1992

Linear Assignment Problem Revisited: Average (1 + 1/e)-out Degree Bipartite Graph has a Perfect Match

Michael Karoński

Wojciech Szpankowski
Purdue University, spa@cs.purdue.edu

Report Number:
92-038

LINEAR ASSIGNMENT PROBLEM REVISITED:
Average $(1 + 1/e)$-out Degree Bipartite
Graph Has a Perfect Matching

Michal Karonski
Wojciech Szpankowski

CSD-TR-92-038
June 1992

LINEAR ASSIGNMENT PROBLEM REVISITED:
Average $(1+1/e)$-out Degree Bipartite Graph Has a Perfect Matching

Working Paper: June 23, 1992

Michal Karoński*
Department of Discrete Mathematics
Adam Mickiewicz University
60-769 Poznań
Poland

Wojciech Szpankowski†
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.

Abstract

We have three results to report in this paper. First of all, we improve a result of Walkup who reported in 1979 that 2-out regular bipartite digraph has almost surely a perfect matching. We prove that a bipartite digraph (constructed in a special way) with the average out-degree of $1+1/e \approx 1.37$ has almost surely a perfect matching. This value is only a little higher than the critical average out-degree $e/2 \approx 1.36$ below which almost surely a perfect matching does not exist. Secondly, we use our finding to establish a new constructive upper bound for the *linear assignment problem* (LAP). Namely, when the weights in the LAP are independently and identically uniformly distributed in the interval $[0,1]$, we prove that the expected weight of the LAP is bounded above by $2+1/e+O(1/n) \approx 2.37+O(1/n)$. Finally, we present also a simple iterative scheme that produces a sequence of lower bounds for the LAP. In particular, the first iteration gives Lazarus' lower bound of $1+1/e+O(1/n)$, while the third iteration returns the value of $1.43+O(1/n)$. Lower and upper bounds are derived in a unified manner by various applications of the random allocation model.

*This research was primary done while the author was visiting Department of Computer Science, Purdue University. Support was provided in part by Grant KBN 2-1087-91-01.

†This author was supported in part by NSF Grant CCR-8900305, and INT-8912631, by Grant AFOSR-90-0107, and in part by Grant R01 LM05118 from the National Library of Medicine.

1. INTRODUCTION

The Linear Assignment Problem (LAP) finds a minimum weight perfect matching in a complete bipartite graph. More precisely, let $\mathbf{W} = \{w_{ij}\}_{i,j=1}^{n}$ be $n \times n$ matrix of real numbers that we further call *weights*. In LAP we seek a permutation $\sigma(\cdot)$ of $\{1, \ldots, n\}$ such that it minimizes the sum $Z_{\min} = \min_\sigma\{\sum_{i=1}^{n} w_{i,\sigma(i)}\}$, where the minimum is taken over all $n!$ possible permutations. This is a classical problem that has a variety of applications (cf. Lawler [14], Zuker [24]). It is conveniently viewed as the problem of finding a minimum weight perfect matching in a complete bipartite graph. Then, several efficient algorithms are readily available (cf. [18], [14]).

In this paper, we shall consider LAP in a probabilistic framework under the assumption that the weights w_{ij} are *independently and identically distributed* (i.i.d). random variables with the distribution function $F(\cdot)$. We concentrate, however, on uniformly distributed weights on $[0, 1]$, and show how this easily can be extended to the general case.

The linear assignment problem was addressed in several papers. Borovkov [4] in 1962 analyzed a greedy heuristic to the LAP. For the distribution function of weights of the form $F(x) = (\alpha x)^\beta(1 + o(1))$ for $\beta > 1$ (e.g., a composition of several uniform distributions) he proved that almost surely (a.s.) $Z_{\min} = O(n^{1-1/\beta})$. Szpankowski in [21] observed that for weights with distribution that have exponential *negative* tail,the greedy algorithm is asymptotically optimal (e.g., for negative exponential weights $Z_{\min} \sim n \log n$ (pr.); for normally distributed weights $Z_{\min} \sim \sqrt{2n \log n}$ (pr.)). Kurtzberg [13] showed that for the uniform distribution $U(0, 1)$ of weights the average value $EZ_{\min}$ is bounded as $1 + O(1/n) \leq EZ_{\min} \leq \log n$. It was not untill the Walkup paper [23] (cf. [22]) that a constant upper bound for $EZ_{\min}$ was found. More specifically, Walkup proved the almost sure existence of a perfect matching in a 2-out bipartite digraph (cf. [23]), and using this he showed that $EZ_{\min} < 3 + O(1/n)$ for the uniform distribution of weights (cf. [22]). This was subsequently reduced by Karp [10] to $2 + O(1/n)$. Karp's argument is based on some properties of dual pair of linear programming problems. On the other hand, in 1979 Lazarus showed in an unpublished B.A. thesis that $EZ_{\min} > 1 + 1/e + O(1/n) \approx 1.37$. Finally, in 1985 Mézard and Parisi [17] gave non-rigorous arguments (based on the mean field theory of spin glasses) to support a conjecture that $EZ_{\min} = \zeta(2) + O(1/n) = \pi^2/6 + O(1/n) \approx 1.64$. This was also confirmed by simulations. A rigorous proof of Mézard and Parisi seems to be very hard, if possible at all (cf. [19]).

