A SURVEY ON THE COMPLEXITY INDICES FOR THE TRAVELLING SALESMAN PROBLEM

Vera Kovačević-Vujčić

Faculty of Organizational Sciences, University of Belgrade, verakov@fon.rs

Coauthors:

Dragoš Cvetković, Mathematical Institute SANU, Belgrade, ecvetkod@etf.rs

Mirjana Čangalović, Faculty of Organizational Sciences, University of Belgrade, canga@fon.rs

Zorica Dražić, Faculty of Mathematics, University of Belgrade, lolaz@sezampro.rs

Applications of graph spectra in combinatorial optimization

For example, one of early heuristics for graph bisection uses the *Fiedler vector*, i.e. the eigenvector belonging to the second smallest eigenvalue of the graph Laplacian. This eigenvalue is called *algebraic connectivity* of the graph and was introduced by M. Fiedler in the paper

Algebraic connectivity of graphs, Czech. J. Math., 23(98)(1973), 298-305.

The algebraic connectivity has been used in

D. Cvetković, M. Čangalović and V. Kovačević-Vujčić, Semidefinite programming methods for the symmetric travelling salesman problem, *Integer Programming and Combinatorial Optimization, Proc. 7th Internat. IPCO Conf.*, Graz, Austria, June 1999, Lecture Notes Comp. Sci. 1610, Springer, Berlin, 1999, 126-136.

to formulate the following discrete semidefinite programming model of the symmetric *travelling salesman problem* (STSP):

STSP semidefinite programming model:

minimize $F(X) = \Sigma_{i=1}^n \Sigma_{j=1}^n \left(-\frac{1}{2}d_{ij}\right) x_{ij} + \frac{\alpha}{2} \Sigma_{i=1}^n \Sigma_{j=1}^n d_{ij}$

subject to

$x_{ii} = 2 + \alpha - \beta \;\; (i = 1, \ldots, n),$

$\Sigma_{j=1}^n x_{ij} = n\alpha - \beta, \;\; (i = 1, \ldots, n),$

$x_{ij} \in \{\alpha - 1, \alpha\} \;\; (j = 1, \ldots, n : \; i < j), \qquad X \geq 0$

Here $X \geq 0$ means that the matrix X is symmetric and positive semidefinite, while α and β are chosen so that $\alpha > h_n/n$ and $0 < \beta \leq h_n$ with $h_n = 2 - 2\cos(2\pi/n)$ being the algebraic connectivity of the cycle C_n.

Solution X essentially gives the adjacency matrix of an optimal Hamiltonian cycle.

A natural semidefinite relaxation of the travelling salesman problem is obtained when discrete conditions are replaced by inequality conditions.

SDP relaxation:

minimize $F(X)$

subject to

$x_{ii} = 2 + \alpha - \beta \ \ (i = 1, \ldots, n),$

$\Sigma_{j=1}^{n} x_{ij} = n\alpha - \beta, \ \ (i = 1, \ldots, n),$

$\alpha - 1 \leq x_{ij} \leq \alpha, \ i, j = 1, \ldots, n, i < j, \quad X \geq 0$

Branch and bound algorithm

Definition of a complexity index

Let A be an (exact) algorithm for solving an NP-hard combinatorial optimization problem C and let I be an instance of C of dimension n. A complexity index of I for C with respect to A is a real r, computable in polynomial time from I, by which we can predict (in a well defined statistical sense) the execution time of A for I.

Correlation coefficient

The efficiency of the complexity index can be statistically estimated measuring the linear correlation between the index value and the number of relaxation tasks solved within the B&B algorithm.

The coefficient of linear correlation for two sequences (b_i) and (c_i) is defined by

$$C_{BC} = \frac{1}{\sqrt{\vartheta_B}\sqrt{\vartheta_C}} \sum_{i=1}^{|\mathcal{N}|} (b_i - \bar{m}_B)(c_i - \bar{m}_C),$$

where $\bar{m}_B$, $\bar{m}_C$ and ϑ_B, ϑ_C are mean values and variances of the corresponding sequences (b_i) and (c_i), respectively.

TSP – short edge subgraph

We consider the symmetric travelling salesman problem with instances I represented by complete graphs G with distances between vertices (cities) as edge weights (lengths).

Intuitively, the hardness of an instance G depends on the distribution of short edges within G.

