Article

A General Statistical Physics Framework for Assignment Problems

Patrice Koehl [1,*] and Henri Orland [2]

1   Department of Computer Science, University of California, Davis, CA 95616, USA
2   CNRS, CEA, Institut de Physique Théorique, Université Paris-Saclay, 91191 Gif-sur-Yvette, France; henri.orland@cea.fr
*   Correspondence: koehl@cs.ucdavis.edu

Abstract: Linear assignment problems hold a pivotal role in combinatorial optimization, offering a broad spectrum of applications within the field of data sciences. They consist of assigning "agents" to "tasks" in a way that leads to a minimum total cost associated with the assignment. The assignment is balanced when the number of agents equals the number of tasks, with a one-to-one correspondence between agents and tasks, and it is and unbalanced otherwise. Additional options and constraints may be imposed, such as allowing agents to perform multiple tasks or allowing tasks to be performed by multiple agents. In this paper, we propose a novel framework that can solve all these assignment problems employing methodologies derived from the field of statistical physics. We describe this formalism in detail and validate all its assertions. A major part of this framework is the definition of a concave effective free energy function that encapsulates the constraints of the assignment problem within a finite temperature context. We demonstrate that this free energy monotonically decreases as a function of a parameter β representing the inverse of temperature. As β increases, the free energy converges to the optimal assignment cost. Furthermore, we demonstrate that when β values are sufficiently large, the exact solution to the assignment problem can be derived by rounding off the elements of the computed assignment matrix to the nearest integer. We describe a computer implementation of our framework and illustrate its application to multi-task assignment problems for which the Hungarian algorithm is not applicable.

Keywords: assignment problems; statistical physics; discrete optimization

1. Introduction

An assignment problem can be interpreted as a problem of resource allocation in which a certain number of "tasks" are performed by some "agents" so that the total cost of pairing tasks with agents is minimized. There are many ways to constrain this problem, such as imposing one-to-one mapping (the balanced problem), searching for a specific number of assignments (the *k*-cardinality assignment problem), imposing constraints on the resources (the generalized assignment problem), or allowing agents to perform multiple tasks or reversely having some tasks performed by more than one agent. For reviews on the variety of assignment problems and the algorithms that have been proposed to solve them, we refer the reader to Refs. [1–3] and references therein. Assignment problems are fundamental combinatorial optimization problems. As such, solving such problems is a significant concern in operational research, economics, and data sciences, among other disciplines. Assignment problems have been and continue to be a focal point of investigation for mathematicians, statisticians, computer scientists, and even physicists. In this paper, we propose a general framework that regroups all the assignment problems mentioned above. We develop an approximate solution to this problem based on statistical physics and prove that it converges efficiently to an exact solution for non-degenerate as well as degenerate assignment problems.

Let us first define the assignment problem. Let T be the set of tasks and A the set of agents, with possibly different cardinalities $|T| = N_2$ and $|A| = N_1$. If we define $C(i,j)$ as the cost of assigning task i to the agent j, the balanced assignment problem can be stated as finding a transportation matrix G between T and A whose elements are non-negative integer numbers and that minimizes

$$U = \sum_{i \in T} \sum_{j \in A} C(i,j) G(i,j). \tag{1}$$

In the special case of $N_1 = N_2$, with a one-to-one assignment of tasks to agents (the *balanced assignment problem*), G corresponds to a permutation of $\{1, \ldots, N\}$. In the more general case, N_1 and N_2 can be different (unbalanced assignment), the number of pairings k may be imposed with $0 < k \le \min(N_1, N_2)$ (*k*-cardinality assignment problem), and/or tasks and agents may be assigned to more than one agents and tasks, respectively. In those general cases, G remains binary but may not be a permutation matrix (in the specific case of the *k*-cardinality problem, for example, G is a partial permutation matrix of rank k).

As mentioned above, assignment problems are basically combinatorial optimization problems that can be solved numerically using a linear programming model or, more specifically, a linear integer programming method. Those methods, however, are theoretically NP-complete. There are fast heuristics that can solve such problems, such as Dantzig's simplex method [4]. The assignment problem, however, is a specialized case of integer linear programming that can be recast and solved by an exact polynomial algorithm. For example, the most common algorithm for solving the balanced assignment problem has its origin in the work of Jacobi [5]. It was "rediscovered" 60 years later by [6] and is now dubbed the Hungarian algorithm. It is a global algorithm that iteratively identifies assignments between agents and tasks. It is polynomial in running time ($O(N^4)$ or $O(N^3)$ depending on the implementation). While it is serial by nature, there are many recent efforts to parallelize this algorithm (see, for example [7–9], and references therein).

The Hungarian algorithm can be adapted to many unbalanced assignment problems. If there are N_2 tasks and N_1 agents, with $N_2 > N_1$, for example, it is possible to add $N_2 - N_1$ "dummy" agents, solve the balanced assignment problem of size N_2, and then only retain the assignments to the actual N_1 agents. If each agent can perform up to M tasks, the textbook solution is then to create M copies of each agent prior to using the Hungarian algorithm. This approach, however, may not lead to the expected solution. Consider a problem with fiv tasks and three agents, with each agent allowed to perform up to three tasks. The option of creating three copies of each agent and then solving the corresponding unbalanced assignment problem may lead to one agent performing two tasks, a second agent performing three tasks, and the last agent not performing any task, possibly violating a constraint that each agent needs to perform at least one task. Several methods have been proposed to circumvent this problem, either by dividing the problem into multiple subproblems [10,11] or by using random algorithms such as the ant colony algorithms [12,13]. It is unclear whether these algorithms can identify the optimal solution (they have been shown to fail in some cases, see for example [14]) or how they scale for large problems.

In this paper, we propose a radically different approach to solving a general class of assignment problems using continuous systems. Our approach is motivated by statistical physics. It is a full generalization of the method developed in a previous paper to solve balanced assignment problems [15]. In the following, we will use the generic term "multi-assignment" to represent an assignment problem with agents possibly performing multiple tasks, or with tasks assigned to multiple agents. In this paper, we aim to

- Introduce and substantiate a continuous framework for addressing potentially multifaceted assignment problems involving multiple tasks and/or multiple agents, leveraging principles from statistical physics;
- Demonstrate that, under generic circumstances where the multi-assignment problem possesses a unique solution, the aforementioned framework ensures convergence to

that solution with arbitrarily high accuracy, in terms of both cost (referred to as energy) and assignment matrix,

- Present a modified approach capable of identifying at least one solution for degenerate multi-assignment problems that feature multiple solutions;
- Demonstrate that the implementation of this framework can be efficiently parallelized.

We emphasize that this formalism is not a mere adaptation but a full generalization of the framework we develop for solving assignment problems [16]. In particular, new features of this paper are as follows:

- A method to account for the fact that each task can be assigned to a variable number of agents and, on the other hand, that each agent can be assigned a variable number of tasks. To that end, we introduce indicator functions over the sets of tasks and agents that are optimized along with the transportation plan.
- The establishment of proofs of validity and convergence of our algorithm for those general multi-assignment problems. The main results are provided in the text, while the proofs themselves are relegated to the Appendices for more clarity. For the balanced assignment problem, these proofs rely heavily on the fact that the corresponding transportation matrices are permutation matrices that are the extreme points of the well-characterized convex set of doubly stochastic matrices (according to the Birkhoff–von Neumann theorem [17,18]). For more general multi-assignment problems, the transportation matrices also belong to a convex set, with properties akin to the Birkhoff–von Neumann theorem (these properties are discussed in Appendix A). The use of these properties to derive the convergence of our algorithm for solving multi-assignment problems is new.

The paper is organized as follows. In sections 2 to 4, we describe in detail the general framework we propose for solving multi-assignment problems. Proofs of all important properties of this framework are provided in the appendices. Section 5 briefly describes the implementation of the method in a C++ program, Matching V1.0. In Section 6, we present some applications, along with a comparison to current algorithms.

2. A General Formulation of the Assignment Problem

We consider two sets of points, S_1 and S_2, with cardinalities N_1 and N_2, respectively. We represent the cost of transportation between S_1 and S_2 as a matrix C whose elements $C(i, j)$ are positive, for all $(i, j) \in [1, N_1] \times [1, N_2]$. We set the number of assignments between points in S_1 and S_2 to be a strictly positive integer number k, a constant given as input to the problem. A point i is assigned to $n_1(i)$ points in S_2, with $n_1(i)$ belonging to $[p_{min}(i), \ldots, p_{max}(i)]$, where $p_{min}(i)$ and $p_{max}(i)$ are non-negative integers satisfying $0 \le p_{min}(i) \le p_{max}(i) \le N_2$. Similarly, a point j is assigned to $n_2(j)$ points in S_1, with $n_2(j)$ belonging to $[q_{min}(j), \ldots, q_{max}(j)]$, where $q_{min}(j)$ and $q_{max}(j)$ are non-negative integers integers satisfying $0 \le q_{min}(j) \le q_{max}(j) \le N_1$. Using the traditional "task-agent" for the formulation of an assignment problem, S_1 represents the tasks, and S_2 the agents. $n_1(i)$ is the number of agents needed to perform task i, while $n_2(j)$ is the number of tasks that can be performed by agent j. $p_{min}(i)$ and $p_{max}(i)$ represent the minimum and maximum number of agents that are assigned to a task i, respectively, while $q_{min}(j)$ and $q_{max}(j)$ represent the minimum and maximum number of tasks that can be assigned to an agent j, respectively. We use a general definition of these boundaries, in that we allow each task and each agent to have their own limits. The multi-assignment problem is then framed as the search for an assignment matrix G that defines the correspondence between points in S_1 and points in S_2. This matrix is found by minimizing the assignment cost U, defined as

$$U(G, C) = \sum_{i,j} G(i, j) C(i, j). \tag{2}$$

The summations encompass all i in S_1 and j in S_2. The objective is to locate the minimum of U given the values of $G(i, j)$ that adhere to the following constraints:

$$\forall i, \quad \sum_j G(i,j) = n_1(i)$$

$$\forall j, \quad \sum_i G(i,j) = n_2(j)$$

$$\sum_i \sum_j G(i,j) = k$$

$$\forall (i,j), \quad G(i,j) \in \{0,1\} \tag{3}$$

$$\forall (i), \quad n_1(i) \in [p_{min}(i), \dots, p_{max}(i)]$$

$$\forall (j), \quad n_2(j) \in [q_{min}(j), \dots, q_{max}(j)],$$

where k is the actual number of task–agent pairs. The numbers $p_{min}(i)$, $p_{max}(i)$, $q_{min}(j)$, $q_{max}(j)$, and k are problem-specific and given. They do satisfy some constraints, such as $0 \le p_{min}(i) \le p_{max}(i) \le N_2$ and $0 \le q_{min}(j) \le q_{max}(j) \le N_1$, as described above, as well as $0 < k \le \sum_{i=1}^{N_1} p_{min}(i)$ and $k \le \sum_{j=1}^{N_2} q_{min}(j)$. We store those numbers in four vectors, $P_{min} = (p_{min}(1), \dots, p_{min}(N_1))$, $P_{max} = (p_{max}(1), \dots, p_{max}(N_1))$, $Q_{min} = (q_{min}(1), \dots, q_{min}(N_2))$, $Q_{max} = (q_{max}(1), \dots, q_{max}(N_2))$.

Equation (3) recap most of the standard linear assignment problems:

(i)   If $k = N_1 = N_2$ and $p_{min}(i) = p_{max}(i) = q_{min}(j) = q_{max}(j) = 1$ for all i and j, we recover the balanced assignment problem,

(ii) If $p_{min}(i) = q_{min}(j) = 0$ and $p_{max}(i) = q_{max}(j) = 1$ for all i and j, the equations correspond to the k-cardinality assignment problem [19–21].

In Equation (3), $G(i,j)$, $n_1(i)$ and $n_2(j)$ are unknown, defining a total of $N_1 N_2 + N_1 + N_2$ variables. The solution to the general assignment problem is given by the functions n_1^* and n_2^* on S_1 and S_2, respectively, which identify which tasks and which agents are involved in the optimal assignment, the transportation matrix G^* that defines these correspondences, and the minimum assignment cost $U^* = U(G^*, C)$.

Optimizing Equation (2) subject to the constraints Equation (3) constitutes a discrete optimization challenge, specifically an integer linear programming problem. To address this challenge, we employ a statistical physics methodology, transforming it into a temperature-dependent problem involving real variables. The integer optimal solution is then obtained as the temperature approaches zero.