Neither Walkup's approach nor Karp's proof were algorithmic. A minimum matching in a weighted complete bipartite graph can be found in $O(n^3)$ steps ([14]) by the *Hungarian*

Method. This is, however, quite complicated to implement, as assessed by Lawler [14]. On the other hand, a perfect matching in an unweighted general graph can be found in $O(m\sqrt{n})$ where m is the number of edges (cf. Micali and Vazirani [18]). In view of this, several heuristics have been proposed to solve efficiently the linear assignment problem. For a survey see Avis [1]. Recently, Avis and Lai [2] proposed an $O(n^2)$ heuristic for LAP which returns a matching of the expected weight smaller than $6 + O(1/n)$. This was further improved by Karp, Rinnooy Kan and Vohra [11] whose $O(n^2)$ heuristic generates a matching with the expected weight less than $3 + O(n^{-\epsilon})$.

Our contribution is twofold. First of all, we deal with a perfect matching in a sparse bipartite digraph. For such a digraph we improve the Walkup result by showing that average $1+1/e \approx 1.37$ degree special bipartite digraphs have almost surely a perfect matching. These bipartite digraphs are constructed in such a way that zero in-degree vertices are eliminated. In other words, a our bipartite digraph with $2n(1+1/e) \approx 2.74n$ arcs on the average has a perfect matching with high probability. Such a matching, if exists, can be found in $O(n^{1.5})$ (a.s.) steps (cf. [18]). Our second result concerns the linear assignment problem. We propose an $O(n^2)$ heuristic that returns a minimum perfect matching with the expected weight $EZ_{\min} < 2 + 1/e + O(\log n/n) \approx 2.37$. This improves the Walkup best estimate. We also elaborate on the lower bound for LAP. We present an iterative scheme that provides a sequence of lower bounds. For example, the first iteration returns Lazarus' bound $EZ_{\min} > 1 + 1/e + O(1/e) \approx 1.37$, while the third iteration gives $EZ_{\min} > 1.43 + O(1/n)$. Our method is extremely simple. Finally, we extend our results to general distribution of weights.

It should be stressed that our method of constructing the upper bound for LAP could lead to several new algorithms and estimates for various combinatorial problems; e.g., the traveling salesman problem, and so forth.

2. MAIN RESULTS

In this section, we present our main results. As indicated before, we have two kinds of results. The first deals with the existence of a perfect matching in a sparse bipartite graph. This finding is further used to construct a solution for the linear assignment problem (LAP).

We start with some notation. Let $B = (V_0, V_1, E)$ be a bipartite graph with vertex sets V_0 and V_1 each of cardinality n, and edge set E of cardinality m. A matching $\mathcal{M}$ in the graph B is a set of edge such that no two edge in $\mathcal{M}$ are incident to the same vertex of B. A matching $\mathcal{M}$ is perfect if every vertex of B is adjacent to some edge in $\mathcal{M}$. In the rest of the paper, we consider only perfect matchings, and we denote them by $\mathcal{M}$. Finally, we associate with every edge $\{i,j\}$ of B a weight w_{ij}. A minimum weight perfect matching for the graph is a perfect matching for which the sum of weights is minimum. The linear assignment problem (LAP) can be alternatively posed as finding a minimum weight perfect matching.

Walkup [23] investigated a *random* out-regular directed bipartite graph $\vec{B}_d(n)$ with $|V_0| = |V_1| = n$ in which every vertex has d neighbors selected independently and equally likely among n possible candidates (i.e., every vertex has out-degree d). He proved that the probability of a perfect matching $P(\mathcal{M}) = \Pr\{\text{there exists a perfect matching in } \vec{B}_d(n)\}$ tends to 1 as $n \to \infty$ provided $d \geq 2$. Furthermore, Walkup also noticed that $P(\mathcal{M}) \leq 3\sqrt{n}(2d/e)^n$, that is, almost surely the graph $\vec{B}_d(n)$ does *not* have a perfect matching if $d = 1$.

We extend Walkup's idea by allowing some vertices to have degree one and others to have degree two. So, we are in position to talk about the *average* degree of a vertex, which we also denote as d. We consider only $1 < d \leq 2$ since by the Walkup upper bound we know that there is no perfect matching for the average degree $d < e/2 \approx 1.36$. Our work is motivated by this fact, and our aim is to uncover the behaviour of the digraph for $e/2 < d < 2$. In short, we wonder whether a digraph with the average degree $d < 2$ has a perfect matching or not.

We need, however, a special construction of a random digraph to assure (a.s.) existence of a perfect matching. We start with a generation of a random 1-our regular bipartite digraph $\vec{B}_1(n)$. More precisely, for every vertex from V_0 we draw an arc to a vertex selected at random from V_1, and *vice versa*, for every vertex from V_1 we select an arc connecting a random vertex in V_0. Next, we color red all unselected vertices both in V_0 and V_1, that is, vertices with in-degree zero. Suppose that $V_0' \subset V_0$ and $V_1' \subset V_1$ are sets of red vertices in V_0 and V_1, respectively. Now, for every vertex in V_0' (V_1') we choose its second neighbor ar random in V_1 (V_0). In the resulting digraph, some vertices have out-degree one and others out-degree two, but all vertices with in-degree zero are eliminated. We denote such a digraph as $\vec{B}(n, M)$, where M is a random variable representing the total numbers of extra arcs added in the second phase of our construction. An example of $\vec{B}(3, 2)$ is shown in Figure 1.

