Exact value for the average optimal cost of the bipartite traveling salesman and two-factor problems in two dimensions

Riccardo Capelli[*]

Dipartimento di Fisica, University of Milan and INFN, via Celoria 16, 20133 Milan, Italy
and Computational Biomedicine Section, Institute of Advanced Simulation IAS-5 and Institute of Neuroscience and Medicine INM-9,
Forschungszentrum Jülich, Wilhelm-Johnen-Straße, 52425 Jülich, Germany

Sergio Caracciolo,[†] Andrea Di Gioacchino,[‡] and Enrico M. Malatesta[§]
Dipartimento di Fisica, University of Milan and INFN, via Celoria 16, 20133 Milan, Italy

We show that the average optimal cost for the traveling salesman problem in two dimensions, which is the archetypal problem in combinatorial optimization, in the bipartite case, is simply related to the average optimal cost of the assignment problem with the same Euclidean, increasing, convex weights. In this way we extend a result already known in one dimension where exact solutions are available. The recently determined average optimal cost for the assignment when the cost function is the square of the distance between the points provides therefore an exact prediction $\overline{E_N} = \frac{1}{\pi} \log N$ for large number of points $2N$. As a by-product of our analysis, also the loop covering problem has the same optimal average cost. We also explain why this result cannot be extended to higher dimensions. We numerically check the exact predictions.

I. INTRODUCTION

The traveling salesman problem (TSP) [1,2] can be formulated in a few words: What is the shortest tour which goes through N given points? But this is well known to be a computationally intractable problem. The number of possible solutions increases exponentially with the number of points and there is not a known algorithm able to find the solution in a time that increases less than exponentially with N.

When the emphasis is shifted from the research of the solution (in the worst case) to the typical properties of the solution in a class of possible instances, the statistical properties of the optimal solutions can be described by a zero-temperature statistical model. This approach has been tremendously fruitful [3–8]. The random model in which the distances between the cities are independent and equally distributed random variables has been deeply studied [9–12]. Much less is known for the Euclidean version of the random problem, where the positions of the points are chosen at random in a finite domain of $\mathbb{R}^d$, so that the distances of the points are now correlated [13–15]. Of course, for large d, the effects of these correlations are smaller and smaller and the methods used to deal with the problem in absence of correlations becomes more and more effective.

In the Euclidean version of the problem, we associate the step in the tour from the ith point with coordinate x_i to the jth

point with coordinate x_j a cost

$$c_p(x_i, x_j) = c_p(x_i - x_j) := \|x_i - x_j\|^p, \qquad (1)$$

with $p \in \mathbb{R}$ and $\|x_i - x_j\|$ the Euclidean distance between the two points. In the bipartite version of the problem the set of $2N$ points is partitioned into two subsets each with N points and steps are allowed only from points in one subset to points in the other subset; in the monopartite version all the points can be reached from any other point. Interestingly enough in $d = 1$, when $p > 1$, that is, when the cost function is convex and increasing, the search for the optimal tour can be exactly solved both in the bipartite [16], as well as in the monopartite [17], version of the problem.

There have been, recently, what we consider three relevant advancements in the field:

(i) For other optimization problems similar to the TSP, the monopartite and bipartite versions have different optimal cost properties. For example, for the matching one-factor and two-factor (or loop-covering) problems, the optimal cost is expected to be a self-averaging quantity whose average scales according to

$$\overline{E_N^{(p,d)}} \sim N^{1-(p/d)} \qquad (2)$$

(see [13] for a proof in the case $p = 1$). On the other hand, in the bipartite version [18–20] it is expected that

$$\overline{E_N^{(p,d)}} \sim \begin{cases} N^{1-(p/2)} & \text{for } d = 1 \\ N^{1-(p/2)}(\log N)^{p/2} & \text{for } d = 2 \\ N^{1-(p/d)} & \text{for } d > 2, \end{cases} \qquad (3)$$

which is a larger average cost with respect to the monopartite case when $d \leqslant 2$. Moreover, in the bipartite case the optimal cost is expected to be not self-averaging.

