'_{41} - c''_{31}| = 4, d_{i2} = |c'_{42} - c''_{12}| = 9, \\ d_{i3} = |c'_{43} - c''_{41}| = 5, d_{i4} = |c'_{34} - c''_{14}| = 2,$$

Since then, by (5), we have

$$m = \max(d_{ij}) = d_{i2} = 9.$$

Therefore

$$\min(c_2) = c_{42} = 3.$$

And we obtain another path of the route; it is DB. Since the node B occurs both as the beginning and the end of the path, the payoff matrix reduces now to the form as follows

$$\begin{matrix} \mathbf{A} \\ \mathbf{B} \\ \mathbf{C} \\ \mathbf{D} \end{matrix} \begin{bmatrix} 0 & 0 & 19 & 7 \\ 0 & 0 & 0 & 0 \\ 8 & 0 & 0 & 5 \\ 4 & 0 & 14 & 0 \end{bmatrix} \tag{17}$$

Another path of the route is AD, and it implies, since the node D occurs both as the beginning and the end of paths, that the payoff matrix reduces now to the form as follows

$$\begin{matrix} \mathbf{A} \\ \mathbf{C} \end{matrix} \begin{bmatrix} 0 & 19 \\ 8 & 0 \end{bmatrix} \tag{18}$$

The reduced payoff matrix (18) implies that the last path must be CA, and the minimum (if A is the beginning of the cycle) is

$$\min_C = A \rightarrow D \rightarrow B \rightarrow C \rightarrow A, f(\min_C) = 7 + 3 + 7 + 8 = 25.$$

The maximal sum of weights will be determined for the same payoff matrix (15). According to (6), (7) we have for rows:

$$D_{1j} = |c'_{13} - c''_{12}| = 7, D_{2j} = |c'_{21} - c''_{24}| = 3, \\ D_{3j} = |c'_{31} - c''_{34}| = 3, D_{4j} = |c'_{43} - c''_{41}| = 10,$$

and for columns:

$$D_{i1} = |c'_{21} - c''_{31}| = 3, D_{i2} = |c'_{12} - c''_{42}| = 9, \\ D_{i3} = |c'_{13} - c''_{43}| = 5, D_{i4} = |c'_{24} - c''_{14}| = 1.$$

Since then, by (5), we have

$$M = \max(D_{ij}) = d_{4j} = 10.$$

Therefore

$$\text{Max}(c_{4j}) = c_{43} = 14.$$

And we obtain one path of the maximum route; it is DC. Since then, third column must consist all zeros and, the payoff matrix reduces to the form given below

$$\begin{matrix} \mathbf{A} \\ \mathbf{B} \\ \mathbf{C} \\ \mathbf{D} \end{matrix} \begin{bmatrix} 0 & 12 & 0 & 7 \\ 11 & 0 & 0 & 8 \\ 8 & 1 & 0 & 0 \\ 4 & 3 & 0 & 0 \end{bmatrix} \tag{19}$$

It is easy to check that next paths of the maximum are CA, AB and BD. Therefore (if A is the beginning of the cycle) we obtain

$$\max_C = A \rightarrow B \rightarrow D \rightarrow C \rightarrow A, f(\max_C) = 12 + 8 + 14 + 8 = 42.$$

### DISCUSSION, REMARKS



As it is well known Richard M. Karp showed that the Hamiltonian cycle problem was NP-hard in combinatorial optimization. For  $n$  nodes the number of possible Hamiltonian cycles may be equal to  $(n - 1)! / 2$  what implies increase in very rapid speed with the number of nodes (dimension of the payoff matrix). Evaluating and testing all Hamiltonian cycles has exponential order and is generally too time consuming to be feasible for most problems (for instance, for 20 nodes it takes many months). Therefore, finding a solution with the LMA seems to be useful in practice.