Note that a similar idea to force the minimum degree of a vertex (in an undirected bipartite graph) to exceed a given value was also discussed in Bollobás [3]. On the other hand, Jaworski and Luczak in their recent paper [9] studied some properties of random uniform digraph processes with unrestricted second phase of arcs additions.

It is clear that the digraph $\vec{B}_1(n)$ resulting from the first phase of our construction is really a representation of a *random mapping* $T = T_1 \cup T_2$ such that $T_0 : V_1 \to V_0$

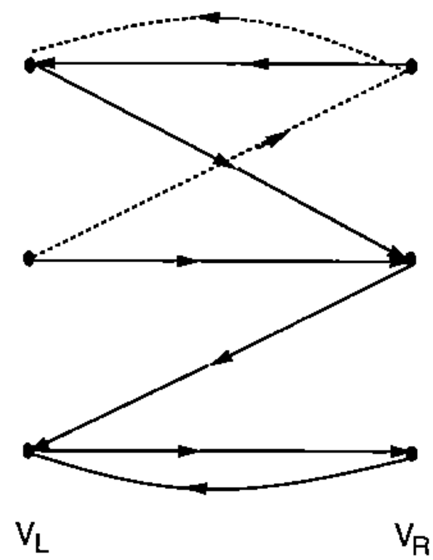

Figure 1: Example of the restricted random digraph $\vec{B}(3,2)$

and $T_1 : V_0 \rightarrow V_1$. Furthermore, the random mappings T_0 and T_1 can be viewed as a representation of a random allocation of n balls into n urns. Partitioning the number of additional arcs M as $M = M_0 + M_1$, where M_0 (M_1) corresponds to the number of in-degree zero vertices in V_0 (V_1), we note that M_0 and M_1 coincide with the number of *empty* urns in the random allocation model (called also further urns model). note that the urns model is another representation for the random mapping. A standard analysis of M_0 and M_1 through the urns model reveals that both random variables satisfy $M_0 \sim M_1 \sim n/e$ (pr.) (cf. Kolchin *et al.* [12]). Hence, the total number of arcs in $\vec{B}(n,M)$ becomes $m = 2n(1 + /1e) \approx 2.74n$ (a.s.), and the average degree of a vetex is $d = 1 + 1/e \approx 1.36$.

Our first main result concerns the existence of a perfect matching $\mathcal{M}$ in $\vec{B}(n,M)$. The proof is provided in the next section.

Theorem 1. PERFECT MATCHING.
Let $P(\mathcal{M})$ be the probability of a matching in the restricted random digraph $\vec{B}(n,M)$. Then,

$$1 - P(\mathcal{M}) = O(n^{-1}) , \tag{1}$$

that is, the graph $\vec{B}(n,M)$ almost surely has a perfect matching. ■

Theorem 1 can be used to design an $O(n^2)$ (a.s.) algorithm that solves the *linear assignment problem* (LAP). To recall, for a given matrix of weights $\mathbf{W} = \{w_{ij}\}_{i,j=1}^{n}$ we search for a permutation $\sigma(\cdot)$ such that the sum $\sum_{i=1}^{n} w_{i,\sigma(i)}$ is minimized. By $Z_{\min}$ we denote the minimum value of the sum returned by the LAP. Clearly, a solution to the LAP can be alternatively found as a perfect matching in a complete bipartite graph. Such a solution, however, can be found in $O(n^3)$ steps, and the algorithm is rather complicated (cf. [14]).

We solve the linear assignment problem in a probabilistic framework. That is, we adopt the following assumption:

(A) The weights w_{ij} are independently and identically distributed according to the distribution function $F(x) = \Pr\{w_{ij} < x\}$. To avoid confusions, we denote *random* weights as W_{ij}.

It turns out that a solution to our problem critically depends on a solution to the LAP with uniformly distributed weights. Therefore, we shall first investigate the LAP under the following assumption, which is a modification of (A):

(A′) The weights W_{ij} are independently and uniformly distributed on the interval $[0,1]$, that is, $W_{ij} \sim U(0,1)$ where $U(0,1)$ represents a generic random variable that is uniformly distributed on $[0,1]$.

Hereafter, we analyze LAP under the assumption (A′) unless otherwise explicitly stated.

Now, we are ready to present our algorithm to solve the LAP. We first discuss it informally in a sequence of steps that are refinments of Walkup ideas:

Step 1. Generate two matrices $\mathbf{X} = \{X_{ij}\}_{i,j=1}^{n}$ and $\mathbf{Y} = \{Y_{ij}\}_{i,j=1}^{n}$ such that for every i and j the random variables X_{ij} and Y_{ij} are independently and identically distributed according to $F(w_{ij}) = 1 - \sqrt{1 - w_{ij}}$. The new weights X_{ij} and Y_{ij} are either generated according to *the inverse transform* method or as in Karp *et al.* [11].

Step 2. Select the smallest elements in every *row* of $\mathbf{X}$ and every *column* of $\mathbf{Y}$. This corresponds to assigning weights to the random mappings in our bipartite digraph $\vec{B}(n,M)$ (e.g., the smallest row elements of $\mathbf{X}$ represent weights of arcs from $V_0 = \{1,\ldots,n\}$ to $V_1 = \{1,\ldots,n\}$, while smallest column elements of $\mathbf{Y}$ are weights of arcs from V_1 to V_0).