---

[*]r.capelli@fz-juelich.de
[†]sergio.caracciolo@mi.infn.it
[‡]andrea.digioacchino@unimi.it
[§]enrico.m.malatesta@gmail.com

(ii) In the bipartite case it is always true [21] that the total optimal cost of the TSP $E_{\mathcal{H}}^*$ is larger than the total optimal cost of the two-factor problem $E_{\mathcal{M}_2}^*$, which is larger than twice the total optimal cost of the corresponding matching problem (assignment) $E_{\mathcal{M}_1}^*$:

$$E_{\mathcal{H}}^* \geqslant E_{\mathcal{M}_2}^* \geqslant 2E_{\mathcal{M}_1}^*. \tag{4}$$

In [16] it has been shown that in $d = 1$, in the asymptotic limit of an infinitely large number of points, this bound is saturated; that is, the total optimal cost of the TSP, rescaled with $N^{1-(p/2)}$, is exactly twice the total rescaled optimal cost of the assignment problem, and therefore all three quantities coincide.

(iii) Through connection with the continuum version of the problem, that is, the well-known *transport problem*, it has been possible to compute, exactly, the total optimal cost of the assignment problem for $d = 2$ and $p = 2$, in the asymptotic limit of an infinitely large number of points [19,22–24]:

$$\overline{E_N^{(2,2)}} = \frac{1}{2\pi} \log N. \tag{5}$$

We considered, therefore, the possibility that also in $d = 2$, and $p > 1$, exactly, thanks to the logarithmic violation present in the bipartite case, the asymptotic total cost of the TSP can be exactly twice that of the assignment, which for $p = 2$ is also exactly known. Indeed, this is the case.

This Rapid Communication is organized as follows. In Sec. II we present the model. In Sec. III we present an argument to justify our strategy. In Sec. IV we provide evidence by numerical simulations of how our result is established for large numbers of points. We also examined the case $p = 1$, which is the most largely considered in the literature.

II. THE MODEL

Consider a generic graph $\mathcal{G} = (V, E)$ where V is its set of vertices and $E \subset V^2$ its set of edges. Let $w_e > 0$ be a weight associated to the edge $e \in E$. We shall consider the complete (monopartite) graph $\mathcal{K}_N$, whose vertex set has cardinality N and $E = V^2$ and the complete bipartite graph $\mathcal{K}_{N,N}$, whose vertex set is $V = V_1 \cup V_2$ with V_1 and V_2 disjoint sets of cardinality N, and $E = V_1 \times V_2$.

Let $\mu = (V, E')$ be a spanning subgraph of $\mathcal{G}$, that is, $E' \subset E$. We can define a total cost associated to μ according to

$$E[\mu] = \sum_{e \in \mu} w_e. \tag{6}$$

We shall consider three different classes of spanning subgraphs $\mathcal{M}$. The set $\mathcal{M}_1$ of one-factor (matching), where each vertex belongs to one and only one edge, the set $\mathcal{M}_2$ of two-factor (two-matching or loop covering), where each vertex belongs to two edges, and the set $\mathcal{H}$ of Hamiltonian cycles, that is, two-factor formed by only one cycle (see Fig. 1). The assignment, respectively two-factor, TSP, problems amounts to the search of the subgraph μ^* in $\mathcal{M}_1$, respectively $\mathcal{M}_2$, $\mathcal{H}$, which is optimal, in the sense of minimal total cost

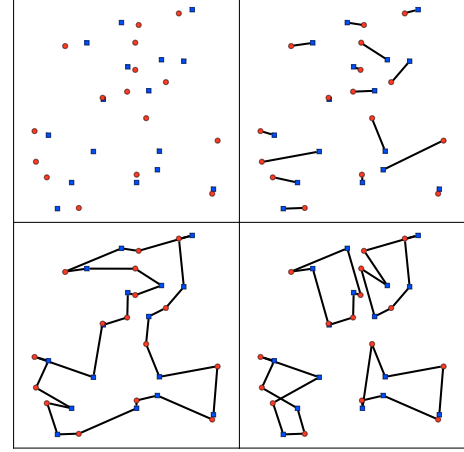$$E_{\mathcal{M}}^* = E_{\mathcal{M}}[\mu^*] = \min_{\mu \in \mathcal{M}} E[\mu] \tag{7}$$

FIG. 1. On the same instance with $N = 16$ blue (squared) and red (disk) points (top left panel), we draw an arbitrarily-chosen example of each class of spanning subgraph we are considering: a one-factor (top right), an Hamiltonian cycle (bottom left) and a two-factor (bottom right).

with $\mathcal{M}$, respectively $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{H}$. In the Euclidean version of our bipartite optimization problems, we consider the immersion of $\mathcal{K}_{N,N}$ in an open subset $\Omega \subset \mathbb{R}^d$. V_1, respectively V_2, will be identified by the set of N points with coordinates r_i, that we shall call the red points (respectively, by the set of N points with coordinates b_j, that we shall call the blue points). Let (i, j) be the edge connecting the ith red vertex with the jth blue vertex. We give a weight $w_{ij} = c_p(r_i - b_j)$, with $p \geqslant 1$ as in Eq. (1).