2.1. Effective Free Energy for the General Assignment Problem

Solving the multi-assignment problem entails determining the minimum of the function defined in Equation (2) across the potential mappings between the two discrete sets of points under consideration. Identifying this minimum is tantamount to identifying the most probable state of the system it characterizes. This "system" encompasses the various binary transportation plans, also referred to as assignment matrices, between S_1 and S_2 that adhere to the constraints Equation (3). Each state within this system corresponds to assignment matrix G with its associated energy $U(G, C)$ as defined in Equation (2). The probability $P(G)$ linked with an assignment matrix G is given by

$$P(G) = \frac{1}{Z(\beta)} e^{-\beta U(G,C)}. \tag{4}$$

In this equation, β is the inverse temperature, namely $\beta = 1/k_B T$, where k_B is the Boltzmann constant and T is the temperature, and $Z(\beta)$ is the partition function computed over all states of the system. As this system is defined by the assignment matrices G and by the marginals of G, n_1, and n_2, the partition function is equal to

$$
\begin{aligned}
Z_\beta &= e^{-\beta \mathcal{F}_\beta} \\
&= \sum_{G(i,j)=0}^{1} \sum_{n_1(i)=p_{min}(i)}^{p_{max}(i)} \sum_{n_2(j)=q_{min}(j)}^{q_{max}(j)} e^{-\beta U(G,C)},
\end{aligned} \tag{5}
$$

where $\mathcal{F}(\beta)$ is the free energy of the system. The practicality of this free energy is limited due to the inability to compute it explicitly. We develop a method for approximating it through the saddle point approximation technique.

Taking into account the constraints in Equation (3), the partition function can be written as

$$Z_\beta = \sum_{G(i,j)=0}^{1} \sum_{n_1(i)=p_{min}(i)}^{p_{max}(i)} \sum_{n_2(j)=q_{min}(j)}^{q_{max}(j)} e^{-\beta \sum_{i,j} C(i,j)G(i,j)}$$

$$\prod_i \delta\left(\sum_j G(i,j) - n_1(i)\right) \prod_j \delta\left(\sum_i G(i,j) - n_2(j)\right)$$

$$\delta\left(\sum_{i,j} G(i,j) - k\right). \tag{6}$$

The three sums impose that $G(i,j)$, $n_1(i)$, and $n_2(j)$ take integer values within some ranges defined by the constraints Equation (3). The additional constraints are imposed through the delta functions. We employ the Fourier representation of those delta functions, thereby introducing additional auxiliary variables x, $\lambda(i)$, and $\mu(j)$, with $i \in [1, N_1]$ and $j \in [1, N_2]$. The partition function is then given, after reorganization, by (up to a multiplicative constant) by

$$Z_\beta = \int_{-\infty}^{+\infty} \prod_i d\lambda(i) \int_{-\infty}^{+\infty} \prod_j d\mu(j) \int_{-\infty}^{+\infty} dx\, e^{\text{ß}\beta k x}$$

$$\sum_{n_1(i)=p_{min}(i)}^{p_{max}(i)} \sum_{n_2(j)=q_{min}(j)}^{q_{max}(j)} e^{\beta\left(\sum_i \text{ß}\lambda(i)n_1(i) + \sum_j \text{ß}\mu(j)n_2(j)\right)}$$

$$\sum_{G(i,j)=0}^{1} e^{-\beta \sum_{i,j} G(i,j)(C(i,j)+\text{ß}\lambda(i)+\text{ß}\mu(j)+\text{ß}x)}, \tag{7}$$

where ß is the imaginary unit ($\text{ß}^2 = -1$). Note that we have rescaled the variables x, λ, and μ by a factor β for consistency with the energy term. Conducting the summations over the variables $G(i,j)$, $n_1(i)$, and $n_2(j)$, we get

$$Z_\beta = \int_{-\infty}^{+\infty} \prod_i d\lambda(i) \int_{-\infty}^{+\infty} \prod_j d\mu_j \int_{-\infty}^{+\infty} dx\, e^{-\beta F_\beta(\lambda,\mu,x)},$$

where $F_\beta(\lambda,\mu,x)$ is a functional, or effective free energy, that depends on the variables λ, μ, and x and is defined by

$$F_\beta(\lambda,\mu,x) =$$

$$-\frac{1}{\beta}\sum_i \ln\left(\frac{e^{\text{ß}\beta p_{min}(i)\lambda(i)} - e^{\text{ß}\beta(p_{max}(i)+1)\lambda(i)}}{1 - e^{\text{ß}\beta\lambda(i)}}\right)$$

$$-\frac{1}{\beta}\sum_j \ln\left(\frac{e^{\text{ß}\beta q_{min}(j)\mu(j)} - e^{\text{ß}\beta(q_{max}(j)+1)\mu(j)}}{1 - e^{\text{ß}\beta\mu(j)}}\right)$$

$$-\frac{1}{\beta}\sum_{i,j} \ln\left[1 + e^{-\beta(C(i,j)+\text{ß}\lambda(i)+\text{ß}\mu(l)+\text{ß}x)}\right] - \text{ß}kx. \tag{8}$$

In contrast to the expression of the internal energy U defined in Equation (2) that depends on the $N_1 N_2 + N_1 + N_2$ constrained binary variables $G(i,j)$, $n_1(i)$, and $n_2(j)$, the expression for the effective free energy $F_\beta(\lambda,\mu,x)$ depends only on $N_1 + N_2 + 1$ uncon-

strained variables, namely $\lambda(i)$, $\mu(j)$, and x. Below, we demonstrate how identifying the extremum of this function enables us to address the multi-assignment problem.

2.2. Optimizing the Effective Free Energy

Let $\overline{G}(i,j)$, $\overline{n_1}(i)$, and $\overline{n_2}(j)$ represent the expected values of $G(i,j)$, $n_1(i)$, and $n_2(j)$, respectively, in accordance with the Gibbs distribution specified in Equation (4). Computing these expected values directly is unfortunately not feasible as the partition function defined in Equation (8) lacks an analytical expression. Instead, we derive a saddle point approximation (SPA) by seeking extrema of the effective free energy with respect to the variables λ, μ, and x:

$$\frac{\partial F_\beta(\lambda,\mu,x)}{\partial \lambda_i} = \frac{\partial F_\beta(\lambda,\mu,x)}{\partial \mu_j} = \frac{\partial F_\beta(\lambda,\mu,x)}{\partial x} = 0. \tag{9}$$

After some rearrangements, these two equations can be written as

$$\forall i, \ \sum_j X(i,j) = d_1(i), \tag{10a}$$

$$\forall j, \ \sum_i X(i,j) = d_2(j), \tag{10b}$$

$$\sum_{i,j} X(i,j) = k, \tag{10c}$$

where

$$
\begin{aligned}
X(i,j) &= h[\beta(C(k,l) + \ss\lambda(i) + \ss\mu(j) + \ss x),] \\
d_1(i) &= g(\ss\beta\lambda(i), p_{min}(i), p_{max}(i)), \\
d_2(j) &= g(\ss\beta\mu(j), q_{min}(j), q_{max}(j)),
\end{aligned} \tag{11}
$$

and

$$
\begin{aligned}
h(x) &= \frac{1}{e^x + 1}, \\
g(x,a,b) &= \frac{(b - a + 1)}{e^{(b-a+1)x} - 1} + \frac{1}{1 - e^x} + b.
\end{aligned}
$$

Note that

(i) $g(x,a,a) = a$. If $a = 1$ and $N_1 = N_2 = 1$, we recover Equation (11) from [16] corresponding to the balanced assignment problem.

(ii) $g(x,0,1) = h(-x)$. In this case, Equation (11) refers to the *k*-cardinality problem.

As frequently encountered, the saddle point may be purely imaginary. In this instance, it is evident from Equations (10) and (11) that the variables $\ss\lambda(i)$, $\ss\mu(j)$, and $\ss x$ must be real. Consequently, we will replace $\{\ss\lambda(i), \ss\mu(j), \ss x\}$ by $\{\lambda(i), \mu(j), x\}$ below.

In order to analyze the saddle point approximation, SPA, it is necessary to verify the existence and evaluate the uniqueness of the extrema of the effective free energy. The following theorem proves that $F_\beta(\lambda,\mu,x)$ is strictly concave.

Theorem 1. *The Hessian of the effective free energy $F_\beta(\lambda,\mu,x)$ is negative definite if $p_{min}(i) < p_{max}(i)$ for all i and $q_{min}(j) < q_{max}(j)$, for all j, and negative semidefinite otherwise. The free energy function $F_\beta(\lambda,\mu,x)$ is then strictly concave in the first case and concave in the second case.*

Proof. See Appendix B. □

The following property establishes a relationship between the solutions of the SPA system of equations and the expected values for the assignment matrix and for the indicator functions:

Proposition 1. *Let $\overline{S}_\beta$ be the expected state of the system at the temperature β with respect to the Gibbs distribution given in Equation (4). $\overline{S}_\beta$ is associated with an expected transportation plan $\overline{G}_\beta$ and expected indicator functions $\overline{n}_{1\beta}$ and $\overline{n}_{2\beta}$. Let $\lambda^{MF}(i)$, $\mu^{MF}(j)$, and x^{MF} be the solutions of the system of Equation (10). Then the following identities hold:*

$$
\begin{aligned}
\overline{G}_\beta(i,j) &= h(\beta(C(i,j) + \lambda^{MF}(i) + \mu^{MF}(j) + x^{MF})) \\
&= X^{MF}(k,l) \\
\overline{n}_{1\beta}(i) &= g(\beta\lambda^{MF}(i), p_{min}(i), p_{max}(i)) = d_1^{MF}(i) \\
\overline{n}_{2\beta}(j) &= g(\beta\mu^{MF}(j), q_{min}(j), q_{max}(j)) = d_2^{MF}(j).
\end{aligned}
\tag{12}
$$

To indicate that the solutions are mean field solutions, we use the superscript MF.

Proof. The proof of this proposition is virtually identical to the proof of the same results for the assignment problem, found in Appendix B of [15]. $\square$

For a given value of the parameter β, $\overline{n}_{1\beta}(i)$ and $\overline{n}_{2\beta}(j)$ are the numbers of elements of S_2 and the number of elements in S_1 that are in correspondence with i and j, respectively, and $\overline{G}_\beta$ forms a transportation plan between S_1 and S_2 that is optimal with respect to the free energy defined in Equation (8). Note that these values are mean values and fractional. The matrix $\overline{G}_\beta$ belongs to a polytope, which we define as $\mathcal{U}^k(P_{min}^{max}, Q_{min}^{max})$ (see Appendix A for a full characterization of this polytope). We can link the mean field assignment matrix $\overline{G}_\beta$ to an optimal free energy F_β^{MF} and to an optimal internal energy $U_\beta^{MF} = \sum_{i,j} \overline{G}_\beta(i,j)C(i,j)$. These values serve as mean field approximations of the exact free energy and internal energy of the system, respectively. Here are the key properties of U_β^{MF} and F_β^{MF}:

Proposition 2. *F_β^{MF} and U_β^{MF} are monotonic increasing and monotonic decreasing functions, respectively, of the parameter β.*

Proof. See Appendix C for F_β^{MF} and Appendix D for U_β^{MF}. $\square$

Theorem 1 and Proposition 2 outlined above underscore several advantages of the proposed framework, which reframes the multi-assignment problem as a temperature-dependent process. Firstly, at each temperature, the multi-assignment problem with a generic cost matrix is transformed into a concave problem with a unique solution. This problem exhibits linear complexity in the number of variables, contrasting with the quadratic complexity of the original problem. The concavity facilitates the utilization of straightforward algorithms for identifying a minimum of the effective free energy function (Equation (8)). Additionally, Equation (12) offers robust numerical stability for computing the transportation plan and the functions n_1 and n_2, owing to the characteristics of the functions $h(x)$ and $g(x)$. Lastly, the convergence with respect to temperature is monotonic.

3. Solving Generic Assignment Problems

In the previous section, we proposed an effective free energy, $F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu}, x)$, that depends on $N_1 + N_2 + 1$ unconstrained variables and on the inverse of the temperature, β, whose extremum identifies the solution of a possibly multi-job, multi-agent assignment problem. We have shown that the trajectory of this extremum is monotonic, increasing with respect to β. We now relate these values to the optimal assignment energy U^*:

Theorem 2. *Let F_β^{MF} and U_β^{MF} be the mean field approximations of the exact free energy and internal energy of the system at the inverse temperature β. The optimal assignment energy U^* is then given by,*

$$U^* = \lim_{\beta \to +\infty} F_\beta^{MF},$$

$$U^* = \lim_{\beta \to +\infty} U_\beta^{MF}, \tag{13}$$

namely, the optimal assignment energy, U^*, is equal to the infinite limit with respect to the inverse temperature of both the mean field free energy and the internal energy.