Step 3. Let V_0' and V_1' denote subsets of *rows* of $\mathbf{Y}$ and *columns* of $\mathbf{X}$ respectively that do *not* contain even one smallest element selected in Step 2. Then, for every $v_L \in V_0'$ select the second smallest weight in the row v_L of the matrix $\mathbf{X}$; and for every $v_R \in V_1'$ select the second smallest weight in the column of $\mathbf{Y}$. This step corresponds to generating M arcs in the restricted bipartite digraph $\vec{B}(n,M)$.

Step 4. Find in the restricted bipartite digraph *without weights* $\vec{B}(n,M)$ constructed above a perfect matching in $O(n\sqrt{n})$ (a.s.) steps by Micali and Vazirani algorithm [18].

Step 5. If the graph $\vec{B}(n, M)$ does *not* have a perfect matching (which happens with the probability $O(1/n)$), then use an $O(n^2)$ *greedy algorithm* applied to the original weight matrix $\mathbf{W}$ to find a perfect matching that approximates the minimum perfect matching.

Our second main result, which improves the Walkup upper bound for the LAP problem, is presented in the theorem below.

Theorem 2. UPPER BOUND.
Consider the linear assignment problem under assumption (A') . Then, the above algorithm solves the LAP problem in $O(n^2)$ (a.s.) steps, and it returns the minimum value $Z_{\min}$ such that

$$EZ_{\min} \leq 2 + \frac{1}{e} + O\left(\frac{\log n}{n}\right) . \tag{2}$$

Furthermore, consider now the LAP problem under the general assumption (A). Let $E \max\{W_1, \ldots, W_n\} \leq \alpha(n)$ where the random variables W_i are distributed according to the general distribution function $F(\cdot)$ as specified in (A). Then, our algorithm returns $Z_{\min}$ that satisfies

$$EZ_{\min} \leq n\alpha(n/2) + \frac{n}{e} \cdot \frac{\alpha(n/4) - \alpha(n/2)}{2} + o(n\alpha(n)) . \tag{3}$$

The above estimate becomes (2) when $F(x) = x$, hence $\alpha(n) = 1/n$.

Proof. This proof is a repetition of the Walkup ideas and a simple-minded analysis of the greedy heuristic. Indeed, we first split the matrix $\mathbf{W}$ into two independent matrices $\mathbf{Y}$ and $\mathbf{X}$, as discussed above. Walkup [22] suggested to define these matrices such that for every pair (i, j)

$$W_{ij} = \min\{Y_{ij}, X_{ij}\} . \tag{4}$$

In fact, many other relationships between these two matrices do the job equally well. The main point is to observe that the k-th order statistics $W_{i,(k)}$ and $Y_{i,(k)}$ of W_{ij} and Y_{ij} for fixed i are related as follows (cf. [22])

$$EY_{i,(k)} \leq 2EW_{i,(k)} , \tag{5}$$

and for uniformly distributed weights on $[0, 1]$ the above becomes $EY_{i,(k)} \leq 2k/(n+1)$.

The next step is to construct the restricted random graph $\vec{B}(n, M)$ such that the arcs from V_0 to V_1 represent the smallest and/or the second smallest weight of the matrix $\mathbf{Y}$, while the weights for arcs from V_1 to V_0 represent the smallest and/or the second smallest elements in the matrix $\mathbf{X}$. Then by Theorem 1, with probability $1 - O(1/n)$ the graph $\vec{B}(n, M)$ has a perfect matching, and by (5)

$$EZ_{\min} \leq (n - n/e) \cdot 2/n + n/e \cdot 3/n + O(1/n) = 2 + 1/e + O(1/n) .$$

If, however, the graph $\vec{B}(n, M)$ does not possess a perfect matching, then we apply a greedy algorithm. Such an algorithm works as follows: we select the smallest element in the first column, and delete all elements from the first column and the row into which the smallest element has fallen. We repeat the procedure for the second, third , ..., last column. It is easy to observe that such a greedy procedure costs $EZ_{gr} = \log n + O(1)$. Since $\vec{B}(n, M)$ fails to have a perfect matching with probability $O(1/n)$, we finally obtain the estimate from our Theorem 2. ∎

We now present the main idea of the lower bound. We consider a sequence of the *urns* (i.e., random allocation) schemes. Think of balls and urns as rows and columns of the matrix **W**, respectively. Selecting the smallest element (or the k-th smallest element) in a row corresponds to throwing a ball randomly and equally likely to the urn that coincides with the column selected by this smallest element. More precisely, in the first iteration, we select only the smallest values in every row of the matrix **W**. It is very likely that several smallest elements fall into the same column, and some columns will be empty. But in terms of the urns model, this corresponds to to the number of urns with at least two balls. To express our ideas more precisely, we introduce some new notation. Consider $N = n$ urns which receive in the k-th iteration n_k balls, where $n_1 = n$ and n_k is defined in sequel. Let also

$N_0(n_k)$, $N_1(n_k)$ – be the number of urns with exactly zero and one ball respectively in the k-th iteration,

$N_2(n_k)$ – be the number of urns with more than one ball in the k-th iteration.

