Elect. Comm. in Probab. **13** (2008), 258–265

ELECTRONIC COMMUNICATIONS in PROBABILITY

RANDOM MATCHING PROBLEMS ON THE COMPLETE GRAPH

JOHAN WÄSTLUND[1]
Department of Mathematical Sciences, Chalmers University of Technology, S-412 96 Gothemburg, Sweden
email: wastlund@chalmers.se

Submitted April 10, 2008, *accepted in final form* April 28, 2008

AMS 2000 Subject classification: 60C05, 90C27, 90C35.
Keywords: Minimum matching, exponential, expectation, mean field, network.

Abstract
The edges of the complete graph on n vertices are assigned independent exponentially distributed costs. A k-matching is a set of k edges of which no two have a vertex in common. We obtain explicit bounds on the expected value of the minimum total cost $C_{k,n}$ of a k-matching. In particular we prove that if $n = 2k$ then $\pi^2/12 < EC_{k,n} < \pi^2/12 + \log n/n$.

1 The random matching problem

The edges of the complete graph on n vertices $v_1, \dots, v_n$ are assigned independent costs from exponential distribution with rate 1. A *k-matching* is a set of k edges of which no two have a vertex in common. We let $C_{k,n}$ denote the minimum total cost of the edges of a k-matching. In 1985 Marc Mézard and Giorgio Parisi [5, 6] gave convincing evidence that as $n \to \infty$,

$$E\left(C_{\lfloor n/2 \rfloor,n}\right) \to \frac{\pi^2}{12}. \tag{1}$$

This was proved in 2001 by David Aldous [1, 2]. He considered the related *assignment problem* on the complete bipartite graph, which is technically simpler. It is known that (1) follows from the results of [2] by a slight modification of Proposition 2 of [1].
We give a simple proof of (1) by establishing explicit upper and lower bounds on $E\left(C_{k,n}\right)$ valid for arbitrary k and n. For *perfect* matchings, that is, when n is even and $k = n/2$, we prove that

$$\frac{\pi^2}{12} < E\left(C_{n/2,n}\right) < \frac{\pi^2}{12} + \frac{\log n}{n}. \tag{2}$$

Notice that the difference between the upper and lower bounds in (2) is much smaller than the random fluctuations of $C_{n/2,n}$. It is not hard to show that the standard deviation of $C_{n/2,n}$ is at least of order $n^{-1/2}$.

[1]RESEARCH SUPPORTED BY THE SWEDISH RESEARCH COUNCIL

2 The extended graph

In the *extended graph* there is an extra vertex v_{n+1}, and the costs of the edges from this vertex are exponentially distributed with rate $\lambda > 0$. In the end, λ will tend to zero. We say that a vertex v *participates* in a matching if the matching contains an edge incident to v.
In the following, we shall assume that the edge costs are such that no two distinct matchings have the same cost (this holds with probability 1), and we let σ_k be the minimum cost k-matching in the extended graph. We let $P_k(n)$ denote the normalized probability that v_{n+1} participates in σ_k. More precisely,

$$P_k(n) = \lim_{\lambda\to 0} \frac{1}{\lambda} P(v_{n+1} \text{ participates in } \sigma_k).$$

Lemma 2.1.

$$E\left(C_{k,n}\right) - E\left(C_{k-1,n-1}\right) = \frac{1}{n} P_k(n), \tag{3}$$

and consequently

$$E\left(C_{k,n}\right) = \frac{1}{n} P_k(n) + \frac{1}{n-1} P_{k-1}(n-1) + \cdots + \frac{1}{n-k+1} P_1(n-k+1). \tag{4}$$

Proof. The right hand side of (3) is the normalized probability that a particular edge from v_{n+1}, say the edge $e = (v_n, v_{n+1})$, belongs to σ_k. Naturally, $C_{k,n}$ denotes the cost of the minimum k-matching on the vertices $v_1, \dots, v_n$. We couple $C_{k,n}$ and $C_{k-1,n-1}$ by letting $C_{k-1,n-1}$ be the cost of the minimum $(k-1)$-matching on the vertices $v_1, \dots, v_{n-1}$.
Let w be the cost of the edge e. If e participates in σ_k, then we must have $w \le C_{k,n} - C_{k-1,n-1}$. Conversely, if $w \le C_{k,n} - C_{k-1,n-1}$ then e will participate in σ_k unless there is some other edge from v_{n+1} that does. This can happen only if both e and some other edge from v_{n+1} have costs smaller than $C_{k,n}$. As $\lambda \to 0$, the probability for this is $O(\lambda^2)$. Hence

$$\begin{aligned} \frac{1}{n} P_k(n) &= \lim_{\lambda\to 0} \frac{1}{\lambda} P(w \le C_{k,n} - C_{k-1,n-1}) \\ &= \lim_{\lambda\to 0} \frac{1}{\lambda} E\left(1 - e^{-\lambda(C_{k,n} - C_{k-1,n-1})}\right) = E\left(C_{k,n}\right) - E\left(C_{k-1,n-1}\right). \end{aligned} \tag{5}$$

□

We therefore wish to estimate $P_k(n)$ for general k and n. For this purpose we design a random process driven by the edge costs. A convenient way to think about this process is to imagine that there is an oracle who knows all the edge costs. We ask questions to the oracle in such a way that we can control the conditional distribution of the edge costs while at the same time being able to determine whether v_{n+1} participates in σ_k or not. The following lemma is well-known in matching theory, but for completeness we include a proof.