Proof. See Appendix E. □

This theorem illustrates that as the inverse temperature approaches infinity (or equivalently, as the temperature tends towards zero), the internal energy and free energy of the system converge to the optimal assignment energy. This confirms the validity of our statistical physics approach, particularly emphasizing the effectiveness of the saddle-point approximation. Note, however, that it does not define the behavior of the coupling matrix $\overline{G}_\beta = X^{MF}$. As $\overline{G}_\beta(i,j) = h(\beta(C(i,j) + \lambda^{MF}(i) + \mu^{MF}(j) + x^{MF}))$ and $0 < h(x) < 1$, the coupling matrix at any finite temperature is fractional. We need to show that as $\beta \to +\infty$, the corresponding matrix $\overline{G}_\infty$ does converge to the solution matrix G^* and not to a fractional matrix that would lead to the same low energy U^*.

We first establish bounds on the internal energy and free energy at the SPA. Let us define

$$A(N_1, N_2) = N_1 N_2 \ln(2) + \sum_{i=1}^{N_1} \ln(p_{max}(i) - p_{min}(i) + 1)$$

$$+ \sum_{j=1}^{N_2} \ln(q_{max}(j) - q_{min}(j) + 1). \tag{14}$$

Theorem 3. *F_β^{MF} and U_β^{MF}, the mean field approximations of the exact free energy and internal energy of the system at the inverse temperature β, satisfy the following inequalities:*

$$U^* - \frac{A(N_1, N_2)}{\beta} \leq F_\beta^{MF} \leq U^*, \tag{15}$$

$$U^* \leq U_\beta^{MF} \leq U^* + \frac{A(N_1, N_2)}{\beta}, \tag{16}$$

where U^* is the optimal assignment energy and $A(N_1, N_2)$ is defined in Equation (14).

Proof. See Appendix F. □

Now, let us assume that the multi-assignment problem linked with the $N_1 \times N_2$ cost matrix C possesses a unique optimal assignment matrix, denoted as G^*. We have the following theorem:

Theorem 4. *Let $\overline{G}_\beta$ be the coupling matrix solution of the SPA equations at the inverse temperature β. We assume that the multi-assignment problem has a unique solution, with G^* being the associated coupling matrix. If Δ is the difference in total cost between the optimal solution and the second-best solution, then*

$$\max_{i,j} |\overline{G}_\beta(i,j) - G^*(i,j)| \leq \frac{A(N_1, N_2)}{\beta \Delta}. \tag{17}$$

Proof. See Appendix G. □

This theorem validates that, in the generic case in which the solution to the assignment problem is unique, the converged solution matrix $\overline{G}_\infty$ is this unique solution.

4. Solving Degenerate Assignment Problems

The method we propose is basically a relaxation approach to the general assignment problem. Indeed, we build a collection of transportation matrices $\overline{G}_\beta$ that belong to $\mathcal{U}^k(P_{min}^{max}, Q_{min}^{max}) \smallsetminus \mathcal{P}^k(P_{min}^{max}, Q_{min}^{max})$ (see Appendix A). Entries of these matrices are non-integer, strictly in the interval $(0, 1)$. If the general assignment problem is known to have a unique integer solution, we have shown that these matrices converge to an element of $\mathcal{P}^k(P_{min}^{max}, Q_{min}^{max})$ when $\beta \to +\infty$. The question remains as to what happens when the problem is degenerate, i.e., when it may have multiple integer solutions.

The general assignment problem is a linear programming problem. Checking if such a problem is degenerate is unfortunately often *NP*-complete ([22,23]). The degeneracies occur due to the presence of cycles in the linear constraints, i.e., in the cost matrix for the assignment problem. If this is the case, we propose randomly perturbing that matrix to bring it back to the generic problem. Megiddo and colleagues [24] have shown that an ε-perturbation of a degenerate linear programming problem reduces this problem to a non-degenerate one.

5. Matching: A Program for Solving Assignment Problems

We have implemented the multi-assignment framework described here in a C++ program, Matching, that is succinctly described in Algorithm 1.

Matching relies on an iterative process in which the parameter β is gradually increased. At each value of β, the nonlinear system of equations defined by Equation (10) is solved. We write this system as

$$\begin{cases} \mathbf{A}_\lambda = 0, \\ \mathbf{A}_\mu = 0, \\ A_x = 0, \end{cases}$$

where $\mathbf{A} = (\mathbf{A}_\lambda, \mathbf{A}_\mu, A_x)$ is a vector of predicates defined as

$$A_\lambda(i) = \sum_j h(\beta(C(i,j) + \lambda(i) + \mu(j) + x))$$
$$-g(\beta\lambda(i), p_{min}(i), p_{max}(i)),$$
$$A_\mu(j) = \sum_i h(\beta(C(i,j) + \lambda(i) + \mu(j) + x))$$
$$-g(\beta\mu(j), q_{min}(j), q_{max}(j)),$$
$$A_x = \sum_{i,j} h(\beta(C(i,j) + \lambda(i) + \mu(j) + x)) - k.$$

This system has $N_1 + N_2 + 1$ equations, with the same number of variables. It is solved using an iterative Newton–Raphson method (for details, see, for example [16,25]). Once the SPA system of equations is solved, the assignment matrix $\overline{G}_\beta$, the functions $\overline{n}_{1\beta}$, $\overline{n}_{2\beta}$, and the corresponding transportation energy $U^{MF}(\beta)$ are computed. The iterations over β are stopped if the mean field energy U_β^{MF} does not change anymore (within a tolerance TOL usually set to 10^{-6}). Otherwise, β is increased, and the current values of λ, μ and x are used as input for the following iteration. At convergence, the values of the assignment matrix are rounded to the nearest integer (indicated as $\lfloor\rceil$ in the output of Algorithm 1). The minimal energy is then computed using the corresponding integer matrix.

The primary computational expense of this algorithm arises from solving the nonlinear set of equations corresponding to the SPA at every β value. We use for this purpose an iterative Newton–Raphson method (see [15]).

Algorithm 1 Matching: a temperature-dependent framework for solving the multi-assignment problem

---

Input: N_1 and N_2, the number of agents and tasks; $p_{min}(i)$ and $p_{max}(i)$, the minimum and maximum number of agents needed to perform task i, and $q_{min}(j)$ and $q_{max}(j)$, the minimum and maximum number of tasks that can be performed by an agent j; k, the expected number of assignments; the cost matrix C. Initial value β_0 for β, the inverse of the temperature
Initialize: Initialize arrays λ and μ to 0 and initialize $x = 0$. Set $STEP = \sqrt{10}$.
for $i = 1, \ldots$ until convergence **do**
    (1) Initialize $\beta^i = STEP * \beta^{i-1}$.

    (2) Solve non linear Equation (10) for λ, μ and x at saddle point

    (3) Compute corresponding $\overline{G}_\beta^i$, $\overline{n}_{1\beta}^i$, $\overline{n}_{2\beta}^i$, and $U_\beta^{i,MF}$

    (4) Check for convergence: if $|U_\beta^{i,MF} - U_\beta^{i-1,MF}| < TOL$, stop
end for
Output: The converged assignment matrix $\lfloor \overline{G}_\beta \rfloor$, the functions $\lfloor \overline{n}_{1\beta} \rfloor$, $\lfloor \overline{n}_{2\beta} \rfloor$ over the agents and tasks, and the minimal associated cost U_β.

As for any annealing scheme, the initial temperature, or, in our case, the initial value β_0, is a parameter that significantly impacts the efficiency of the algorithm. Setting β_0 to be too small (i.e., a high initial temperature) will lead to inefficiency as the system will spend a significant amount of time at high temperatures, while setting β_0 too high will require many steps to converge at the corresponding low temperature, thereby decreasing the efficiency brought about by annealing. The value of β scales the cost matrix C and as such is related to the range of this matrix, more specifically to its largest value, C_{max}. We found that setting $\beta_0 C_{max} = 1$ provides satisfactory annealing efficiency. The value for $STEP$ was chosen empirically.

6. Examples

The framework we propose is general: it allows us to solve balanced and unbalanced assignment problems, including those that allow for multi-agent or multi-task assignments. We illustrate our method on the latter types of problems, as, to our knowledge, there are currently no solutions to those that are guaranteed to reach the optimal assignment.

As a first illustration of our framework, we ran Matching on the cost matrix C_1 given in Table 1. This matrix has been used in previous studies of multi-agent assignment problems [10,11,13,14,26,27]. Matching was run with the following parameters: $p_{min} = 0$ for all agents (i.e., some agents may stay idle), $p_{max} = 4$ for all agents (i.e., an agent can perform up to 4 tasks), $q_{min} = q_{max} = 1$ (i.e., a task is only performed by one agent), and $k = 8$ (all tasks are performed). In Figure 1, we show the corresponding trajectories of the internal energy U_β^{MF} and free energy F_β^{MF}.

Table 1. The cost matrix C_1.

Agents	Tasks							
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
A_1	300	250	180	320	270	190	220	260
A_2	290	310	190	180	210	200	300	190
A_3	280	290	300	190	190	220	230	260
A_4	290	300	190	240	250	190	180	210
A_5	210	200	180	170	160	140	160	180
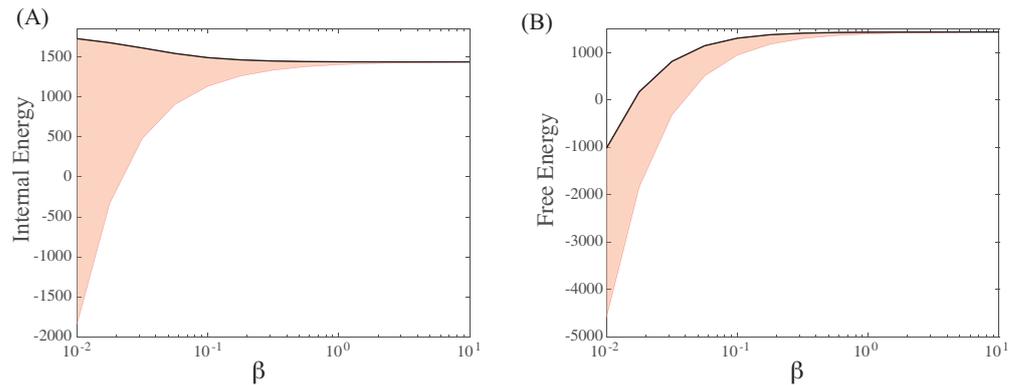
Figure 1. Convergence of the internal energy U_β^{MF} (**A**) and free energy F_β^{MF} (**B**) as a function of β when solving the multi-task assignment problem with the cost matrix C_1, such that each agent can perform up to 4 tasks, may be idle, and all 8 tasks are performed. On both panels, we show the bounds on the expected value for the optimal cost U^* as shaded areas (see text for details).

As expected, the internal energy is monotonically decreasing while the free energy is monotonically increasing, and both converge to the same value, 1440. Theorem 3 provides bounds on those energy values. Note that it can be rewritten as

$$U_\beta^{MF} - \frac{A(N_1, N_2)}{\beta} \le U^* \le U_\beta^{MF},$$

$$F_\beta^{MF} \le U^* \le U_\beta^{MF} + \frac{A(N_1, N_2)}{\beta},$$

i.e., at each value of β, we have bounds based on internal energy and free energy for the actual optimal cost of the assignment problem. These bounds are shown as shaded areas in Figure 1. Note that the widths of the intervals defined by the bounds are inversely proportional to β and therefore decrease monotonically as β increases.

In Table 2, we compare the assignment found by Matching on C_1 with p_{max} set to 4 (i.e., up to four tasks per agent), and $p_{min} = 0$ (agents can be idle) or $p_{min} = 1$ (each agent must execute at least one task). The removal of the latter constraint leads to a better optimal assignment in which agent 3 remains idle. When we maintain this constraint, our results are the same as those found by Wang et al. [13].

As a second more general test of our method, we compared the results of the framework we propose with the results of our own version of the Hungarian algorithm, as well as with those found in previous studies. We used both the matrix C1 and the matrix C2 given in Table 3 in those computing experiments.

Table 2. Assignments for C_1 with $p_{max} = 4$.

	$p_{min} = 0$	$p_{min} = 1$	Wang [a]
	$A_1 \rightarrow T_3$	$A_1 \rightarrow T_3$	$A_1 \rightarrow T_3$
	$A_2 \rightarrow T_4, T_8$	$A_2 \rightarrow T_8$	$A_2 \rightarrow T_8$
	A_3	$A_3 \rightarrow T_4$	$A_3 \rightarrow T_4$
	$A_4 \rightarrow T_7$	$A_4 \rightarrow T_7$	$A_4 \rightarrow T_7$
	$A_5 \rightarrow T_1, T_2, T_5, T_6$	$A_5 \rightarrow T_1, T_2, T_5, T_6$	$A_5 \rightarrow T_1, T_2, T_5, T_6$
Total cost	1440	1450	1450

[a] Assignment based on an ant colony algorithm [13].