Returning to our lower bound, we note that $N_0(n_1)$ is the number of empty columns and $N_2(n_1)$ represents the number of columns with at least two smallest values, where $n_1 = n$. Note also that $N_2(n_1) = n - N_0(n_1) - N_1(n_1)$. If $N_2(n_1) > 0$, then the entries chosen in the first iteration do not constitute a permutation. Nevertheless, we can obtain a lower bound for the objective function $Z_{\min}$ of LAP. Indeed, let $U_{(k)}$ be the k-th order statistic of n uniformly distributed random variables $U(0,1)$, where $U_{(1)} = \min_{1 \le i \le n}\{U_i\}$ and $U_{(n)} = \max_{1 \le i \le n}\{U_i\}$. Define

$$Z_1 = (N_1(n_1) + N_2(n_1))U_{(1)} + (n_1 - N_1(n_1) - N_2(n_1))U_{(2)}$$

It should be clear that $Z_1 \le_{st} Z_{\min}$, where $\le_{st}$ means stochastically smaller [20], that is, $X \le_{st} Y$ if and only if $\Pr\{X > x\} \le \Pr\{Y > x\}$. In particular, $EZ_1 \le EZ_{\min}$, and EZ_1 is easy to estimate.

In the second iteration, we consider only those positions (balls) that fall into columns already occupied by others. The number of those elements (balls) is equal to $n_2 = n_1 -$

$N_1(n_1) - N_2(n_1)$. We throw again these balls into n urns which corresponds to selecting second smallest element (and those who again fall into the same column will select the third smallest element). Define

$$Z_2 = (N_1(n_1) + N_2(n_1))U_{(1)} + (N_1(n_2) + N_2(n_2))U_{(2)} + (n_2 - N_1(n_2) - N_2(n_2))U_{(3)} \quad (6)$$

It is easy to see that $Z_1 \leq_{\rm st} Z_2 \leq_{\rm st} Z_{\min}$, hence we shall obtain a better lower bound.

In general, in the k-th iteration we throw n_k balls into n urns where

$$n_k = n_{k-1} - N_1(n_{k-1}) - N_2(n_{k-1}) = n_{k-1} - n + N_0(n_{k-1}) \ . \quad (7)$$

We also define Z_k as

$$Z_k = \sum_{i=1}^{k}(N_1(n_i) + N_2(n_i))U_{(i)} + n_{k+1}U_{(k+1)} \ . \quad (8)$$

Our third main result can be summarized as follows.

Theorem 3. LOWER BOUND.
Under the assumption (A′), we have the following sequence of lower bounds on $EZ_{\min}$ for any integer k

$$EZ_k \leq EZ_{k+1} \leq EZ_{\min} \ , \quad (9)$$

where the quantities $N_l(n_i)$ for $l = 0, 1, 2$ appearing in (8) can be estimated from the urns model as follows

$$N_0(n_i) = ne^{-\alpha_i} \qquad (a.s.) \qquad N_1(n_i) = n\alpha_i e^{-\alpha_i} \ , \quad (10)$$

and $N_2(n_i) = n - N_0(n_i) - N_1(n_i)$, with $\alpha_i = n_i/n$ where n_i is defined in (7). In particular,

$$EZ_k = \sum_{i=1}^{k} i(1 - e^{-\alpha_i}) + (k+1)(\alpha_k + e^{-\alpha_k} - 1) + O(1/n) \ . \quad (11)$$

For example,

$$EZ_{\min} \geq EZ_1 = 1 + \frac{1}{e} + O(1/n) \approx 1.368 \quad (12)$$

$$EZ_{\min} \geq EZ_2 = \frac{2}{e} + \frac{1}{e^{1/e}}O(1/n) \approx 1.428 \quad (13)$$

$$EZ_{\min} \geq EZ_3 = \frac{3}{e} + \frac{2}{e^{1/e}} + \exp(-1/e - 1/e^{1/e} - 1) - 2 + O(1/n) \approx 1.430 \ . \quad (14)$$

The first lower bound coincides with the Lazarus bound.

Proof. We prove first that the sequence of estimates Z_k defined in (8) is a nondecreasing sequence. Indeed, by (8) and (7) we have

$$\begin{aligned} Z_k &= Z_{k-1} - n_k U_{(k)} + (B_1(n_k) + B_2(n_k))U_{(k)} + n_{k+1} U_{(k+1)} \\ &= Z_{k-1} + n_{k+1}(U_{(k+1)} - U_{(k)}) \geq Z_{k-1} \ , \end{aligned} \tag{15}$$

where the last inequality is a a simple consequence of $U_{(k+1)} \geq U_{(k)}$.

To prove that $EZ_k \leq_{st} EZ_{\min}$ we use the sample path theorem [20]. For this we note that starting from the iteration $k = 0$, in which we take all the smallest elements in every column, we add in the next iteration the smallest increment that does not yet produce a permutation, that is, a feasible solution to the LAP. Note that $Z_{k-1} = Z_k$ when $n_{k+1} = 0$, that is, when there is no urn with two or more balls. This observation completes the proof of Theorem 3. ∎

It should be noted that the above sequence of lower bounds works also under the general assumption (A) provided one replaces the i-th order statistic $U_{(i)}$ in (8) by the i-th order statistic of n random variables distributed according to general distribution function $F(\cdot)$.

3. ANALYSIS OF THE PERFECT MATCHING RESULT

This section provides a proof of Theorem 1. It turns out that it can be established along the lines of the Walkup proof [23]. In the proof we often use the random allocation paradigm. The reader is referred to the book of Kolchin *et al.* [12] for a detailed account of the theory of random allocations.

