

Exact value for the average optimal cost of bipartite traveling-salesman and 2-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

(Dated: September 13, 2018)

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 byproduct of our analysis also the loop covering problem has the same optimal average cost. We also explain why this result cannot be extended at higher dimensions. We numerically check the exact predictions.

INTRODUCTION

The Traveling Salesman Problem (TSP) [1, 2] can be formulated in 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 position 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 to the step in the tour from the i -th point with coordinate x_i to the j -th point with coordinate x_j a cost

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

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 in 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 progresses 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, 1-factor and 2-factor (or loop-covering) problems, the optimal cost is expected to be a self-averaging quantity whose average scales according to

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

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

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

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

- 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 2-factor problem $E_{\mathcal{M}_2}^*$, which is larger than twice the total optimal cost of the corresponding matching problem (assignment) $E_{\mathcal{M}_1}^*$

$$E_{\mathcal{H}}^* \geq E_{\mathcal{M}_2}^* \geq 2E_{\mathcal{M}_1}^*. \quad (4)$$

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

- iii) in the bipartite case, in $d = 2$ and $p = 2$, thanks to a deep 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 in the asymptotic limit of an infinitely large number points [19, 22–24]:

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

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 the one of the assignment, that for $p = 2$ is also exactly known. Indeed, this is the case!

The paper is organized as follows. In Section we present the model. In Section we present an argument to justify our strategy. In Section we provide evidence by numerical simulations how our result is established for large number of points. We also examined the case $p = 1$, which is the most largely considered in the literature.

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

$$E[\mu] = \sum_{e \in \mu} w_e. \quad (6)$$

We shall consider three different classes of spanning subgraphs $\mathcal{M}$. The set $\mathcal{M}_1$ of 1-factor (matching), where each vertex belongs to one and only one edge, the set $\mathcal{M}_2$ of 2-factor (2-matching or loop covering), where each vertex belongs to two edges, and the set $\mathcal{H}$ of Hamiltonian cycles, that is 2-factor formed by only one cycle (see Fig. 1). The assignment, respectively 2-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

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

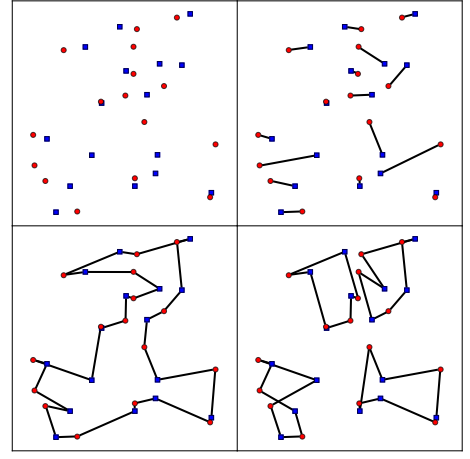


Figure 1: On the same instance with $N = 16$ blue (squared) and red (disk) points (top left panel), we draw an arbitrarily-chosen example of each class of spanning subgraph we are considering: a 1-factor (top right), an Hamiltonian cycle (bottom left) and a 2-factor (bottom right).

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 's, that we shall call the red points, respectively by the set of N points with coordinates b_j 's, that we shall call the blue points. Let (i, j) be the edge connecting the i -th red vertex with the j -th blue vertex. We give a weight $w_{ij} = c_p(r_i - b_j)$, with $p \geq 1$ 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^* .

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 point 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

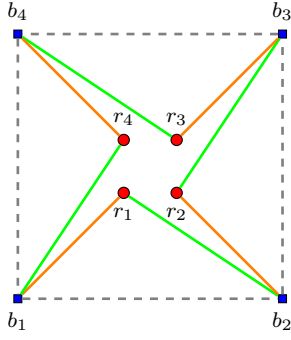


Figure 2: The optimal assignment μ^* is given by the orange edges $\{(r_1, b_1), (r_2, b_2), (r_3, b_3), (r_4, b_4)\}$. The monopartite TSP (gray dashed edges) among blue points provides the necessary ordering. In order to obtain the TSP $b_1, r_1, b_2, r_2, b_3, r_3, b_4, r_4, b_1$ in the bipartite graph we have to add the green edges $\{(r_1, b_2), (r_2, b_3), (r_3, b_4), (r_4, b_1)\}$.

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 to zero with N)

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

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

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

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

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

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

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

and we have shown that, for large N , $E_{\mathcal{H}}[h^A]$ goes to $2 E_{\mathcal{M}_1}^*$ and therefore also $E_{\mathcal{H}}^*$ must behave in the same way. As a byproduct 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 2-factor goes for large N to 2 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 $N \rightarrow \infty$. 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 2-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 2-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.