Table 3. The cost matrix C_2.

	Tasks									
Agents	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
A_1	10	2	14	9	6	7	21	32	18	11
A_2	7	12	9	3	5	6	9	16	54	12
A_3	4	8	6	12	21	9	21	14	45	13
A_4	21	9	12	9	32	10	19	25	16	10
A_5	10	12	30	15	12	17	30	12	12	9
A_6	15	7	34	17	7	16	14	17	9	5

We ran Matching as follows. For all tasks j, we set $q_{min}(j) = q_{max}(j) = 1$, i.e., a task is only performed by one agent, and all tasks are expected to be executed. The latter is further enforced by setting k to be the total number of tasks (8 for C_1 and 10 for C_2). For an agent i, we either set all $p_{min}(i)$ to be 1 (in which case all agents must perform at least one task) or 0 (in which case some agents may not be assigned). We considered all $p_{max}(i)$ to be equal to a given value p and ran Matching with p varying from 1 to k. Note that both matrices C_1 and C_2 contain cycles. As such, optimal assignments based on those matrices are not unique. To remove the degeneracy, we added random uniform noise between $[0, \epsilon]$ with $\epsilon = 0.001$ to all values of the cost matrices (see Section 4 for details).

The Hungarian algorithm remains a standard for solving balanced as well as unbalanced assignment problems. It needs to be adapted, however, when solving multi-task assignment problems. The corresponding textbook solution that was used, for example, in Ref. [14] is to make copies or clones of the agents, solve the augmented unbalanced assignment problem using the Hungarian algorithm, and then regroup all clones of an agent to define the tasks that it is assigned to. We consequently created multiple versions of the cost matrices C_1 and C_2, with all agents copied p times, where p defines the maximum number of tasks that an agent can perform. We ran our own version of the Hungarian algorithm on those matrices. This version is adapted from the serial code written by C. Ma and is publicly available at https://github.com/mcximing/hungarian-algorithm-cpp, (accessed on 1 May 2021).

Comparisons of the results of Matching, of the Hungarian algorithm, and of previous studies are provided in Table 4 for the cost matrix C1 and in Table 5 for the cost matrix C2.

Table 4. Solving the multi-task assignment problems defined by the cost matrix C_1.

Method \ p [a]	2	3	4	5	6	7	8
Matching, $p_{min} = 1$	1520	1470	1450	1450	1450	1450	1450
Matching, $p_{min} = 0$	1520	1470	1440	1420	1410	1400	1400
Hungarian	1520	1470	1440	1420	1410	1400	1400
Majumdar GA [b]	1520	1470	1450	NA [c]	NA	NA	NA
Wang AC [d]	1520	1470	1450	NA	NA	NA	NA

[a] Each agent is allowed to execute up to p tasks. Note that as there are 8 tasks and 5 agents only, the minimum value of p is 2 if all the tasks must be executed. [b] Genetic algorithm [26]. [c] Not available in either [26] or [13]. [d] Ant colony algorithm [13].

Table 5. Solving the multi-task assignment problems defined by the cost matrix C_2.

Method \ p [a]	2	3	4	5	6	7	8
Matching, $p_{min} = 1$	66	65	65	65	65	65	65
Matching, $p_{min} = 0$	66	62	61	61	61	61	61
Hungarian	66	62	61	61	61	61	61
Majumdar GA [b]	66	67	66	74	NA [c]	NA	NA
Wang AC [d]	66	65	65	NA	NA	NA	NA

[a] Each agent is allowed to execute up to p tasks. Note that as there are 10 tasks and 6 agents only, and the minimum value of p is 2 if all the tasks must be executed. [b] Genetic algorithm [26]. [c] Not available in either [26] or [13]. [d] Ant colony algorithm [13].

The results of Matching have been proven to be optimal (see Section 3). As expected, those results are strongly dependent on the constraints imposed on the problem: we can find assignments with a lower total cost if we allow agents to remain idle by setting p_{min} to 0. For example, for the cost matrix C_1, in the extreme case with $p_{min} = 0$ and $p_{max} = 8$ for all agents, we find that all eight tasks have been assigned to agent 5, while the other agents have no task to perform, with a total cost of 1400. If instead we set $p_{min} = 1$ and keep $p_{max} = 8$, we find that agent 5 is assigned tasks 1, 2, 5, and 6; agent 1 is assigned task 3; agent 2 is assigned task 8; agent 3 is assigned task 4; and agent 4 is assigned task 7, for a total cost of 1450. If we instead run the Hungarian algorithm with agents that have been cloned to allow for multi-task assignments, we find optimal assignments that match those found by Matching with p_{min} set to 0 for both cost matrices. For matrix C_1, for example, if each agent is represented with 8 clones, we find that agent 5 is assigned tasks 1, 2, 4, 5, 6, 7, and 8 while agent 1 is assigned task 3, for a total cost of 1400. This illustrates two points. First, as expected, the optimal assignment is not unique, as we find two different assignments with the same cost, 1400. Second, and more importantly, this shows the difficulty of applying the Hungarian algorithm for such problems. It works fine if we want to find the optimal assignment without consideration of constraints, such as the constraint each agent needs to perform at least one task, but does not fit if such a constraint is necessary.

The optimal solutions found by Matching when p_{min} is set to 1 match those found by Wang et al. [13] for both matrices and are better than those found by Majumdar et al. [26] for matrix C_2. Note that the latter are obtained using either an ant colony algorithm or a genetic algorithm. Both are probabilistic algorithms that do not guarantee that the true minimum is found.

7. Conclusions

We have developed a general framework for solving balanced and unbalanced and constrained and unconstrained assignment problems. Given two sets of points S_1 and S_2 with the possibly different cardinalities N_1 and N_2, constraints on the number of possible assignments for each element of S_1 and of S_2, and a cost matrix defining the penalties of on assignment, we have constructed a concave free energy parameterized by temperature that captures those constraints. The definition of this free energy is general enough that it includes balanced and unbalanced cases. Its extremum establishes an optimal assignment between the two sets of points. We have demonstrated that this free energy consistently decreases as a function of β (the inverse of temperature) towards the optimal assignment cost. Moreover, we have established that for sufficiently large β values, the precise solution to the generic multi-assignment problem can be directly derived through straightforward rounding to the nearest integer of the elements of the assignment matrix linked to this extremum. Additionally, we have developed a method that guarantees convergence for addressing degenerate assignment problems.

The formalism introduced in this paper was designed to generalize the Hungarian algorithm for a large range of assignment problems. We expect it to be useful for a much larger set of problems, especially those associated with graphs. Graphs capture data

relationships and, as such, are essential to numerous applications. They are particularly useful in domains such as Web search [28], neural [29] and social network analysis [30], gene networks [31], etc. More generally, they are designed to represent complex knowledge. The scale of modern graphs leads to a need to develop more efficient methods to process very large graphs. The formalism we propose is well adapted to tackle this problem for applications such as bipartite or simple graph matching. We will also consider extensions to higher-order matching problems, such as *k*-matching problems [32] (for example the three-assignment problem [33]) that are known to be *NP*-complete [34]. These problems have their own applications for graph analyses.

Appendix A. A Generalized Birkhoff–von Neumann Theorem

An $N \times N$ matrix $A = (a(i,j))$ is said to be doubly stochastic if and only if it satisfies the following conditions,

$$
\begin{aligned}
a(i,j) &\geq 0, \\
\sum_{i=1}^{N} a(i,j) &= 1, \quad \sum_{j=1}^{N} a(i,j) = 1,
\end{aligned}
$$

for all $(i,j) \in [1,N]^2$. The set Ω_N of doubly stochastic matrices is a convex polytope whose vertices are the permutation matrices with the same size. This is expressed by the following theorem, established by [17,18]:

Theorem A1. *An $N \times N$ matrix A is doubly stochastic if and only if it can be written as a weighted sum of permutation matrices, i.e.,*

$$
A = \sum_{\pi \in \Pi_N} a_\pi \pi,
$$

where Π_N is the set of permutation matrices of size N, a_π is a positive real number and $\sum_{\pi \in \Pi_N} a_\pi = 1$.

Doubly stochastic matrices and their properties associated with the Birkhoff–von Neumann theorem above proved useful to establish convergence properties of statistical physics frameworks for solving the balanced assignment problem [15,35].

For the more general assignment problem considered here, however, we need to consider a different but related set of matrices, defined as follows. Let N and M be two natural numbers. Let $R_{min} = (r(1), \ldots, r(N))$, $R_{max} = (R(1), \ldots, R(N))$, $C_{min} = (c(1), \ldots, c(M))$, and $C_{max} = (C(1), \ldots, C(M))$ be four non-negative integral vectors satisfying

$$
\begin{aligned}
\forall i, \quad 0 &\leq r(i) \leq R(i), \\
\forall j, \quad 0 &\leq c(j) \leq C(j).
\end{aligned}
$$

Let A be a non-negative matrix of size $N \times M$, and let us denote the row sum vector of A as RA and the column sum vector of A as CA. Let us also define $\sigma(A)$ as the sum of all elements of A, i.e., $\sigma(A) = \sum_{i=1}^{N} \sum_{j=1}^{M} A(i, j)$. The transportation polytope $\mathcal{U}(R_{min}^{max}, C_{min}^{max})$ is the set of $N \times M$ matrices A that satisfy

$$\forall(i, j), 0 \leq A(i, j) \leq 1,$$
$$\forall i, \quad r(i) \leq RA(i) \leq R(i),$$
$$\forall j, \quad c(j) \leq CA(j) \leq C(j).$$

The transportation polytope $\mathcal{U}^k(R_{min}^{max}, C_{min}^{max})$ is the set of $N \times M$ matrices A that satisfy

$$\mathcal{U}^k(R_{min}^{max}, C_{min}^{max}) = \{A \in \mathcal{U}(R_{min}^{max}, C_{min}^{max}) | \quad \sigma(A) = k\}$$

Denote $\mathcal{P}(R_{min}^{max}, C_{min}^{max})$ as the set of all matrices in $\mathcal{U}(R_{min}^{max}, C_{min}^{max})$ whose entries are either 0 or 1 with a similar definition for $\mathcal{P}^k(R_{min}^{max}, C_{min}^{max})$ with respect to $\mathcal{U}^k(R_{min}^{max}, C_{min}^{max})$.

Relaxed solutions to the general assignment problem considered in this paper belong to $\mathcal{U}^k(P_{min}^{max}, Q_{min}^{max})$, while integer solutions belong to $\mathcal{P}^k(P_{min}^{max}, Q_{min}^{max})$, where P_{min}, represents the vectors containing the constraints on the number of tasks that can be assigned to an agent or the number of agents that can be assigned to a task. In Koehl [36], we show the following theorem:

Theorem A2. *$\mathcal{U}^k(R_{min}^{max}, C_{min}^{max})$ is the convex hull of all matrices in $\mathcal{P}^k(R_{min}^{max}, C_{min}^{max})$.*

Theorem A2 can be stated as any matrix in $\mathcal{U}^k(R_{min}^{max}, C_{min}^{max})$ that can be written as a linear combination of matrices in $\mathcal{P}^k(R_{min}^{max}, C_{min}^{max})$. This result will be useful for some of the proofs below.

Remark A1. *(i)   If $R_{min} = R_{max} = 1_N$, $C_{min} = C_{max} = 1_M$ (where 1_N and 1_N are vectors of one of size N and M, respectively, $N = M = k$, $\mathcal{U}(R_{min}^{max}, C_{min}^{max})$ is the set of doubly stochastic matrices, $\mathcal{P}^k(R_{min}^{max}, C_{min}^{max})$ is the set of permutation matrices, and Theorem A2 is then equivalent to the Birkhoff–von Neumann theorem for doubly stochastic matrices [17,18].*
(ii)   If $R_{min} = 0_n$, $C_{min} = 0_m$, $R_{max} = 1_n$, $C_{max} = 1_m$, $\mathcal{U}^k(R_{min}^{max}, C_{min}^{max})$ is the set of doubly sub-stochastic matrices with sum k and $\mathcal{P}^k(R_{min}^{max}, C_{min}^{max})$ is the set of sub-permutation matrices of rank k; a specific version of Theorem A2 was established by Mendelssohn and Dulmage for square matrices ([37]), and later by Brualdi and Lee for rectangular matrices ([38]).