In order to prove the existence with high probability of a perfect matching in a random digraph $\vec{B}(n, M)$, we need to introduce a probabilistic framework for our digraphs. Let $\vec{\mathcal{G}}$ be a family of all bipartite directed graphs $\vec{G} = (V_0, V_1, \vec{E})$ with $|V_0| = V_1| = n$ such that each vertex has out-degree equal to either one or two but no vertex of in-degree greater than zero has out-degree zero. Then, our graph $\vec{B}(n, M)$ introduced above can be viewed as an element of $\vec{\mathcal{G}}$ selected randomly from this family.

Denote by G the underlying simple (undirected) graph of $\vec{G} \in \vec{\mathcal{G}}$. Furthermore, let $\vec{\Gamma}(A)$ and $\Gamma(A)$ denote the set of neighbours of $A \subset V_i$ $(i = 0, 1)$ in $\vec{G}$ and G respectively. Note that if G has no perfect matching, then by Hall's theorem (cf. [15]) there exists a set $A \subset V_i$ such that $|\Gamma(A)| < |A|$. This implies that in $\vec{G}$ we also have $\vec{\Gamma}(A) \subset \Gamma(A)$. Moreover, all arcs emanating from $B = V_{1-i} - \Gamma(A)$ point out to the vertices in the set $A' = V_i - A$, i.e., $\vec{\Gamma}(B) \subset A'$.

Definition 1. A pair (A, B) is called a *blocking pair* if $A \subset V_i$ and $B \subset V_{1-i}$ for $i = 0, 1$ such that $\vec{\Gamma}(A) \subset B'$ and $\vec{\Gamma}(B) \subset A'$ where $A' = V_i - A$ and $B' = V_{1-i} - B$ with $|B'| = |A| - 1$. A blocking pair (A, B) with $|A| = a$ is called a blocking a-pair. □

It should be clear from the above definition that we can restrict our analysis to $2 \leq a \leq \lfloor (n+1)/2 \rfloor$. Then, Hall's criterion says that a directed bipartite graph has a perfect matching if *there exists no blocking pair* [15]. For our purpose, however, we need a refinment of Hall's theorem, as already discussed in Walkup [23]. We define a *critical* blocking pair as follows.

Definition 2. A blocking a-pair (A, B) is *critical* if there is no $(a-1)$-pair $(A - \{s\}, B \cup \{t\})$ which is blocking for any $s \in A$ and $t \in B'$. □

We now can estimate the probability $1 - P(\mathcal{M})$ that the restricted random graph $\vec{B}(n, M)$ does not possess a perfect matching. Let $\varepsilon > 0$ be an arbitrary small positive number. Then,

$$
\begin{aligned}
1 - P(\mathcal{M}) &= \sum_{k=0}^{2n} \Pr\{\vec{B}(n,M) \text{ has no } \mathcal{M} \mid M = k\} \Pr\{M = k\} \\
&\leq \sum_{(1-\varepsilon)2n/e \leq k \leq (1+\varepsilon)2n/e} \Pr\{M = k\} \qquad (16) \\
&+ \sum_{(1-\varepsilon)2n/e < k < (1+\varepsilon)2n/e} \Pr\{\vec{B}(n,M) \text{ has no } \mathcal{M} \mid M = k\} \Pr\{M = k\} \\
&\leq \delta + \Pr\{\vec{B}(n,M) \text{ has no } \mathcal{M} \text{ and } 2n(1-\varepsilon)/e < M < (1+\varepsilon)2n/e\}
\end{aligned}
$$

where $\delta = O(1/n)$ (cf. Kolchin *et al.* [12]) can be made arbitrary small as $n \to \infty$. Certainly, the existence of a blocking pair implies the existance of at least one critical blocking a-pair for some a. This, together with the above, lead to our next estimate

$$
\begin{aligned}
1 - P(\mathcal{M}) &\leq \delta + 2 \sum_{a=2}^{\lfloor (n+1)/2 \rfloor} \Pr\{\vec{B}(n,M) \text{ has a critical blocking } a\text{-pair}, \\
&\qquad\qquad \text{provided } 2n(1-\varepsilon)/e < M < (1+\varepsilon)2n/e\} \\
&\leq \delta + 2 \sum_{a=2}^{\lfloor (n+1)/2 \rfloor} \beta_n(a) \gamma_n(a, K_0) , \qquad (17)
\end{aligned}
$$

where $\beta_n(a)$ represents the expected number of blocking a-pairs, and $\gamma_n(a, K_0)$ denotes the conditional probability that a blocking a-pair is critical blocking pair provided $M \in K_0 = (2n(1-\varepsilon)/e \, , \, 2n(1+\varepsilon)/e)$.

In sequel we estimate $\beta_n(a)$ and $\gamma_n(a, K_0)$. We start with $\beta_n(a)$. The following lemma is crucial.

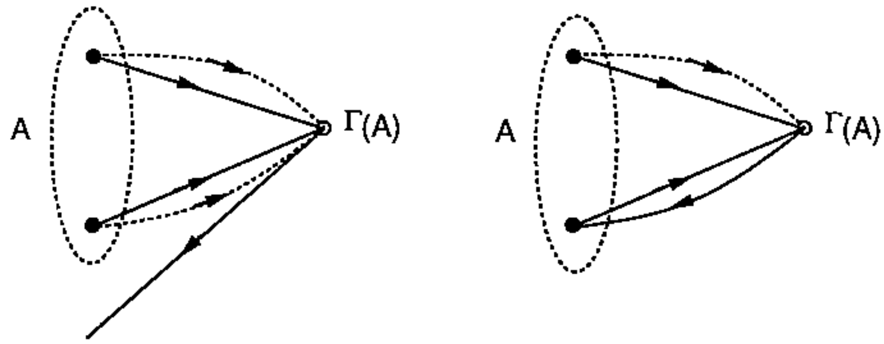

Figure 2: Illustration for the calculation of $\beta(2)$

Lemma 4. *Suppose that $\vec{G} \in \vec{\mathcal{G}}$ and (A, B) is a blocking a-pair in $\vec{G}$. Then, the subgraph induced by (A, b') and (A', B) have at least $2a$ and $2(n-a+1)$ arcs, respectively. Clearly, this is particularly true for the random digraph $\vec{B}(n, M)$.*