NUMERICAL RESULTS

We have confirmed our theoretical predictions performing numerical simulations on all the three models previously presented: assignment, bipartite 2-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, 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 the Edmonds' blossom algorithm [27]. In the case of the 2-factor and TSP, the most efficient way to tackle numerically those problems is to exploit their *linear* or *integer programming* formulation.

To validate our argument, we solved for both assignment and 2-factor problem (with $p = 1, 2$), 10^5 independent instances for $2 \leq N \leq 125$, 10^4 independent instances for $150 \leq N \leq 500$, and 10^3 independent instances for $600 \leq N \leq 1000$. In the TSP case, the computational cost is dramatically larger; for this reason the maximum number of points we were able to achieve with a 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

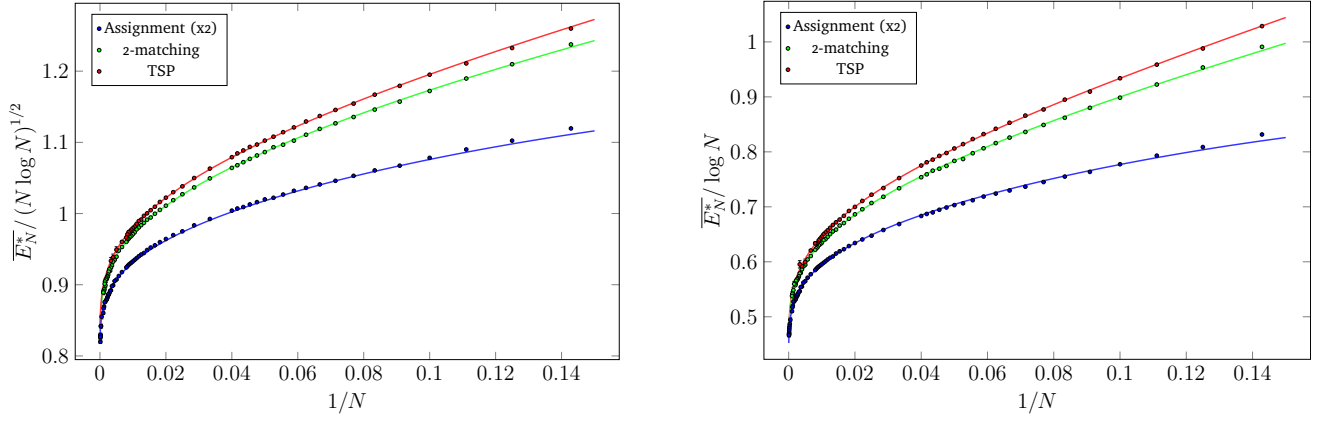


Figure 3: Numerical results for $p = 1$ (left panel) and $p = 2$ (right panel) for the TSP (red points, top), the 2-factor (green points, middle) and 2 times the assignment problem (blue points, bottom) in the open boundary condition case. Continuous lines are numerical fit to the data.

$p = 1$	a_1	a_2	a_3
TSP	0.717(2)	1.32(1)	-0.513(1)
2-factor	0.714(2)	1.31(1)	-0.58(2)
Assignment	0.714(2)	1.17(2)	-0.77(2)

$p = 2$	a_1	a_2	a_3
TSP	0.321(5)	1.603(2)	-0.428(6)
2-factor	0.319(4)	1.577(2)	-0.547(7)
Assignment	0.31831	1.502(2)	-1.05(1)

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 $1/\pi$.

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 3 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 Figure 3 we plot the data and fit in the case of open boundary conditions. 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.

CONCLUSIONS

In this work we have considered three combinatorial optimization problems, the matching problem, 2-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:

$$\lim_{N \rightarrow \infty} \frac{\overline{E_{\mathcal{H}}^*}}{\overline{E_{\mathcal{M}_1}^*}} = \lim_{N \rightarrow \infty} \frac{\overline{E_{\mathcal{M}_2}^*}}{\overline{E_{\mathcal{M}_1}^*}} = 2. \quad (13)$$

This implies, for the special case $p = 2$, by using (5), our main exact result, that is $\lim_{N \rightarrow \infty} (\overline{E_{\mathcal{H}}^*} / \log N) = 1/\pi$. 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 byproduct of our analysis, we provided in Sec. two approximate algorithms, for the bipartite TSP and the bipartite 2-factor: both are guaranteed to give a solution with optimal cost for large N . The first algorithm allows to solve the bipartite TSP on N points solving the monopartite TSP with N points (notice that, on principle, the bipartite version consists of $2N$ points). The second allows to exploit the fast Hungarian algorithm to obtain an approximate solution of the 2-factor problem.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 785907 (HBP SGA2). We thank Gabriele Sicuro for sharing with us his numerical results on the matching.