Of course, in the monopartite Euclidean version there is only one set of points.

In the random Euclidean version of the problem, each possible instance is obtained by choosing at random, with a given law, the position of the points in Ω. For example, we shall consider $\Omega = [0, 1]^d$ and the flat distribution. We denote by $\overline{E^*}$ the average, over all instances, of the optimal total cost E^*.

III. SCALING ARGUMENT

In this section we will provide a scaling argument to support our claim; that is, also in two dimensions, for any given choice of the positions of the points, in the asymptotic limit of large N, the cost of the bipartite TSP converges to twice the cost of the assignment.

Given an instance, let us consider the optimal assignment μ^* on them. Let us now consider N points which are taken between the red and blue points of each edge in μ^* and call $\mathcal{T}^*$ the optimal *monopartite* TSP solution on these points. For simplicity, as these N points we take the blue points.

We shall use $\mathcal{T}^*$ to provide an ordering among the red and blue points. Given two consecutive points in $\mathcal{T}^*$, for example, b_1 and b_2, let us denote by (r_1, b_1) and (r_2, b_2) the two edges in μ^* involving the blue points b_1 and b_2 and let us consider also the new edge (r_1, b_2). We know that, in the asymptotic limit of large N, the typical distance between two matched points in μ^* scales as $(\log N/N)^{1/2}$ while the typical distance between two points matched in the monopartite case scales only as $1/N^{1/2}$, that is (for all points but a fraction which goes
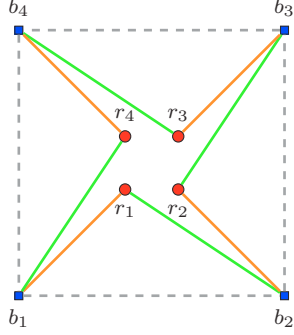
FIG. 2. The optimal assignment μ^* is given by the orange edges $\{(r_1, b_1), (r_2, b_2), (r_3, b_3), (r_4, b_4)\}$. The monopartite TSP (gray dashed edges) among the blue points provides the necessary ordering. In order to obtain the TSP $b_1, r_1, b_2, r_2, b_3, r_3, b_4, r_4, b_1$ in the bipartite graph we have to add the green edges $\{((r_1, b_2), (r_2, b_3), (r_3, b_4), (r_4, b_1)\}$.

to zero with N),

$$w_{(b_1, r_1)} = \left(\alpha_{11} \frac{\log N}{N} \right)^{p/2},$$

$$w_{(b_2, r_1)} = \left[\beta_{22} \frac{1}{N} + \alpha_{11} \frac{\log N}{N} - \gamma \frac{\sqrt{\log N}}{N} \right]^{p/2}, \quad (8)$$

where $(\alpha_{11} \log N/N)^{1/2}$ is the length of the edge (r_1, b_1) of μ^*, $(\beta_{22}/N)^{1/2}$ is the length of the edge (b_1, b_2) of $\mathcal{T}^*$, and $\gamma = 2\sqrt{\alpha_{11}\beta_{22}} \cos\theta$, where θ is the angle between the edges (r_1, b_1) of μ^* and (b_1, b_2) of $\mathcal{T}^*$.

This means that, typically, the difference in cost

$$\Delta E = w_{(b_2, r_1)} - w_{(b_1, r_1)} \sim \frac{(\log N)^{(p-1)/2}}{N^{p/2}} \quad (9)$$

is small as compared to the typical cost $(\log N/N)^{p/2}$ of one edge in the bipartite case. To obtain a valid TSP solution, which we call h^A, we add to the edges $\mu^* = \{(r_1, b_1), \ldots, (r_N, b_N)\}$ the edges $\{(r_1, b_2), \ldots, (r_{N-1}, b_N), (r_N, b_1)\}$ (see Fig. 2).