Proof. For the reader convenience, we first discuss the case $a = 2$ (see Figure 1). Then $|B'| = 1$ and $\vec{\Gamma}(A) = B' = v$. Since (A, B) is blocking, there is no arc emanating from B and pointing out to a vertex in A. By definition of $\vec{G}$ either both vertices from A have out-degree two (when the vertex v has a neighbor in A') or one of the vertices of A have out-degree two while the second one is a neighbor of v, and therefore has in-degree and out-degree equal to one. In both cases, (A, B') induces a subgraph with four arcs. Similarly, in a subgraph induced by (A', B) there are at least $n-1$ arcs from B to A'. In the worst case, when $\vec{\Gamma}(B) = A'$ each vertex has out-degree one, which in turn can prevent at most $n-2$ vertices of B to have out-degree two (since (A, B) is blocking there is no arc leading from A to B). Therefore, the subgraph induced by $A', B)$ has at least $(n-1) + (n-2) + 1 = 2(n-1)$ arcs.

The case $a > 2$ can be argued in a similar manner. Indeed, if $|A| = a$ and there are k arcs ($0 \leq k \leq a-1$) from B' to A, then at least $2a - k$ arcs must go from A to B'. Hence, (A, B') induces at least $2a$ arcs, and as in the case of $a = 2$ we can easily see that (A', B) induces at least $2(n-a+1)$ arcs. ∎

Using Lemma 4, we immediately obtain the following estimate on $\beta_n(a)$

$$\beta_n(a) \leq \binom{n}{a}\binom{n}{a-1}\left(\frac{a}{n}\right)^{2a}\left(1-\frac{a}{n}\right)^{2(n-a+1)} . \tag{18}$$

Applying Stirling's inequality in the form (cf. [23])

$$n^{n+1/2}\sqrt{2\pi}e^{-n} \leq n! \leq n^{n+1/2}\sqrt{2\pi}e^{-n}(1 + 1/(12n-1)) ,$$

we finally arrive at

$$\beta_n(a) \leq \frac{c}{n-a+1} , \tag{19}$$

where the constant $c \leq 144/(242 \cdot \pi) \approx 0.19$.

To estimate the probability $\gamma_n(a, K_0)$ we need the following observation already suggested by Walkup [23].

Lemma 5. (Walkup [23]). *If a blocking pair is critical, then $B' = B^*$, where $B' = V_{1-i} - B$ and $B^* = \bar{\Gamma}(A) \cup B''$ and B'' consists of all vertices of B' that have exactly two neighbours in A.* ■

From the above lemma, we can bound

$$\gamma_n(a, K_0) \leq \Pr\{B' = B^*\} \leq (\Pr\{v \in B^*\})^a \leq (1 - P_1^a P_2)^{a-1} ,$$

where

$$P_1 = \Pr\{v \notin \Gamma(A)\} = \frac{a-2}{a-1}\left(1 - \frac{1+\varepsilon}{e(a-1)}\right) ,$$

and

$$P_2 = \Pr\{|\bar{\Gamma}(v)| = 2 \ , \ v \in B'\} = 1 - \frac{1+\varepsilon}{e}\frac{a^2}{n^2} .$$

Putting everything together, we finally obtain

$$\begin{aligned}
\gamma_n(a, K_0) &\leq 2 \sum_{a=3}^{\lfloor (n+1)/2 \rfloor} \left(1 - \left(\left(\frac{a-2}{a-1}\right)\left(1 - \frac{1+\varepsilon}{e(a-1)}\right)\right)^a \left(1 - \frac{1+\varepsilon}{e}\frac{a^2}{n^2}\right)\right)^{a-1} \\
&\leq 2 \sum_{a=3}^{\lfloor (n+1)/2 \rfloor} (1 - 125/1728)^a + O(1/n) < 28 + O(1/n) . \qquad (20)
\end{aligned}$$

Finally, the above estimates (19) and (20) lead to

$$1 - P(\mathcal{M}) \leq \frac{C}{n} , \tag{21}$$

where the constant $C \leq 5.4$. This completes the proof of Theorem 1.

4. CONCLUSIONS AND FURTHER RESEARCH

In this paper we constructed a new random bipartite graph called restricted random out-regular bipartite graph that assures the existence of a perfect matching. One interesting feature of this graph is its fractional degree of a vertex. This simple trick allow us to improve several existing estimates for the linear assignment problem. In particular, we propose an heuristic for the LAP that in $O(n^2)$ (a.s.) steps returns the value $Z_{\min}$ such that $1.43 \leq EZ_{\min} \leq 2 + 1/e + O(1/n) \approx 2.368$.