* r.capelli@fz-juelich.de

† sergio.caracciolo@mi.infn.it

‡ andrea.digioacchino@unimi.it

§ enrico.m.malatesta@gmail.com

- [1] R. M. Karp and M. Steele. *The Travelling Salesman Problem*. John Wiley and Sons, New York, 1985.
- [2] A. M. Frieze and J. E. Yukich. *The Travelling Salesman Problem and its Variations*. Kluwer Academic Publishers B.V., 2002.
- [3] Scott Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983.
- [4] Jean Vannimenus and Marc Mézard. On the statistical mechanics of optimization problems of the travelling salesman type. *Journal de Physique Lettres*, 45(24):L1145–L1153, 1984.
- [5] Yaotian Fu and Philip W. Anderson. Application of statistical mechanics to np-complete problems in combinatorial optimisation. *Journal of Physics A: Mathematical and General*, 19(9):1605, 1986.
- [6] Nicolas Sourlas. Statistical Mechanics and the Travelling Salesman Problem. *Europhysics Letters*, 2(12):919–923, 1986.
- [7] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.
- [8] Marc Mézard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.
- [9] Henri Orland. Mean-field theory for optimization problems. *Le Journal de Physique - Lettres*, 46(17), 1985.
- [10] Marc Mézard and Giorgio Parisi. Mean-field equations for the matching and the travelling salesman problems. *Europhysics Letters*, 2(12):913–918, 1986.
- [11] Marc Mézard and Giorgio Parisi. A replica analysis of the travelling salesman problem. *Journal de Physique*, 47(1986):1285–1296, 1986.
- [12] Werner Krauth and Marc Mézard. The Cavity Method and the Travelling-Salesman Problem. *Europhysics Letters*, 8(3):213–218, 1989.
- [13] Jillian Beardwood, John H Halton, and John M. Hammersley. The shortest path through many points. *Proc. Cambridge Philos. Soc.*, 55:299–327, 1959.
- [14] Allon G. Percus and Olivier C. Martin. Finite size and dimensional dependence in the Euclidean traveling salesman problem. *Physical Review Letters*, 76(8):1188–1191, feb 1996.
- [15] Nicolas J. Cerf, Jacques H. Boutet de Monvel, Oriol Bohigas, Olivier C. Martin, and Allon G. Percus. The random link approximation for the Euclidean traveling salesman problem. *Journal de Physique I*, 7(January):117–136, 1997.
- [16] Sergio Caracciolo, Andrea Di Gioacchino, Marco Gherardi, and Enrico M. Malatesta. Solution for a bipartite euclidean traveling-salesman problem in one dimension. *Phys. Rev. E*, 97:052109, May 2018.
- [17] Sergio Caracciolo, Andrea Di Gioacchino, Enrico M. Malatesta, and Carlo Vanoni. Monopartite euclidean tsp in one dimension. 2018.
- [18] Miklós Ajtai, János Komlós, and Gabor Tusnády. On optimal Matchings. *Combinatorica*, 4(4):259–264, 1984.
- [19] Sergio Caracciolo, Carlo Lucibello, Giorgio Parisi, and Gabriele Sicuro. Scaling hypothesis for the euclidean bipartite matching problem. *Phys. Rev. E*, 90:012118, 2014.
- [20] Elena Boniolo, Sergio Caracciolo, and Andrea Sportiello. Correlation function for the grid-poisson euclidean matching on a line and on a circle. *J. Stat. Mech.*, 11:P11023, 2014.
- [21] Sergio Caracciolo, Andrea Di Gioacchino, and Enrico M. Malatesta. Plastic number and possible optimal solutions for an euclidean 2-matching in one dimension. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(8):083402, 2018.
- [22] Sergio Caracciolo and Gabriele Sicuro. Scaling hypothesis for the Euclidean bipartite matching problem. II. Correlation functions. *Phys. Rev. E*, 91(062125), 2015.
- [23] Sergio Caracciolo and Gabriele Sicuro. Quadratic stochastic euclidean bipartite matching problem. *Phys. Rev. Lett.*, 115(23):230601, 2015.
- [24] Luigi Ambrosio, Federico Stra, and Dario Trevisan. A PDE approach to a 2-dimensional matching problem. *Probab. Theory Relat. Fields*, 2018.
- [25] Harold W Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics (NRL)*, 2(1-2):83–97, 1955.
- [26] Balázs Dezső, Alpár Jüttner, and Péter Kovács. LEMON – an open source C++ graph template library. *Electronic Notes in Theoretical Computer Science*, 264(5):23–45, 2011.
- [27] Jack Edmonds. Paths, trees, and flowers. *Canadian Journal of mathematics*, 17(3):449–467, 1965.
- [28] David Applegate, Ribert Bixby, Vasek Chvatal, and William Cook. *Concorde TSP solver*, 2006.
- [29] Manfred Padberg and Giovanni Rinaldi. A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems. *SIAM review*, 33(1):60–100, 1991.