Appendix B. Proof of Theorem 1: Concavity of the Effective Free Energy

The free energy associated with the multi-assignment problem can be written as

$$F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu}, x) = \quad -\frac{1}{\beta} \sum_i G(\beta\lambda(i), p_{min}(i), p_{max}(i))$$
$$-\frac{1}{\beta} \sum_j G(\beta\mu(j), q_{min}(j), q_{max}(j))$$
$$+\frac{1}{\beta} \sum_{i,j} H(\beta(C(i, j) + \lambda(i) + \mu(l) + x)) - kx,$$

with:

$$H(x) = -\ln(1 + e^{-x}),$$
$$G(x, a, b) = \ln\left(\frac{e^{ax} - e^{(b+1)x}}{1 - e^x}\right).$$

where a and b are two positive integers with $0 \le a \le b$. The functions $H(x)$, its derivatives, and some associated functions have been fully characterized in [15]. In parallel, the functions $G(x, a, b)$, its derivatives, and associated functions are characterized in Appendix H.

We prove that this free energy is weakly concave by showing that its Hessian HE is negative-definite. HE is a symmetric matrix of size $(N_1 + N_2 + 1) \times (N_1 + N_2 + 1)$, such that its rows and columns correspond to all N_1 λ values first, followed by all N_2 μ values, and finally to the value x. Let $h(x) = H'(x)$, and let h' be its derivative, i.e.,

$$h'(x) = -\frac{e^x}{(1 + e^x)^2}. \tag{A1}$$

In [15], we have shown that $h'(x) \in [\frac{-1}{4}, 0) \quad \forall x \in \mathbb{R}$, i.e., that $h'(x)$ is always strictly negative. Similarly, let $g(x) = G'(x)$, and let g' be its derivative, i.e.,

$$g'(x, a, b) = -\frac{(b - a + 1)^2 e^{(b-a+1)x}}{(e^{(b-a+1)x} - 1)^2} + \frac{e^x}{(e^x - 1)^2} \tag{A2}$$

In Appendix H, we show that when $b = a$, $g'(x, a, a) = 0 \quad \forall x \in \mathbb{R}$, and, when $b > a$, $g'(x, a, b)$ is strictly positive $\forall x \in \mathbb{R}$.

We define the matrix X' and the vector $\mathbf{d_1'}$ and $\mathbf{d_2'}$ such that

$$
\begin{aligned}
X'(i, j) &= h'(\beta(C(i, j) + \lambda(i) + \mu(j) + x), \\
d_1'(i) &= g'(\beta\lambda(i), p_{min}(i), p_{max}(i)), \\
d_2'(j) &= g'(\beta\mu(j), q_{min}(j), q_{max}(j)).
\end{aligned}
$$

From Equation (10), we obtain

$$
\begin{aligned}
HE(i, i') &= \frac{\partial^2 F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu}, x)}{\partial \lambda(i) \partial \lambda(i')} = \beta \delta_{ii'} \left(\sum_j X'(i, j) - d_1'(i) \right), \\
HE(i, j) &= \frac{\partial^2 F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu}, x)}{\partial \lambda(i) \partial \mu(j)} = \beta X'(i, j), \\
HE(i, N) &= \frac{\partial^2 F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu}, x)}{\partial \lambda(i) \partial x} = \beta \sum_j X'(i, j), \\
HE(j, j') &= \frac{\partial^2 F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu})}{\partial \mu(j) \partial \mu(j')} = \beta \delta_{jj'} \left(\sum_i X'(i, j) - d_2'(j) \right), \\
HE(j, N) &= \frac{\partial^2 F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu}, x)}{\partial \mu(j) \partial x} = \beta \sum_i X'(i, j), \\
HE(N, N) &= \frac{\partial^2 F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu}, x)}{\partial x^2} = \beta \sum_i \sum_j X'(i, j),
\end{aligned}
$$

where δ are Kronecker functions, the indices i and i' belong to $[1, N_1]$ and the indices j and j' belong to $[1, N_2]$, and we have defined $N = N_1 + N_2 + 1$.

Let $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2}, x_3)$ be an arbitrary vector of size N. The quadratic form $Q(\mathbf{x}) = \mathbf{x}^T H \mathbf{x}$ is equal to

$$
\begin{aligned}
Q(\mathbf{x}) = \quad & \sum_{i,i'} x_1(i)HE(i,i')x_1(i') + \sum_{j,j'} x_2(j)HE(j,j')x_2(j') + \\
& x_3 HE(N,N)x_3 + 2\sum_{i,j} x_1(i)HE(i,j)x_2(j) + \\
& 2\sum_{i} x_1(i)HE(i,N)x_3 + 2\sum_{j} x_2(j)HE(j,N)x_3 \\
= \quad & \beta \sum_{i,j} (x_1(i) + x_2(j) + x_3)^2 X'(i,j) \\
& -\beta \left(\sum_{i} x_1(i)^2 d_1'(i) + \sum_{j} x_2(j)^2 d_2'(j) \right).
\end{aligned}
$$

As $X'(i,j)$ is based on the function h' that is negative and $d_1'(i)$, and $d_2'(j)$ are based on the function g' that is positive, the summands in the equation above are all negative for all $i \in [1, N_1]$ and $j \in [1, N_2]$, and therefore, $Q(\mathbf{x})$ is negative for all vector $\mathbf{x}$. The Hessian H is negative, semidefinite.

To check for definitiveness, let us note first that as $Q(\mathbf{x})$ is a sum of negative terms, it is 0 if and only if all the terms are equal to 0. As the function $h'(x)$ is strictly negative, this means that $\forall (i,j)$,

$$
(x_1(i) + x_2(j) + x_3)^2 = 0. \tag{A3}
$$

For the two other terms, we consider two cases:

(i) $p_{min}(i) < p_{max}(i) \quad \forall i$ and $q_{min}(j) < q_{max}(j) \quad \forall j$. Then $g'(x, p_{min}(i), p_{max}(i)) < 0$ and $g'(x, q_{min}(j), q_{max}(j)) < 0$. This leads to $\forall (i,j)$:

$$
\begin{aligned}
x_1(i)^2 &= 0, \\
x_2(j)^2 &= 0, \tag{A4}
\end{aligned}
$$

Equations (A3) and (A4) are satisfied when all $x_1(i)$, $x_2(j)$, and when x_3 are zero, namely that $\mathbf{x} = \mathbf{0}$. Therefore in this case, H is negative, definite, and the free energy $F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu}, x)$ is strictly concave.

(ii) For all i $p_{min}(i) = p_{max}(i)$ or for all j $q_{min}(j) = q_{max}(j)$. Then, either $d_1'(x) = 0 \quad \forall x \in \mathbb{R}^{N_1}$ or $d_2'(x) = 0 \quad \forall x \in \mathbb{R}^{N_2}$, or both. There are then non-zero solutions to the equation $Q(\mathbf{x}) = 0$. The Hessian is then only semidefinite.

Appendix C. Monotonicity of F_β^{MF}

The effective free energy $F_\beta(\boldsymbol{\lambda}, \boldsymbol{\mu}, x)$ defined in Equation (8) is a function of the cost matrix C and of real unconstrained variables $\lambda(i)$, $\mu(j)$, and x. For the sake of simplicity, for any $(i,j) \in [1, N_1] \times [1, N_2]$, we define

$$
\begin{aligned}
y(i,j) &= C(i,j) + \lambda(i) + \mu(j) + x, \\
y^{MF}(i,j) &= C(i,j) + \lambda^{MF}(i) + \mu^{MF}(j) + x^{MF}.
\end{aligned}
$$

The effective free energy is then

$$F_\beta(\lambda, \mu, x) = \quad -\frac{1}{\beta} \sum_i G(\beta\lambda(i), p_{min}(i), p_{max}(i))$$

$$-\frac{1}{\beta} \sum_j G(\beta\mu(j), q_{min}(j), q_{max}(j))$$

$$+\frac{1}{\beta} \sum_{i,j} H(\beta y(i,j)) - kx.$$

As written above, $F_\beta(\lambda, \mu, x)$ is a function of the independent variables β, $\lambda(i)$, $\mu(j)$, and x. However, under the saddle point approximation, the variables $\lambda(i)$, $\mu(j)$, and x are constrained by the conditions

$$\frac{\partial F_\beta(\lambda, \mu, x)}{\partial \lambda(i)} = \frac{\partial F_\beta(\lambda, \mu, x)}{\partial \mu(j)} = \frac{\partial F_\beta(\lambda, \mu, x)}{\partial x} = 0,$$

and the free energy under those constraints is written as F_β^{MF}. In the following, we use the notations $\frac{dF_\beta^{MF}}{d\beta}$ and $\frac{\partial F_\beta^{MF}}{\partial \beta}$ to differentiate between the total derivative and partial derivative of F_β^{MF} with respect to β, respectively. It can be shown easily that (see Appendix C of [15])

$$\frac{dF_\beta^{MF}}{d\beta} = \frac{\partial F_\beta\left(\lambda^{MF}, \mu^{MF}, x^{MF}\right)}{\partial \beta},$$

namely that the total derivative with respect to β is, in this specific case, equal to the corresponding partial derivative, which is easily computed to be

$$\frac{\partial F_\beta^{MF}}{\partial \beta} =$$

$$\frac{1}{\beta^2} \sum_i G(\beta\lambda^{MF}(i), p_{min}(i), p_{max}(i)) -$$

$$\frac{1}{\beta} \lambda^{MF}(i) g(\beta\lambda^{MF}(i), p_{min}(i), p_{max}(i)) +$$

$$\frac{1}{\beta^2} \sum_j G(\beta\mu^{MF}(j), q_{min}(j), q_{max}(j)) -$$

$$\frac{1}{\beta} \mu^{MF}(j) g(\beta\mu^{MF}(j), q_{min}(j), q_{max}(j))$$

$$-\frac{1}{\beta^2} \sum_{i,j} \left[H(\beta y^{MF}(i)) - \beta y^{MF}(i,j) h(\beta y^{MF}(i,j)) \right].$$

Let $t(x) = H(x) - xh(x)$. The function $t(x)$ is continuous and defined over all real values x and is bounded above by 0, i.e., $t(x) \leq 0 \quad \forall x \in \mathbb{R}$. Similarly, let us define $l(x, a, b) = G(x, a, b) - xg(x, a, b)$. In Appendix H, we show that $l(x, a, b)$ is positive over $\mathbb{R}$. As

$$
\begin{aligned}
\frac{dF_\beta^{MF}}{d\beta} = \quad & \frac{1}{\beta^2} \sum_i l(\beta \lambda^{MF}(i), p_{min}(i), p_{max}(i)) + \\
& \frac{1}{\beta^2} \sum_j l(\beta \mu^{MF}(j), q_{min}(j), q_{max}(j)) \\
& -\frac{1}{\beta^2} \sum_{i,j} t(\beta y^{MF}(i,j)),
\end{aligned} \tag{A5}
$$

we conclude that

$$
\frac{dF_\beta^{MF}}{d\beta} \geq 0,
$$

namely that F_β^{MF} is a monotonically increasing function of β.

Appendix D. Monotonicity of U_β^{MF}

Let

$$
U_\beta(\lambda, \mu, x) = \sum_{i,j} C(i,j) \bar{G}(i,j),
$$

and let the corresponding mean field approximation of the internal energy at the saddle point,

$$
U_\beta^{MF} = U_\beta \left(\lambda^{MF}, \mu^{MF}, x^{MF} \right).
$$

Before computing $\frac{dU_\beta^{MF}}{d\beta}$, we prove the following property

Proposition A1.

$$
U_\beta^{MF} = F_\beta^{MF} + \beta \frac{dF_\beta^{MF}}{d\beta}, \tag{A6}
$$

i.e., it extends the well-known relationship between the free energy and the average energy to their mean field counterparts.