In literature can be found many algorithms used to determine a solution to traveling salesman problem. One of the first was The Nearest Neighbour Algorithm (see, for example [2], [3], [6]) in which, general idea, the tourist starts at a random node and repeatedly visits the nearest vertex until all have been visited. It quickly yields a short tour, but there are some disadvantages – it might not give the optimal route or might lead to an incomplete route or might not find a feasible route at all, even when one exists. Very successful in solving the traveling salesman problem are evolutionary algorithms like Genetic Algorithms – which consist of Encoding, Evaluation, Crossover, Mutation and Decoding (more details can be found in [1], [8]) or Ant Colony Algorithms, proposed by Marco Dorigo in his PhD thesis (see [5] or [7]).

In LMA if the payoff matrix is non symmetric, then in the payoff matrix one row or one column can be reduced, when the vertex stands for the beginning or the end of a path.

It is easy to see that  $\min_C$  and  $\max_C$  are dual concepts and searching  $\max_C$  can be easily replaced by searching  $\min_C$ . Formally, let  $\mathbf{A} = [a_{ij}]$  be a payoff matrix. Construct the payoff matrix  $\mathbf{B} = [b_{ij}]$  of the same dimensions, where for any fixed number  $k > \max(a_{ij})$  we have

$$b_{ij} = 0, \text{ if } a_{ij} = 0, \text{ and } b_{ij} = k - a_{ij}, \text{ if } a_{ij} \neq 0.$$

Then the following equations hold

$$\begin{aligned} \min_{\mathbf{A}} &= \max_{\mathbf{B}}; \\ f(\min_{\mathbf{A}}) &= kn - f(\max_{\mathbf{B}}). \end{aligned}$$

If in an investigating graph there is no Hamiltonian cycle then the LMA method will stop at some step.

If in some rows (columns) the biggest differences are equal then the biggest (the smallest element in the row (column) with smallest sum of weights (biggest sum of weights, respectively) should be chosen

Tourism, for example, has become a very popular global and local leisure activity and designing for a tourist route is very important. Having a lot of options of a route for walking, cycling, horse riding, hiking, ... is beneficial to a decision maker in tourism industry. The decision maker has to propose the optimal cycle for his clients. From that point of view, or more precisely, in combinatorial optimization the problem is related to the travelling salesman problem which is one of the most intensively studied problems in computational mathematics and might seem to be of interest to tourist companions.

The problem appears when a tourist companion wants to design a route containing interesting cultural and historic attractions (nodes or vertices of a graph) in a city. Therefore the resulting route will start from one node and pass all other vertices of the graph and come back to the beginning node. The route should be optimal, giving minimum or maximum total weight, from some point of view (for example, should be the most interesting, not necessarily the shortest) and passing through every edge of the graph (a street of the city) only once. In other words, the route should create Hamiltonian cycle.

## REFERENCES

- [1] Y. Alkhalifaf and R. L. Wainwright. (2004). A Genetic Algorithm Applied to Graph Problems Involving Subsets of Vertices. *Evolutionary Computation*, Vol. 1, pp. 303–308
- [2] J. Bang-Jensen, G. Gutin and A. Yeo. (2004). When the greedy algorithm fails. *Discrete Optimization* 1, pp. 121-127.
- [3] G. Bendall and F. Margot. (2006). Greedy Type Resistance of Combinatorial Problems. *Discrete Optimization* 3, pp. 288-298.
- [4] D. Davendra, editor. (2010) *Travelling Salesman Problem, Theory and Applications*, InTech Publisher
- [5] M. Dorigo (1992). *Optimization. Learning and Natural Algorithms*, PhD thesis. Politecnico di Milano
- [6] G. Gutin, A. Yeo and A. Zverovich. (2002). Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the TSP. *Discrete Applied Mathematics* 117, pp. 81-86.
- [7] K. Li et al. (2008). Comparative Analysis of Genetic Algorithm and Ant Colony Algorithm on Solving Traveling Salesman Problem. *Semantic: Computing and Systems*, pp. 72-75
- [8] Z. Michalewicz. (1994). *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag
- [9] H. E. Stanley and S. V. Buldyrev. (2001). The salesman and the tourist. *Nature* 413, pp. 373–374