Of course h^A is not, in general, the optimal solution of the TSP. However, because of Eq. (4), we have that

$$E_{\mathcal{H}}[h^A] \geqslant E_{\mathcal{H}}^* \geqslant E_{\mathcal{M}_2}^* \geqslant 2 E_{\mathcal{M}_1}^*, \quad (10)$$

and we have shown that, for large N, $E_{\mathcal{H}}[h^A]$ goes to $2E_{\mathcal{M}_1}^*$ and therefore also $E_{\mathcal{H}}^*$ must behave in the same way. As a by-product of our analysis, also $E_{\mathcal{M}_1}^*$ for the loop covering problem has the same optimal average cost. Note also that our argument is purely local and therefore it does not depend in any way on the type of boundary conditions adopted. Since in the case of periodic boundary conditions, as shown in [23], it holds (5), we get that the average optimal cost of both the TSP and two-factor goes for large N to two times the optimal assignment.

Notice that an analogous construction can be used in any number of dimensions. However, the success of the procedure lies in the fact that the typical distance between two points in μ^* goes to zero slower than the typical distance between two consecutive points in the monopartite TSP. This is true only in one and two dimensions, and it is related to the importance

of fluctuations in the number of points of different kinds in a small volume.

This approach allowed us to find also an approximated solution of the TSP which improves as $N \to \infty$. However, this approximation requires the solution of a *monopartite* TSP on $N/2$ points, corroborating the fact that the bipartite TSP is a hard problem (from the point of view of complexity theory).

A similar construction can be used to achieve an approximated solution also for the two-factor problem. In this case, instead of solving the monopartite TSP on a point chosen within each edge of μ^*, one should solve the monopartite matching problem on this set of points, obtaining a matching $\mathcal{M}^*$. Once more let us denote by (r_1, b_1) and (r_2, b_2) the two edges in μ^* which give rise to two matched points in $\mathcal{M}^*$, and collect them together with the edges (r_1, b_2) and (r_2, b_1). Repeating the above procedure for each couple of points matched in $\mathcal{M}^*$, the union of the edges obtained gives a valid two-factor whose cost tends, in the limit of large N, to twice the cost of the optimal assignment in one and two dimensions. Notice that, in this case, the procedure is much more efficient because the solution of the matching problem is polynomial in time.

IV. NUMERICAL RESULTS

We have confirmed our theoretical predictions by performing numerical simulations on all three models previously presented: assignment, bipartite two-factor, and bipartite TSP. We have considered the case of open boundary conditions.

For what concerns the assignment problem, many polynomial-time algorithms are available in the literature, such as the famous Hungarian algorithm [25]. We have implemented an in-house assignment solver based on the LEMON optimization library [26], which is based on Edmonds' blossom algorithm [27]. In the case of the two-factor and TSP, the most efficient way to numerically tackle those problems is to exploit their *linear* or *integer programming* formulation.

To validate our argument, we solved for both the assignment and the two-factor problem (with $p = 1, 2$), 10^5 independent instances for $2 \leqslant N \leqslant 125$, 10^4 independent instances for $150 \leqslant N \leqslant 500$, and 10^3 independent instances for $600 \leqslant N \leqslant 1000$. In the TSP case, the computational cost is dramatically larger; for this reason the maximum number of points we were able to achieve with good numerical precision using integer programming was $N = 300$, also reducing the total number of instances.

An estimate of the asymptotic average optimal cost and finite-size corrections has been obtained using the fitting function for $p = 1$

$$f^{(p=1)}(N) = \sqrt{N \log N} \left(a_1 + \frac{a_2}{\log N} + \frac{a_3}{\log^2 N} \right), \quad (11)$$

while, for $p = 2$

$$f^{(p=2)}(N) = \log N \left(a_1 + \frac{a_2}{\log N} + \frac{a_3}{\log^2 N} \right). \quad (12)$$