Proof. The proof of this proposition is virtually identical to the proof of the same results for the assignment problem, found in Appendix C of [15]. $\square$

Based on the chain rule,

$$
\begin{aligned}
\frac{dU_\beta^{MF}}{d\beta} = \quad & \frac{\partial U_\beta^{MF}}{\partial \beta} + \sum_i \frac{\partial U_\beta^{MF}}{\partial \lambda(i)} \frac{\partial \lambda(i)}{\partial \beta} + \sum_j \frac{\partial U_\beta^{MF}}{\partial \mu(j)} \frac{\partial \mu(j)}{\partial \beta} + \\
& \frac{\partial U_\beta^{MF}}{\partial x} \frac{\partial x}{\partial \beta}.
\end{aligned}
$$

Using Proposition A1, we find that the partial derivatives of U_β^{MF} with respect to λ, μ, and x are all zeros, and therefore,

$$
\frac{dU_\beta^{MF}}{d\beta} = \quad \frac{\partial U_\beta^{MF}}{\partial \beta}.
$$

Using again Proposition A1, we obtain

$$
\begin{aligned}
\frac{dU_\beta^{MF}}{d\beta} &= \frac{dF_\beta^{MF}}{d\beta} + \frac{d}{d\beta}\left(\beta\frac{dF_\beta^{MF}}{d\beta}\right) \\
&= 2\frac{\partial F_\beta^{MF}}{\partial\beta} + \beta\frac{\partial}{\partial\beta}\left(\frac{\partial F_\beta^{MF}}{\partial\beta}\right).
\end{aligned}
$$

Using Equation (A5),

$$
\begin{aligned}
\beta^2\frac{dU_\beta^{MF}}{d\beta} &= \sum_i \beta\lambda(i)l'(\beta\lambda^{MF}(i), p_{min}(i), p_{max}(i)) \\
&+ \sum_j \beta\mu^{MF}(j)l'(\beta\mu^{MF}(j), q_{min}(j), q_{max}(j)) \\
&- \sum_{i,j} \beta y^{MF}(i,j)t'(\beta y^{MF}(i,j)),
\end{aligned}
$$

where $t(x)$ and $l(x, a, b)$ are the functions defined above, and $t'(x)$ and $l'(x, a, b)$ are their derivatives with respect to x. Let us define $r(x) = xt'(x)$ and $s(x, a, b) = xl'(x, a, b)$. Then,

$$
\begin{aligned}
\beta^2\frac{\partial U_\beta^{MF}}{\partial\beta} &= \sum_i s(\beta\lambda^{MF}(i), p_{min}(i), p_{max}(i)) + \\
&\sum_j s(\beta\mu^{MF}(j), q_{min}(j), q_{max}(j)) - \\
&\sum_{i,j} r(\beta y^{MF}(i,j)).
\end{aligned}
$$

Note that $r(x) = \dfrac{x^2 e^x}{(1+e^x)^2}$. Therefore, $r(x)$ is positive, bounded below by 0. In Appendix H, we show that $s(x, a, b)$ is negative, bounded above 0. Therefore,

$$
\frac{dU_\beta^{MF}}{d\beta} = \frac{\partial U_\beta^{MF}}{\partial\beta} \leq 0,
$$

and the function U_β^{MF} is a monotonically decreasing function of β.

Appendix E. Proof of Theorem 2: Convergence of the Mean Field Free Energy and the Internal Energy to the Optimal Assignment Cost

For simplicity in notation, we define $F^{MF}(\infty) = \lim_{\beta\to+\infty} F_\beta^{MF}$ and $U^{MF}(\infty) = \lim_{\beta\to+\infty} U_\beta^{MF}$.

Appendix E.1. Defining Entropy

Recall that we have defined

$$
\begin{aligned}
y(i,j) &= C(i,j) + \lambda(i) + \mu(j) + x, \\
X(i,j) &= h(\beta y(i,j)), \\
d_1(i) &= g(\beta\lambda(i), p_{min}(i), p_{max}(i)), \\
d_2(j) &= g(\beta\mu(j), q_{min}(j), q_{max}(j)).
\end{aligned}
$$

The free energy is given by

$$
\begin{aligned}
F_\beta(\pmb{\lambda},\pmb{\mu},x) \quad &= -\frac{1}{\beta}\sum_i G(\beta\lambda(i), p_{min}(i), p_{max}(i)) \\
&\quad -\frac{1}{\beta}\sum_j G(\beta\mu(j), q_{min}(j), q_{max}(j)) \\
&\quad +\frac{1}{\beta}\sum_{i,j} H(\beta y(i,j)) - kx,
\end{aligned}
$$

while the internal energy is

$$
U_\beta(\pmb{\lambda},\pmb{\mu},x) = \sum_{i,j} C(i,j) X(i,j).
$$

By adding and subsequently subtracting the internal energy in the equation defining the free energy, and then replacing $C(i,j)$ with $y(i,j) - \lambda(i) - \mu(j) - x$, we obtain (after some reorganization),

$$
\begin{aligned}
F_\beta(\pmb{\lambda},\pmb{\mu},x) &= U_\beta(\pmb{\lambda},\pmb{\mu},x) \\
&\quad -\frac{1}{\beta}\sum_i l(\beta\lambda(i), p_{min}(i), p_{max}(i)) \\
&\quad -\frac{1}{\beta}\sum_j l(\beta\mu(j), q_{min}(j), q_{max}(j)) + \frac{1}{\beta}\sum_{i,j} t(\beta y(i,j)) \\
&\quad +x\left(\sum_{i,j} X(i,j) - k\right) + \sum_i \lambda(i)\left(\sum_j X(i,j) - d_1(i)\right) \\
&\quad + \sum_j \mu(j)\left(\sum_i X(i,j) - d_2(j)\right),
\end{aligned}
$$

where $t(x)$ and $l(x,a,b)$ are defined above. Let us define

$$
\begin{aligned}
S_\beta(\pmb{\lambda},\pmb{\mu},x) = \sum_i l(\beta\lambda(i), p_{min}(i), p_{max}(i)) + \\
\sum_j l(\beta\mu(j), q_{min}(i), q_{max}(i)) - \sum_{i,j} t(\beta y(i,j)).
\end{aligned} \tag{A7}
$$

Then,

$$
\begin{aligned}
F_\beta(\pmb{\lambda},\pmb{\mu},x) = U_\beta(\pmb{\lambda},\pmb{\mu},x) - \frac{1}{\beta} S_\beta(\pmb{\lambda},\pmb{\mu},x) \\
+x\left(\sum_{i,j} X(i,j) - k\right) + \sum_i \lambda(i)\left(\sum_j X(i,j) - d_1(i)\right) + \\
\sum_j \mu(j)\left(\sum_i X(i,j) - d_2(j)\right).
\end{aligned} \tag{A8}
$$

The form of the free energy given in Equation (A8) has an intuitive physical interpretation. The first term is the original unbalanced assignment energy, the second is -T times an entropy term (defined in Equation (A7)), and the third, fourth, and fifth terms impose the constraints via Lagrange multipliers. At the saddle point, these constraints are satisfied, and the free energy has the form

$$
F_\beta^{MF} = U_\beta^{MF} - \frac{1}{\beta} S_\beta^{MF}. \tag{A9}
$$

Let us find the bound for the entropy term. $t(x)$ is negative, bounded below by $-\ln(2)$. Similarly, in Appendix H, we show that $l(x, a, b)$ is positive, bounded above by $\ln(b - a + 1)$. Let us define

$$
\begin{aligned}
A(N_1, N_2) = \quad & N_1 N_2 \ln(2) + \\
& \sum_{i=1}^{N_1} \ln(p_{max}(i) - p_{min}(i) + 1) + \\
& \sum_{j=1}^{N_2} \ln(q_{max}(j) - q_{min}(j) + 1).
\end{aligned}
\tag{A10}
$$

Then, the entropy satisfies the following constraints:

$$
0 \leq S_\beta^{MF} \leq A(N_1, N_2).
\tag{A11}
$$

Appendix E.2. $F^{MF}(\infty) = U^{MF}(\infty)$

Proof. Using Equation (A22), after rearrangements, we obtain

$$
U_\beta^{MF} - \frac{1}{\beta} A(N_1, N_2) \leq F_\beta^{MF} \leq U_\beta^{MF}.
$$

Taking the limits when $\beta \to +\infty$, we obtain

$$
F^{MF}(\infty) = U^{MF}(\infty).
\tag{A12}
$$

□

Appendix E.3. $U^* \leq F^{MF}(\infty)$

Proof. Let $U^{MF}(\beta)$ be the mean field internal energy at the inverse temperature β:

$$
U_\beta^{MF} = \sum_{i,j} C(i, j) X_\beta^{MF}(i, j),
$$

where X_β^{MF} is the solution the the SPA system of equations. At a finite inverse temperature β, X_β^{MF} is strictly non-integral. In addition, X_β^{MF} satisfies constraints on row sums and column sums that make it an element of $\mathcal{U}^k(R_{min}^{max}, C_{min}^{max})$; see appendix A for details. Based on Theorem A2, X_β^{MF} can be written as a linear combination of matrices in $\mathcal{P}^k(R_{min}^{max}, C_{min}^{max})$,

$$
X_\beta^{MF} = \sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max})} a_\pi \pi,
$$

with all $a_\pi \in [0, 1]$ and $\sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max})} a_\pi = 1$. Therefore,

$$
\begin{aligned}
U_\beta^{MF} &= \sum_{i,j} C(i, j) X_\beta^{MF}(i, j) \\
&= \sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max})} a_\pi \sum_{i,j} C(i, j) \pi(i, j).
\end{aligned}
\tag{A13}
$$

As U^* is the minimum matching cost over all possible matrices in $\mathcal{P}^k(R_{min}^{max}, C_{min}^{max})$, we have

$$
\sum_{i,j} C(i, j) \pi(i, j) \geq U^*,
\tag{A14}
$$

for all $\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max})$. Combining Equations (A13) and (A14), we obtain

$$U_{\beta}^{MF} \geq \sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max})} a_{\pi} U^*$$

$$\geq \left(\sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max})} a_{\pi} \right) U^*,$$

from which we conclude that at each β,

$$U^* \leq U_{\beta}^{MF}.$$

Therefore $U^* \leq U^{MF}(\infty)$, and consequently $U^* \leq F^{MF}(\infty)$, based on Equation (A12). $\square$

Appendix E.4. $U^* \geq F^{MF}(\infty)$

Proof. Let us first recall the definition of the free energy,

$$
\begin{aligned}
F_{\beta}(\boldsymbol{\lambda}, \boldsymbol{\mu}, x) =\ & -\frac{1}{\beta} \sum_i \ln \left(\frac{e^{\beta p_{min}(i)\lambda(i)} - e^{\beta(p_{max}(i)+1)\lambda(i)}}{1 - e^{\beta\lambda(i)}} \right) \\
& -\frac{1}{\beta} \sum_j \ln \left(\frac{e^{\beta q_{min}(j)\mu(j)} - e^{\beta(q_{max}(j)+1)\mu(j)}}{1 - e^{\beta\mu(j)}} \right) \\
& -\frac{1}{\beta} \sum_{i,j} \ln \left[1 + e^{-\beta y(i,j)} \right] - kx.
\end{aligned}
$$

Note these two properties of limits:

$$\lim_{\beta \to +\infty} \frac{\ln(1 + e^{-a\beta})}{\beta} = \begin{cases} 0 & \text{if } a \geq 0 \\ -a & \text{if } a \leq 0, \end{cases}$$

and

$$\lim_{\beta \to +\infty} \frac{1}{\beta} \ln \left(\frac{e^{xa\beta} - e^{x(b+1)\beta}}{1 - e^{x\beta}} \right) = \begin{cases} xb & \text{if } x \geq 0 \\ xa & \text{if } x \leq 0, \end{cases}$$

when a and b are non-negative integers. Therefore,

$$
\begin{aligned}
\lim_{\beta \to +\infty} F_{\beta}(\boldsymbol{\lambda}, \boldsymbol{\mu}, x) =\ & \\
& - \sum_{i \mid \lambda(i) \geq 0} p_{max}(i)\lambda(i) - \sum_{i \mid \lambda(i) < 0} p_{min}(i)\lambda(i) \\
& - \sum_{j \mid \mu(j) \geq 0} q_{max}(j)\mu(j) - \sum_{j \mid \mu(j) < 0} q_{min}(j)\mu(j) \\
& + \sum_{(i,j) \mid y(i,j) \leq 0} y(i,j) - kx.
\end{aligned}
\tag{A15}
$$

Let us consider now a matrix π in $\mathcal{P}^k(R_{min}^{max}, C_{min}^{max})$ (see above). We can write

$$
\begin{aligned}
\sum_{i,j} C(i,j)\pi(i,j) =\ & \sum_{i,j} y(i,j)\pi(i,j) - \sum_i \lambda(i) RS(i) \\
& - \sum_j \mu(j) CS(j) - kx,
\end{aligned}
\tag{A16}
$$

where RS and CS are the row sums and column sums of π, respectively. For each index i, the summand included in the first term on the right is always larger than or equal to the sum of all the corresponding terms that are negative:

$$\sum_{i,j} y(i,j)\pi(i,j) \geq \sum_{i|y(i,j)\leq 0} y(i,j). \tag{A17}$$

As π belongs to $\mathcal{P}^k(R_{min}^{max}, C_{min}^{max})$, its row sums satisfy the constraints:

$$p_{min}(i) \leq RS(i) \leq p_{max}(i). \tag{A18}$$

We multiply this equation separately for positive and negative $\lambda(i)$. We find

$$\sum_{i|\lambda(i)>0} p_{min}(i)\lambda(i) \leq \sum_{i|\lambda(i)>0} \lambda(i)RS(i) \leq \sum_{i|\lambda(i)>0} p_{max}(i)\lambda(i)$$

$$\sum_{i|\lambda(i)<0} p_{max}(i)\lambda(i) \leq \sum_{i|\lambda(i)<0} \lambda(i)RS(i) \leq \sum_{i|\lambda(i)<0} p_{min}(i)\lambda(i).$$