The idea of our construction is *not* restricted to bipartite graphs and to the linear assignment problem. In particular, we believe we can provide the best up-to-date estimate for the existence of a hamiltonian cycle and the traveling salesman problem. In recent paper Frieze and Luczak [8] proved that $4+\varepsilon$-degree digraph contains a hamiltonian cycle. This result was established by the authors of [8] by constructing two perfect matchings each using 2-out-regular random graph, and then by using additional $\log n$ edges to patch these two matchings to create a hamiltonian cycle. This result easily implies an upper bound for the traveling salesman problem with uniform weights, namely, $EZ_{\min} \leq 5$, where $Z_{\min}$ is the objective function for the traveling salesman problem. Our approach, if proved correct, would lead to the upper bound $3+2/e \approx 3.72$ for the traveling salesman problem. This problem will be attacked in a forthcoming paper.

Finally, there are interesting generalization of the linear assignment problem. Consider an ℓ-dimensional assignment problem in which an ℓ dimensional matrix $n\times\ldots\times n$ of weights $\mathbf{W}$ is given and one is asked to select n elments of $\mathbf{W}$ such that any two do not lie on the same coordinate, and the sum of them is minimal. It is plausible that for uniformly distributed weights the optimal value $Z_{\min}$ satisfies $EZ_{\min} = O(1/n^{\ell-2})$. This is also true for the one-dimensional case. In fact, this is the only case for which we know the exact constant hidden in the $O(\cdot)$-notation. The general ℓ-dimensional assignment problem can be attacked in a similar manner as we did with the linear (two-dimensional) assignment problem. However, for such a general problem one needs a condition that assures the existence of a perfect matching in an ℓ-partite graph. Even the three-dimensional case is not easy. In particular, in this case one needs a generalization of the Hall's criterion to construct a perfect matching in a tripartite hypergraph. The problem seems not to be a trivial one, and our solution proposed for the two-dimensional assignment problem can possible lead to some results. It can be easily checked, however, that the lower bound technique based on the urn-and-ball model does1 not work particularly well in this case. This makes the problem even more interesting.

References

[1] D. Avis, A Survey of Heuristics for the Weighted Matching Problem, *Networks*, 13, 475-493, 1983.

[2] D. Avis and C.W. Lai, The Probabilistic Analysis of a Heuristic for the Assignment Problem, *SIAM J. Computing*, 17, 732-741, 1988.

[3] B. Bollobas, *Random Graphs*, Academic Press, London, 1985.

[4] A.A. Borovkov, A Probabilistic Formulation of Two Economic Problems, *Soviet Mathematics*, 3, 419-427, 1962.

[5] M. Dyer, A. Frieze, and C.J.H. McDiarmid, On Linear Programs with Random Costs, *Mathematical Programming*, 35, 3-16, 1986.

[6] J. Frenk, M. van Houweninge, and A. Rinnooy Kan, Order Statistics and the Linear Assignment Problem, *Computing*, 39, 165-174, 1987.

[7] A. Frieze, On Matching and Hamilton Cycles in Random Graphs, in *Surveys in Combinatorics, 1989* (ed. J. Siemons), 84-114, Bambridge University Press, 1989.

[8] A. Frieze and T. Luczak, Hamiltonian Cycles in a Class of random Graphs: One Step Further, preprint 1989.

[9] J. Jaworski, and T. Luczak, Cycles in a Uniform Graph Process, Technical Report, Adam Mickiewicz University, Poznań, 1991.

[10] R. Karp, An Upper Bound on the Expected Cost of an Optimal Assignment, in *Discrete Algorithms and Complexity*, ed. D.S. Johnson, T. Nishizeki, A. Nozaki and H.S. Wilf, Academic Press, New York, 1987.

[11] R. Karp, A. Rinnooy Kan, and R. Vohra, Average Case Analysis of a Heuristic for the Assignment Problem, Technical report, Ohio State University, WPS 91-83, 1991.

[12] V. Kolchin, B. Sevatianov, and V. Chistyakov, *Random Allocations*, John Wiley&Sons, New York, 1978.

[13] J. Kurtzberg, On Approximate Methods for the Assignment Problem, *J. of the ACM*, 9, 419-439, 1962.

[14] E. Lawler, *Combinatorial Optimization: Networks and Matroids*, Saunders College Publishing, Fort Worth, 1976.

[15] L. Lovász and M. Plummer, *Matching Theory*, North-Holland, Amsterdam, 1986.

[16] A. Lazarus, The Assignment Problem with Uniform $(0,1)$ Cost Matrix, B.A. Thesis, Dept. Mathematics, Princeton University, 1979.

[17] M. Mézard, and G. Parisi, Replicas and Optimization, *J. Physique Lettre*, 46, L-771-L-778, 1985.

[18] S. Micali and Vazirani, V., An $O(\sqrt{|V|}\cdot E)$ Algorithm for Finding Maximum Matching in General Graphs, *Proc. 21-st Annual IEEE Symp. on Found. of Comput. Sci.*, 17-27, 1980.

[19] P. Shor, Private Communication.

[20] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models*, John Wiley&Sons, New York 1983.

[21] W. Szpankowski, Combinatorial optimization through order statistics, *Second Annual International Symposium on Algorithms*, Taiwan, 1991.

[22] D. Walkup, On the Expected Value of a Random Assignment Problem, *SIAM J. Computing*, 8, 440-442, 1979.

[23] D. Walkup, Matching in Random Regular Bipartite Digraphs, *Discrete Mathematics*, 31, 59-64, 1980.

[24] M. Zuker, Computer Prediction of RNA Structure, *Methods in Enzymology*, 180, 262-288, 1989.