These are the first three terms of the asymptotic behavior of the cost of the assignment problem [18,19]. Parameters a_2 and a_3 for $p = 2$ were obtained fixing a_1 to $1/\pi$. In Fig. 3 we plot
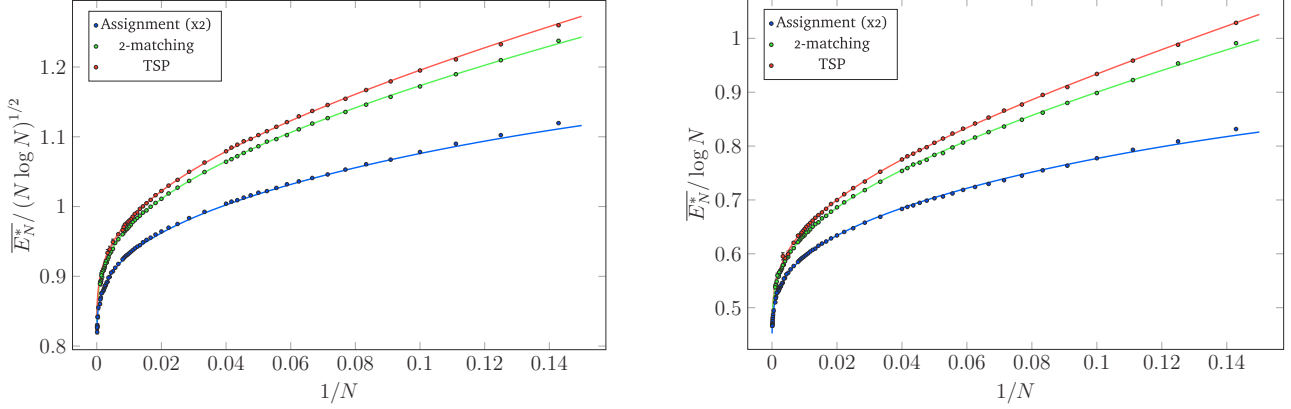
FIG. 3. Numerical results for $p = 1$ (left panel) and $p = 2$ (right panel) for the TSP (red points, top), the two-factor [green points (middle)], and two times the assignment problem [blue points (bottom)] in the open boundary condition case. Continuous lines are numerically fit to the data.

the data and fit in the case of open boundary conditions. The results are reported in Table I.

To better confirm the behavior of the average optimal cost of the TSP, we also performed some numerical simulations using a much more efficient solver, that is, the Concorde TSP solver [28], which is based on an implementation of the branch-and-cut algorithm proposed by Padberg and Rinaldi [29]. The results for the leading term of the asymptotic average optimal cost are confirmed while a small systematic error due to the integer implementation of the solver is observed in the finite-size corrections.

V. CONCLUSIONS

In this work we have considered three combinatorial optimization problems: the matching problem, two-factor problem, and TSP, where the cost is a convex increasing function of the point distances. Previous investigations have been performed in the one-dimensional case, by means of exact solutions [16]. Here we analyzed the bipartite version of these problems in two dimensions, showing that, as already obtained in one dimension,

$$\lim_{N \to \infty} \frac{\overline{E^*_{\mathcal{H}}}}{\overline{E^*_{\mathcal{M}_1}}} = \lim_{N \to \infty} \frac{\overline{E^*_{\mathcal{M}_2}}}{\overline{E^*_{\mathcal{M}_1}}} = 2. \qquad (13)$$

This implies, for the special case $p = 2$, by using (5), our main exact result is $\lim_{N \to \infty}(\overline{E^*_{\mathcal{H}}}/\log N) = 1/\pi$. In general, the evaluation of $\overline{E^*_{\mathcal{H}}}$ and $\overline{E^*_{\mathcal{M}_2}}$ for large N is reduced to the solution of the matching problem which requires only polynomial time. This seems to be a peculiar feature of the bipartite problem: the monopartite TSP cannot be approached in a similar way. As a by-product of our analysis, we provided in Sec. III two approximate algorithms, for the bipartite TSP and the bipartite two-factor: both are guaranteed to give a solution with optimal cost for large N. The first algorithm allows one to solve the bipartite TSP on N points solving the monopartite TSP with N points (notice that, in principle, the bipartite version consists of $2N$ points). The second allows one to exploit the fast Hungarian algorithm to obtain an approximate solution of the two-factor problem.

TABLE I. Comparison between fit factors in assignment and TSP, for $p = 1, 2$. We have doubled the factors for the assignment to verify our hypothesis. For $p = 2$, we have reported the theoretical value of a_1, which is $1/\pi$.

$p = 1$	a_1	a_2	a_3	$p = 2$	a_1	a_2	a_3
TSP	0.717(2)	1.32(1)	$-0.513(1)$	TSP	0.321(5)	1.603(2)	$-0.428(6)$
Two-factor	0.714(2)	1.31(1)	$-0.58(2)$	Two-factor	0.319(4)	1.577(2)	$-0.547(7)$
Assignment	0.714(2)	1.17(2)	$-0.77(2)$	Assignment	0.31831	1.502(2)	$-1.05(1)$

[1] R. M. Karp and M. Steele, *The Travelling Salesman Problem* (Wiley, New York, 1985).