Therefore,

$$-\sum_i \lambda(i)RS(i) \geq -\sum_{i|\lambda(i)\geq 0} p_{max}(i)\lambda(i) - \sum_{i|\lambda(i)<0} p_{min}(i)\lambda(i). \tag{A19}$$

Similarly, we can show that

$$-\sum_j \mu(j)CS(j) \geq -\sum_{j|\mu(j)\geq 0} q_{max}(j)\mu(j) - \sum_{j|\mu(j)<0} q_{min}(j)\mu(j). \tag{A20}$$

Combining Equations (A15)–(A17), (A19) and (A20), we obtain

$$\sum_{i,j} C(i,j)\pi(i,j) \geq \lim_{\beta\to+\infty} F_\beta(\lambda,\mu,x). \tag{A21}$$

Equation (A21) is valid for all matrices π in $\mathcal{P}^k(R_{min}^{max}, C_{min}^{max})$. It is therefore valid for the optimal π^* that solves the general assignment problem. Since $U^* = \sum_{i,j} C(i,j)\pi^*(i,j)$, we have

$$U^* \geq \lim_{\beta\to+\infty} F_\beta(\lambda,\mu,x).$$

As this equation is true for all λ, μ, and x, it is true in particular for $\lambda = \lambda^{MF}$, $\mu = \mu^{MF}$, and $x = x^{MF}$, leading to

$$U^* \geq \lim_{\beta\to+\infty} F_\beta^{MF} = F^{MF}(\infty).$$

$\square$

We have shown that $U^* \leq F^{MF}(\infty)$ and $F^{MF}(\infty) \leq U^*$, and therefore $U^* = F^{MF}(\infty)$. The corresponding result for the internal energy, $U^* = U^{MF}(\infty)$ follows directly from Equation (A12).

Appendix F. Proof of Theorem 3: Bounds on the Internal Energy and Free Energy

Appendix F.1. Bounds on the Free Energy

Proposition A1 from Appendix D states that

$$\beta \frac{dF_\beta^{MF}}{d\beta} = -F_\beta^{MF} + U_\beta^{MF}.$$

Using this equation and the relationship between free energy, energy, and entropy at SPA (see Equation (A22)), we obtain

$$\frac{dF_\beta^{MF}}{d\beta} = \frac{1}{\beta^2} S_\beta^{MF}.$$

From the bounds on the entropy (Equation (A11)),

$$0 \le \frac{dF_\beta^{MF}}{d\beta} \le \frac{A(N_1, N_2)}{\beta^2}.$$

By integrating over β between $+\infty$ and β,

$$0 \le F^{MF}(\infty) - F_\beta^{MF} \le \frac{A(N_1, N_2)}{\beta}.$$

Finally, as $F^{MF}(\infty) = U^*$,

$$U^* - \frac{A(N_1, N_2)}{\beta} \le F_\beta^{MF} \le U^*. \tag{A22}$$

Appendix F.2. Bounds on the Energy

As $U_\beta^{MF} = F_\beta^{MF} + \frac{1}{\beta} S_\beta^{MF}$, using the bounds on the free energy and on the entropy, we obtain

$$U_\beta^{MF} \le U^* + \frac{A(N_1, N_2)}{\beta}. \tag{A23}$$

In addition, U_β^{MF} is monotonic, decreasing, with the limit U^* as $\beta \to +\infty$, $U^* \le U_\beta^{MF}$. Therefore,

$$U^* \le U_\beta^{MF} \le U^* + \frac{A(N_1, N_2)}{\beta}.$$

Appendix G. Proof of Theorem 4: Bounds on the Transportation Matrix $\overline{G}_\beta$

This proof is inspired by the proof of Theorem 6 in Appendix 2 of [35] and by Appendix F of [15].

We first recall that $\overline{G}_\beta$ belongs to $\mathcal{U}^k(R_{min}^{max}, C_{min}^{max})$ (see above), with $\sigma(\overline{G}_\beta) = k$. As such, it can be written as a linear combination of the matrices $\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max})$,

$$\overline{G}_\beta = \sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max})} a_\pi \pi,$$

with all $a_\pi \in [0, 1]$ and $\sum\limits_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max})} a_\pi = 1$.

We assume that the general assignment problem considered has a unique solution. We want to prove that $\max\limits_{i,j} |\overline{G}_\beta(i, j) - G^*(i, j)| \le \frac{A(N_1, N_2)}{\beta \Delta}$, where G^* is the optimal solution of the assignment problem, $\Delta = U^{2*} - U^*$ is the difference in total cost between the second

best solution and the optimal solution, and $A(N_1, N_2)$ is defined in Equation (A10). We use for that a proof by contradiction. We assume that there exists a pair (i, j) such that

$$\frac{A(N_1, N_2)}{\beta \Delta} < \left| \bar{G}_\beta(i, j) - G^*(i, j) \right|.$$

Let us denote $B(i, j) = \left| \bar{G}_\beta(i, j) - G^*(i, j) \right|$. As G^* is a binary matrix, $G^*(i, j) = 0$ or $G^*(i, j) = 1$.

In the first case,

$$\begin{aligned} B(i, j) &= \bar{G}_\beta(i, j) \\ &= \sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max})} a_\pi \pi(i, j). \end{aligned}$$

Since G^* belongs to $\mathcal{P}^k(R_{min}^{max}, C_{min}^{max})$, it is included in the decomposition of $\overline{G}_\beta$, and therefore,

$$\begin{aligned} B(i, j) &= a_{G^*} G^*(i, j) + \sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max}) - \{G^*\}} a_\pi \pi(i, j) \\ &= \sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max}) - \{G^*\}} a_\pi \pi(i, j) \\ &< \sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max})) - \{G^*\}} a_\pi = 1 - a_{G^*}, \end{aligned}$$

where the final equality follows from the fact that the sum of all coefficients a is equal to 1.

In the second case, $G^*(i, j) = 1$,

$$\begin{aligned} B(i, j) &= 1 - \overline{G}_\beta(i, j) \\ &= 1 - \sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max})} a_\pi \pi(i, j). \end{aligned}$$

Again, as G^* is included in the decomposition of $\overline{G}_\beta$,

$$\begin{aligned} B(i, j) &= 1 - a_{G^*} G^*(i, j) - \sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max}) - \{G^*\}} a_\pi \pi(i, j) \\ &= 1 - a_{G^*} - \sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max}) - \{G^*\}} a_\pi \pi(i, j) \\ &< 1 - a_{G^*}, \end{aligned}$$

where the final inequality follows from the fact that $\sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max}) - \{G^*\}} a_\pi \pi(i, j)$ is positive.

In conclusion, in both cases, we have

$$\frac{A(N_1, A_2)}{\beta \Delta} < 1 - a_{G^*}. \tag{A24}$$

Now, let us look at the energy associated with $\bar{G}_\beta$:

$$
\begin{aligned}
U_\beta^{MF} &= \sum_{i,j} C(i,j)\overline{G}_\beta(i,j) \\
&= \sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max})} a_\pi \sum_{i,j} C(i,j)\pi(i,j) \\
&= a_{G^*}U^* + \sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max}) - \{G^*\}} a_\pi \sum_{i,j} C(i,j)\pi(i,j) \\
&\geq a_{G^*}U^* + \left(\sum_{\pi \in \mathcal{P}^k(R_{min}^{max}, C_{min}^{max}) - \{G^*\}} a_\pi \right) U^{2*} \\
&\geq a_{G^*}U^* + (1 - a_{G^*})U^{2*} \\
&\geq U^* + (1 - a_{G^*})\Delta.
\end{aligned}
$$

In Theorem 3, we have shown that

$$
U^* \leq U_\beta^{MF} \leq U^* + \frac{A(N_1, N_2)}{\beta}.
$$

Therefore,

$$
U^* + (1 - a_{G^*})\Delta \leq U^* + \frac{A(N_1, N_2)}{\beta},
$$

i.e.,

$$
(1 - a_{G^*}) \leq \frac{A(N_1, N_2)}{\beta\Delta}, \tag{A25}
$$

as Δ is strictly positive.

We have shown that $\frac{A(N_1, N_2)}{\beta\Delta} < 1 - a_{G^*}$ (Equation (A24)) and $(1 - a_{G^*}) \leq \frac{A(N_1, N_2)}{\beta\Delta}$ (Equation (A25)); i.e., we have reached a contradiction. Our hypothesis is wrong, and therefore $\max_{i,j}\left|\overline{G}_\beta(i,j) - G^*(i,j)\right| \leq \frac{A(N_1, N_2)}{\beta\Delta}$.

Appendix H. Properties of the General Function $G(x, a, b)$ and Its Derivatives

Theorem A3. *Let $G(x, a, b)$ be the function defined by,*

$$
G(x, a, b) = \ln\left(\frac{e^{ax} - e^{(b+1)x}}{1 - e^x} \right).
$$

where a and b are two positive integers with $0 \leq a < b$. Let $g(x) = G'(x)$, $l(x, a, b) = G(x, a, b) - xg(x, a, b)$, and $s(x, a, b) = xl'(x, a, b)$. We have the following properties:

(1) $g'(x, a, b) > 0 \quad \forall x \in \mathbb{R}.$
(2) $0 \leq l(x, a, b) \leq \ln(b - a + 1) \quad \forall x \in \mathbb{R}.$
(3) $s(x, a, b) \leq 0 \quad \forall x \in \mathbb{R}.$

Note that in Theorem A3, we consider a, b non-negative, with $a < b$. The result can be trivially expanded to the case $a = b$, with $G(x, a, a) = ax$ and therefore $g(x, a, a) = a$, $g'(x, a, a) = 0$, $l(x, a, a) = 0$, and $s(x, a, a) = 0$. In the following, we only consider $a < b$.

Appendix H.1. The Function $G(x, a, b)$

Note first that

$$
\begin{aligned}
G(x, a, b) &= \ln\left(\frac{e^{ax} - e^{(b+1)x}}{1 - e^x}\right) \\
&= \ln\left(\sum_{k=a}^{b} e^{kx}\right).
\end{aligned}
$$

$G(x, a, b)$ is therefore defined, continuous, and differentiable over $\mathbb{R}$. Some special values:

$$
\begin{aligned}
G(0, a, b) &= \ln(b + a - 1), \\
\lim_{x \to -\infty} G(x, a, b) &= -\infty, \\
\lim_{x \to +\infty} G(x, a, b) &= +\infty.
\end{aligned}
$$

Appendix H.2. $g(x, a, b)$ and Its Derivative $g'(x, a, b)$

Note that

$$
\begin{aligned}
g(x, a, b) &= \frac{(b - a + 1)}{e^{(b-a+1)x} - 1} + \frac{1}{1 - e^x} + b, \\
g'(x, a, b) &= -\frac{(b - a + 1)^2 e^{(b-a+1)x}}{(e^{(b-a+1)x} - 1)^2} + \frac{e^x}{(e^x - 1)^2},
\end{aligned}
$$

g and its derivative g' are defined when $e^{(b-a+1)x} - 1 \neq 0$ and $e^x - 1 \neq 0$, i.e., when $x \neq 0$. They can be extended by continuity at 0:

$$
\begin{aligned}
g(0, a, b) &= \frac{a + b}{2}, \\
g'(0, a, b) &= \frac{(b - a + 1)^2 - 1}{12}.
\end{aligned}
$$

We have $g'(0, a, b) > 0$ as $b - a + 1 > 1$. In addition,

$$
\begin{aligned}
\lim_{x \to -\infty} g(x, a, b) &= a, \\
\lim_{x \to +\infty} g(x, a, b) &= b, \\
\lim_{x \to -\infty} g'(x, a, b) &= 0, \\
\lim_{x \to +\infty} g'(x, a, b) &= 0.
\end{aligned}
$$

Figure A1 shows a few examples of g and g' for different values of a and b.
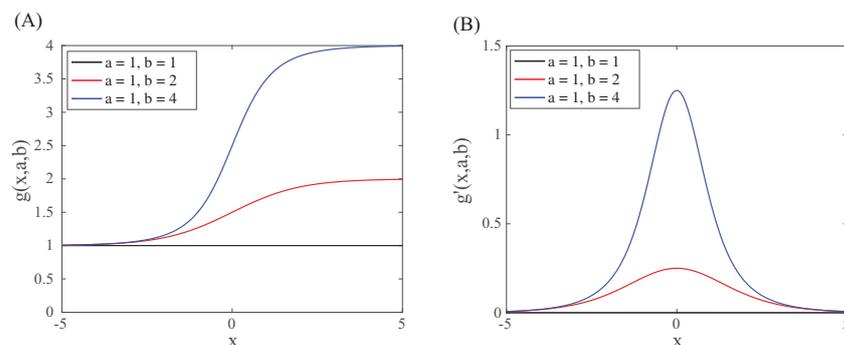

Figure A1. Examples of $g(x, a, b)$ (**A**) and its derivative $g'(x, a, b)$ (**B**) for $a = 1, b = 1$ (black), $a = 1, b = 2$, (red), and $a = 1, b = 4$ (blue).