Therefore we consider some short edge subgraphs of G (minimal spanning tree, critical connected subgraph, and several others) as non-weighted graphs and several their invariants as potential complexity indices.

How short an edge should be to be considered as *short* depends on the context.

Open tour TSP – B&B with MST as relaxation

If a minimal spanning tree is a path, it represents also a solution to the TSP. However, a path is also a tree with a minimal branching extent (in an intuitive sense).

The main idea is based on the expectation that a branch and bound algorithm will run *for longer the more the minimal spanning tree deviates from a path*, i.e. the greater “branching extent” it has.

Accordingly, any graph invariant characterizing well the “branching extent” in an intuitive sense, can be considered as a complexity index for the travelling salesman problem.

Graph invariants as complexity indices

The following invariants have been considered:

- D the number of vertices of degree 2 in the minimal spanning tree;
- λ_1 the largest eigenvalue of the adjacency matrix of the minimal spanning tree.

The quantity D is maximal ($D = n-2$, where n is the number of vertices) if the tree reduces to a path, but it attains its minimal value $D = 0$ on a great number of trees.

The largest eigenvalue λ_1 reflects more precisely the branching extent of a tree.

Given n, the number of vertices of a tree, the quantity λ_1 varies between $2cos\frac{\pi}{n+1}$ and $\sqrt{n-1}$, both bounds being attained on exactly one tree (a path and a star, respectively).

Since the path P_n has the least branching extent in the intuitive sense and the star $K_{1,n-1}$ has the maximal one, the quantity λ_1 has at least a good property that it characterizes extremal trees in the above sense.

Any invariant which is considered as a branching extent parameter should fulfil this criterion.

Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be eigenvalues (of the adjacency matrix) and $d_1, d_2, \ldots, d_n$ vertices degrees of vertices in a graph. The quantity

$$S_k = \sum_{i=1}^{n} \lambda_i^k \quad (k = 0, 1, 2, \ldots)$$

is called the *k-th spectral moment* of the graph.

For trees we have

$$S_4 = 6(n-1) - 2\sum_{i=1}^{n} d_i^2 \quad .$$

It has been shown that S_4 is maximal for the star $K_{1,n-1}$ and minimal for the path P_n.

The following two indices are also considered:
$F_s = d_1! + d_2! + \cdots + d_n!$ and $F_p = d_1! d_2! \cdots d_n!$.

Experiments in 1980's

A number of instances of the TSP have been generated by means of a random generator using a uniform distribution in the interval (0,1) for the weights.

For each instance we have computed the considered indices and the number N of the solved relaxation tasks when running the branch and bound algorithm.

Since the input matrices are randomly generated the indices and the number N are random variables. The linear correlation coefficient and the Spearman correlation coefficient between indices and N have been calculated.

The results are given in Tables 1-4.

	N	$\log N$	N (Spearman)
D	0.361	0.605	0.500
λ_1	0.433	0.598	0.514
S_4	0.501	0.612	0.536
F_s	0.463	0.421	0.568
F_p	0.531	0.534	0.570

Table 1: 200 graphs on 8 vertices

	N	$\log N$	N (Spearman)
D	0.293	0.529	0.439
λ_1	0.378	0.543	0.502
S_4	0.469	0.587	0.545
F_s	0.643	0.375	0.530
F_p	0.690	0.517	0.551

Table 2: 100 graphs on 12 vertices

	N	$\log N$	N (Spearman)
D	0.349	0.499	0.477
λ_1	0.225	0.408	0.407
S_4	0.340	0.505	0.571
F_s	0.036	0.179	0.493
F_p	0.269	0.359	0.575

Table 3: 100 graphs on 14 vertices

	N	$\log N$	N (Spearman)
D	0.372	0.513	0.542
λ_1	0.383	0.333	0.329
S_4	0.498	0.444	0.537
F_s	0.206	0.144	0.473
F_p	0.235	0.168	0.581

Table 4: 100 graphs on 16 vertices

The first and the second column give the correlation coefficient between all mentioned indices and quantities N and $\log N$. The Spearman rank correlation coefficient is given in the third column.

Experiments in 1990's

Several invariants have been considered as complexity indices for the TSP with respect to B&B algorithms based on SDP relaxation:

D. Cvetković, M. Čangalović, V. Kovačević-Vujčić, *Complexity indices for the travelling salesman problem based on a semidefinite relaxation*, SYM-OP-IS '99, Proc. XXVI Yugoslav Symp. Operations Research, Beograd, 1999, 177–180.