[2] A. M. Frieze and J. E. Yukich, *The Travelling Salesman Problem and its Variations* (Kluwer Academic, Dordrecht, 2002).

[3] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, Science **220**, 671 (1983).

[4] J. Vannimenus and M. Mézard, On the statistical mechanics of optimization problems of the travelling salesman type, J. Phys. Lett. **45**, L1145 (1984).

[5] Y. Fu and P. W. Anderson, Application of statistical mechanics to NP-complete problems in combinatorial optimisation, J. Phys. A **19**, 1605 (1986).

[6] N. Sourlas, Statistical mechanics and the travelling salesman problem, Europhys. Lett. **2**, 919 (1986).

[7] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond: An Introduction to the Replica Method and its Applications* (World Scientific, Singapore, 1987), Vol. 9.

[8] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, New York, 2009).

[9] H. Orland, Mean-field theory for optimization problems, J. Phys. Lett. **46**, 763 (1985).

[10] M. Mézard and G. Parisi, Mean-field equations for the matching and the travelling salesman problems, Europhys. Lett. **2**, 913 (1986).

[11] M. Mézard and G. Parisi, A replica analysis of the travelling salesman problem, J. Phys. **47**, 1285 (1986).

[12] W. Krauth and M. Mézard, The cavity method and the travelling-salesman problem, Europhys. Lett. **8**, 213 (1989).

[13] J. Beardwood, J. H. Halton, and J. M. Hammersley, The shortest path through many points, Proc. Cambridge Philos. Soc. **55**, 299 (1959).

[14] A. G. Percus and O. C. Martin, Finite Size and Dimensional Dependence in the Euclidean Traveling Salesman Problem, Phys. Rev. Lett. **76**, 1188 (1996).

[15] N. J. Cerf, J. H. Boutet de Monvel, O. Bohigas, O. C. Martin, and A. G. Percus, The random link approximation for the Euclidean traveling salesman problem, J. Phys. I **7**, 117 (1997).

[16] S. Caracciolo, A. D. Gioacchino, M. Gherardi, and E. M. Malatesta, Solution for a bipartite Euclidean traveling-salesman problem in one dimension, Phys. Rev. E **97**, 052109 (2018).

[17] S. Caracciolo, A. D. Gioacchino, E. M. Malatesta, and C. Vanoni, Monopartite Euclidean TSP in one dimension (unpublished).

[18] M. Ajtai, J. Komlós, and G. Tusnády, On optimal matchings, Combinatorica **4**, 259 (1984).

[19] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro, Scaling hypothesis for the Euclidean bipartite matching problem, Phys. Rev. E **90**, 012118 (2014).

[20] E. Boniolo, S. Caracciolo, and A. Sportiello, Correlation function for the grid-Poisson Euclidean matching on a line and on a circle, J. Stat. Mech. (2014) P11023.

[21] S. Caracciolo, A. D. Gioacchino, and E. M. Malatesta, Plastic number and possible optimal solutions for an Euclidean 2-matching in one dimension, J. Stat. Mech. (2018) 083402.

[22] S. Caracciolo and G. Sicuro, Scaling hypothesis for the Euclidean bipartite matching problem. II. Correlation functions, Phys. Rev. E **91**, 062125 (2015).

[23] S. Caracciolo and G. Sicuro, Quadratic Stochastic Euclidean Bipartite Matching Problem, Phys. Rev. Lett. **115**, 230601 (2015).

[24] L. Ambrosio, F. Stra, and D. Trevisan, A PDE approach to a 2-dimensional matching problem, Probab. Theory Relat. Fields 1 (2018).

[25] H. W. Kuhn, The Hungarian method for the assignment problem, Nav. Res. Logist. **2**, 83 (1955).

[26] B. Dezső, A. Jüttner, and P. Kovács, LEMON—An open source C++ graph template library, Electron. Notes Theor. Comput. Sci. **264**, 23 (2011).

[27] J. Edmonds, Paths, trees, and flowers, Can. J. Math. **17**, 449 (1965).

[28] D. Applegate, R. Bixby, V. Chvatal, and W. Cook, Concorde TSP solver, (2006), http://www.math.uwaterloo.ca/tsp/index.html.

[29] M. Padberg and G. Rinaldi, A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems, SIAM Rev. **33**, 60 (1991).