We now prove property 1 of Theorem A3.

Proof. Let us define $m = b - a + 1$. As $a < b$, we have $m > 1$. Let x be a real number; then,

$$g'(x, a, b) = \quad -\frac{m^2 e^{mx}}{(e^{mx} - 1)^2} + \frac{e^x}{(e^x - 1)^2},$$

which we can rewrite as

$$
\begin{aligned}
&g'(x, a, b) \\
&= \frac{1}{4} \left(-\frac{m^2}{\sinh^2(\frac{mx}{2})} + \frac{1}{\sinh^2(\frac{x}{2})} \right) \\
&= \frac{1}{4} \left(\frac{\sinh^2(\frac{mx}{2}) - m^2 \sinh^2(\frac{x}{2})}{\sinh^2(\frac{mx}{2}) \sinh^2(\frac{x}{2})} \right) \\
&= \frac{1}{4} \left(\frac{(\sinh(\frac{mx}{2}) - m \sinh(\frac{x}{2}))(\sinh(\frac{mx}{2}) + m \sinh(\frac{x}{2}))}{\sinh^2(\frac{mx}{2}) \sinh^2(\frac{x}{2})} \right).
\end{aligned}
$$

From this formulation, it is clear that $g'(x, a, b)$ is even. As $g'(0, a, b)$ is strictly positive (see above), we only need to consider $x > 0$. Let us define $y = x/2$. For $y > 0$, the denominator of $g'(x, a, b)$ and the term $\sinh(my) + m \sinh(y)$ are strictly positive. The only remaining term is $\sinh(my) - m \sinh(y)$, which we call $f(y)$. We have

$$
\begin{aligned}
f'(y) &= m(\cosh(my) - \cosh(y)) \\
&= 2m \sinh((m + 1)y) \sinh((m - 1)y).
\end{aligned}
$$

As $m > 1$, $\sinh((m + 1)y)$ and $\sinh((m - 1)y)$ are strictly positive when $y > 0$: $f(y)$ is monotonically increasing for $y \geq 0$. As $f(0) = 0$, we obtain $f(y) > 0$ for $y > 0$. Therefore $g'(x, a, b) > 0$ for all $x > 0$, and since it is an even function and $g'(0, a, b) > 0$, we find that $\forall x \in \mathbf{R} \quad g'(x, a, b) > 0$. $\square$

Appendix H.3. The Function $l(x, a, b) = G(x, a, b) - xg(x, a, b)$

Recall that

$$l(x, a, b) = G(x, a, b) - xg(x, a, b).$$

We now prove property 2 of Theorem A3.

Proof. We prove first that $l(x, a, b)$ is even. Note that

$$
\begin{aligned}
G(-x, a, b) &= \ln \left(\frac{e^{-ax} - e^{-(b+1)x}}{1 - e^{-x}} \right) \\
&= \ln \left(\frac{e^{(b+1)x} - e^{ax}}{e^x - 1} \right) + \ln \left(\frac{e^x}{e^{(b+1)x} e^{ax}} \right) \\
&= -bx - ax + G(x, a, b),
\end{aligned}
$$

and

$$
\begin{aligned}
g(-x, a, b) &= \frac{(b - a + 1)}{e^{-(b-a+1)x} - 1} + \frac{1}{1 - e^{-x}} + b \\
&= \frac{(b - a + 1)e^{(b-a+1)x}}{1 - e^{(b-a+1)x}} + \frac{e^x}{e^x - 1} + b \\
&= a + b - g(x, a, b).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
l(-x,a,b) &= G(-x,a,b) + xg(-x,a,b) \\
&= G(x,a,b) - xg(x,a,b) \\
&= l(x,a,b).
\end{aligned}
$$

Let now x be a positive real number. Then,

$$
\begin{aligned}
l'(x,a,b) &= G'(x,a,b) - g(x,a,b) - xg'(x,a,b). \\
&= -xg'(x,a,b)
\end{aligned}
$$

We know that $g'(x,a,b)$ is positive for all real numbers x (property 1 of Theorem A3); therefore, $l'(x,a,b)$ is negative on $\mathbb{R}^+$ and $l(x,a,b)$ is monotonic decreasing on this domain. Since:

$$
\begin{aligned}
l(0,a,b) &= \ln(b-a+1), \\
\lim_{x \to +\infty} l(x,a,b) &= 0,
\end{aligned}
$$

we have $0 \leq l(x,a,b) \leq \ln(b-a+1)$ for $x \in \mathbb{R}^+$, and since $l(x,a,b)$ is even, $0 \leq l(x,a,b) \leq \ln(b-a+1)$ for $x \in \mathbb{R}$, which concludes the proof of property 2 of Theorem A3. $\square$

Appendix H.4. The Function $s(x,a,b)$ Is Negative

We now prove the property 3 of A3.

Proof. Recall that

$$
l'(x,a,b) = -xg'(x,a,b).
$$

Therefore,

$$
s(x,a,b) = -x^2 g'(x,a,b).
$$

We know that $g'(x,a,b)$ is positive for all real numbers x (property 1 of Theorem A3); therefore, $s(x,a,b)$ is negative on $\mathbb{R}$. $\square$

References

1. Dell'Amico, M.; Toth, P. Algorithms and codes for dense assignment problems: the state of the art. *Discret. Appl. Math.* **2000**, *100*, 17–48. [CrossRef]
2. Pentico, D.W. Assignment problems: A golden anniversary survey. *Eur. J. Oper. Res.* **2007**, *176*, 774–793. [CrossRef]
3. Burkard, R.; Dell'Amico, M.; Martello, S. *Assignment Problems*; Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, USA, 2009.
4. Dantzig, G. Origins of the simplex method. In *A History of Scientific Computing*; Association for Computing Machinery: New York, NY, USA, 1990; pp. 141–151.
5. Jacobi, C. De investigando ordine systatis aequationum differentialum vulgarium cujuscunque. *J. Reine Angew. Math.* **1890**, *94*, 292–320.
6. Kuhn, H. The Hungarian method for the assignment problem. *Nav. Res. Logist.* **1955**, *2*, 83–97. [CrossRef]
7. Date, K.; Nagi, R. GPU-accelerated Hungarian algorithms for the Linear Assignment Problem. *Parallel Comput.* **2016**, *57*, 52–72. [CrossRef]
8. Lopes, P.A.; Yadav, S.S.; Ilic, A.; Patra, S.K. Fast block distributed CUDA implementation of the Hungarian algorithm. *J. Parallel Distrib. Comput.* **2019**, *130*, 50–62. [CrossRef]
9. Yadav, S.S.; Lopes, P.A.C.; Ilic, A.; Patra, S.K. Hungarian algorithm for subcarrier assignment problem using GPU and CUDA. *Int. J. Commun. Syst.* **2019**, *32*, e3884. [CrossRef]
10. Kumar, A. A modified method for solving the unbalanced assignment problems. *Appl. Math. Comput.* **2006**, *176*, 76–82. [CrossRef]
11. Yadaiah, V.; Haragopal, V. A New Approach of Solving Single Objective Unbalanced Assignment Problem. *Am. J. Oper. Res.* **2016**, *6*, 81–89. [CrossRef]
12. Costa, D.; Hertz, A. Ants can color graphs. *J. Oper. Res. Soc.* **1997**, *48*, 295–305. [CrossRef]

13. Wang, L.; He, Z.; Liu, C.; Chen, Q. Graph based twin cost matrices for unbalanced assignment problem with improved ant colony algorithm. *Results Appl. Math.* **2021**, *12*, 100207. [CrossRef]

14. Betts, N.; Vasko, F.J. Solving the unbalanced assignment problem: Simpler is better. *Am. J. Oper. Res.* **2016**, *6–9*, 296. [CrossRef]

15. Koehl, P.; Orland, H. Fast computation of exact solutions of generic and degenerate assignment problems. *Phys. Rev. E* **2021**, *103*, 042101. [CrossRef] [PubMed]

16. Koehl, P.; Delarue, M.; Orland, H. Physics approach to the variable-mass optimal-transport problem. *Phys. Rev. E* **2021**, *103*, 012113. [CrossRef]

17. Birkhoff, G. Tres observaciones sobre el algebra lineal. *Univ. Nac. Tucuman. Ser. A* **1946**, *5*, 147 –154.

18. Von Neumann, J. A certain zero-sum two-person game equivalent to the optimal assignment problem. *Contrib. Theory Games* **1953**, *2*, 5–12.

19. Dell'Amico, M.; Martello, S. The k-cardinality assignment problem. *Discret. Appl. Math.* **1997**, *76*, 103–121. [CrossRef]

20. Dell'Amico, M.; Lodi, A.; Martello, S. Efficient algorithms and codes for k-cardinality assignment problems. *Discret. Appl. Math.* **2001**, *110*, 25–40. [CrossRef]

21. Volgenant, A. Solving the k-cardinality assignment problem by transformation. *Eur. J. Oper. Res.* **2004**, *157*, 322–331. [CrossRef]

22. Chandrasekaran, R.; Kabadi, S.; Murty, K. Some NP-complete problems in linear programming. *Oper. Res. Lett.* **1982**, *1*, 101–104. [CrossRef]

23. Greenberg, H.J. An analysis of degeneracy. *Nav. Res. Logist. Q.* **1986**, *33*, 635–655. [CrossRef]

24. Megiddo, N.; Chandrasekaran, R. On the ε-perturbation method for avoiding degeneracy. *Oper. Res. Lett.* **1989**, *8*, 305–308. [CrossRef]

25. Koehl, P.; Delarue, M.; Orland, H. Finite temperature optimal transport. *Phys. Rev. E* **2019**, *100*, 013310. [CrossRef] [PubMed]

26. Majumdar, J.; Bhunia, A.K. An alternative approach for unbalanced assignment problem via genetic algorithm. *Appl. Math. Comput.* **2012**, *218*, 6934–6941. [CrossRef]

27. Rabbani, Q.; Khan, A.; Quddoos, A. Modified Hungarian method for unbalanced assignment problem with multiple jobs. *Appl. Math. Comput.* **2019**, *361*, 493–498. [CrossRef]

28. Heist, N.; Hertling, S.; Ringler, D.; Paulheim, H. Knowledge Graphs on the Web—An Overview. *arXiv* **2020**, arXiv:2003.00719.

29. Veličković, P. Everything is connected: Graph neural networks. *Curr. Opin. Struct. Biol.* **2023**, *79*, 102538. [CrossRef] [PubMed]

30. Camacho, D.; Panizo-LLedot, A.; Bello-Orgaz, G.; Gonzalez-Pardo, A.; Cambria, E. The four dimensions of social network analysis: An overview of research methods, applications, and software tools. *Inf. Fusion* **2020**, *63*, 88–120. [CrossRef]

31. Ideker, T. Network genomics. *Ernst Scher. Res. Found. Workshop* **2007**, *61*, 89–115.

32. Pierskalla, W.P. The multidimensional assignment problem. *Oper. Res.* **1968**, *16*, 422–431. [CrossRef]

33. Balas, E.; Saltzman, M.J. Facets of the three-index assignment polytope. *Discret. Appl. Math.* **1989**, *23*, 201–229. [CrossRef]

34. Spieksma, F.C. Multi index assignment problems: complexity, approximation, applications. In *Nonlinear Assignment Problems: Algorithms and Applications*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 1–12.

35. Kosowsky, J.; Yuille, A. The invisible hand algorithm: Solving the assignment problem with statistical physics. *Neural Netw.* **1994**, *7*, 477–490. [CrossRef]

36. Koehl, P. Extreme points of general transportation polytopes. *arXiv* **2024**, arXiv:2404.16791.

37. Mendelsohn, N.S.; Dulmage, A.L. The convex hull of sub-permutation matrices. *Proc. Am. Math. Soc.* **1958**, *9*, 253–254. [CrossRef]

38. Brualdi, R.A.; Lee, G.M. On the truncated assignment polytope. *Linear Algebra Its Appl.* **1978**, *19*, 33–62. [CrossRef]