Let X be the solution of SDP relaxation and $L = X + h_n I - J$.

Then L determines the weighted graph $W_L = (V, E_L, C_L)$, where $E_L = \{\{i,j\} \in E | l_{ij} < 0\}$ and $C_L = 2I - L$,

the corresponding unweighted graph $G_L = (V, E_L)$

and a stochastic matrix $S_L = I - \frac{1}{2}L$.

The most efficient indices are the following:

I_1: the number of edges of G_L

I_2: the second smallest eigenvalue of the Laplacian of G_L

I_3: the entropy of S_L, i.e. value equal to

$$\sum_{(i,j)\in E_L} (l_{ij}/2)\log_2(-l_{ij}/2) - n/2$$

I_4: $\sum_{i=1}^{n} |\mu_i - \mu_i^*|$, where $\mu_1, \mu_2, \ldots, \mu_n$ and $\mu_1^*, \mu_2^*, \ldots, \mu_n^*$ are sequences of nondecreasing eigenvalues of the Laplacians of G_L and a Hamiltonian circuit, respectively.

I_5: the same sum as in I_4 but with eigenvalues of the Laplacian of W_L instead of G_L.

I_6: the number of vertices of the G_L with degrees greater than 2.

The efficiency of indices I_k, $k = 1, \ldots, 6$, has been investigated For each dimension 20, 25, 30, 35 we consider 50 randomly generated TSP instances with distances uniformly distributed in the interval [1,999]. To each instance a B&B algorithm based on SDP relaxation is applied.

The coefficients of the linear correlation between values of indices $I_k (k = 1, \ldots, 6)$ and the number of relaxation tasks for dimensions n= 20, 25, 30, 35 are summarized in Table 6. Results indicate that the most reliable indices are I_1, I_4 and I_6 with almost significant correlation.

Table 5: Values of the linear correlation coefficients

index n	I_1	I_2	I_3	I_4	I_5	I_6
20	0.53	0.35	0.51	0.53	0.53	0.53
25	0.48	0.49	0.21	0.48	0.48	0.49
30	0.29	0.21	0.32	0.29	0.42	0.33
35	0.56	0.52	0.37	0.56	0.38	0.55
average value	0.47	0.39	0.35	0.47	0.45	0.48

The obtained experimental results indicate the luck of theoretical explanations of phenomena with complexity indices, the need for experiments with instances of higher dimensions and, perhaps, the need for better classification of graph invariants than the intuitive approach.

Summary of previous work

Previous work on complexity indices for the travelling salesman problem, has been summarized in

Cvetković D., Čangalović M., Kovačević-Vujčić V., *Optimization and highly informative graph invariants*, Two Topics in Mathematics, ed. B.Stanković, Zbornik radova 10(18), Matematički institut SANU, Beograd 2004, 5-39.

Some new ideas about complexity indices based on spectral clustering have been given in

Cvetković D., *Complexity indices for the travelling salesman problem and data mining*, Transactions of Combinatorics, **1**(2012), No. 1, 35-43.

Recently we started to check and extend all these ideas on complexity indices by extensive computational experiments using the well known programming package CONCORDE TSP Solver.

Instead of counting relaxation tasks (in branch and bound algorithms), here we have recorded execution times for solving randomly generated instances.

We have computed several invariants for the considered instances and calculated correlation coefficients between the sequence of selected invariant and the sequence of execution times.

In particular, we generated 100 instances on 50 vertices with a uniform distribution of integer edge weights in the interval [1, 100]. We have considered several invariants of short edge subgraphs, such as the number of edges, the number of components, the largest eigenvalue of the adjacency matrix, the algebraic connectivity, and several others.

The best result obtained so far is for the product of the numbers of vertices in the components of the short edge subgraph consisting of edges with weights 1 or 2. the corresponding correlation coefficient was equal to 0.46.

We plan to consider several similar invariants.

In fact, a short edge subgraph in our situation is a random graph of the Erdős-Rényi type. Here, each pair of vertices is connected with an edge with a fixed probability (= 0.02 in the mentioned case). We noticed that in many cases a giant component appears as the Erdős-Rényi theory predicts.

In one of the cases a component with 32 vertices appeared.

Experiments with higher dimensions and some specially constructed invariants of short edge subgraphs are planned.

Thank you for your attention