Lemma 2.2. *Every vertex that participates in σ_r also participates in σ_{r+1}.*

Proof. Let H be the symmetric difference $\sigma_r \triangle \sigma_{r+1}$ of σ_r and σ_{r+1}, in other words the set of edges that belong to one of them but not to the other. Since no vertex has degree more than 2, H consists of paths and cycles. We claim that H consists of a single path. If this would not be the case, then it would be possible to find a subset $H_1 \subseteq H$ consisting of one or two components of H, such that H_1 contains equally many edges from σ_r and σ_{r+1}. By assumption, the edge sets $H_1 \cap \sigma_r$ and $H_1 \cap \sigma_{r+1}$ do not have equal total cost. Therefore either $H_1 \triangle \sigma_r$ has smaller cost than σ_r, or $H_1 \triangle \sigma_{r+1}$ has smaller cost than σ_{r+1}, a contradiction.
The fact that H is a path clearly implies the statement of the lemma. □

3 The lower bound

The lower bound on $E(C_{k,n})$ is the simpler one and we establish it first. In Section 4, a modification of the method will yield a fairly good upper bound as well.

3.1 The process

The following protocol for asking questions to the oracle looks like an algorithm for finding the minimum k-matching, but what we are interested in is the probability that v_{n+1} participates in the minimum matching.
At each stage of the process, we say that a certain set of vertices are *exposed*, and the remaining vertices are *unexposed*. We have the following information:

1. We know the costs of all edges between exposed vertices.
2. For each exposed vertex v, we also know the minimum cost of the edges connecting v to the unexposed vertices.
3. Finally, we know the minimum cost of all edges connecting two unexposed vertices.

Another way to put this is to say that for every set A of at most two exposed vertices, we know the minimum cost of the edges whose set of exposed endpoints is precisely A. By well-known properties of independent exponential variables, the minimum is located with probabilities proportional to the rates of the corresponding exponential variables, and conditioning on a certain edge not being the one of minimum cost, its cost is distributed like the minimum plus another exponential variable of the same rate.
We also keep track of a nonnegative integer r which is such that σ_r contains only edges between exposed vertices. Moreover, we shall require that it can be verified from the information at hand that this matching is indeed the minimum r-matching. This is the reason why possibly some more vertices have to be exposed.
Initially, $r = 0$ and no vertex is exposed. At each stage of the process, the following happens:

- We compute a proposed minimum $(r + 1)$-matching under the assumption that for all exposed vertices, their minimum cost edges to an unexposed vertex go to different unexposed vertices.

  By Lemma 2.2, σ_{r+1} will use at most two unexposed vertices. Hence either it contains the minimum cost edge connecting two unexposed vertices, or at most two of the minimum cost edges connecting an exposed vertex to an unexposed one.
- If the proposed minimum $(r + 1)$-matching contains the minimum edge connecting two unexposed vertices, then it must indeed be the minimum $(r + 1)$-matching. The two endpoints of the new edge are exposed (that is, we ask for the information required for them to be exposed). Finally, the value of r increased by 1.
- Otherwise the proposed matching includes up to two edges from exposed vertices to unexposed ones. Then the unexposed endpoints of these edges are revealed and exposed. Unless there are two such edges and they happen to have the same endpoint, the proposed matching is indeed the minimum $(r + 1)$-matching, and the value of r is increased. If there are two edges to unexposed vertices and it turns out that they "collide", that is, they have the same unexposed endpoint, then the proposed matching is not valid. We have then exposed only one more vertex, and we complete the round of the process without updating the value of r.

3.2 A lower bound on $P_k(n)$

We wish to estimate the probability that v_{n+1} participates in σ_k. Suppose that at a given stage of the process there are m ordinary unexposed vertices (that is, not counting v_{n+1}). There are two cases to consider.

Suppose first that an edge between two unexposed vertices is going to be revealed. The total rate of the edges between unexposed vertices is

$$\binom{m}{2} + O(\lambda)$$

and the total rate of the edges from v_{n+1} to the other unexposed vertices is λm. Hence the probability that v_{n+1} is among the two new vertices to be exposed is

$$\frac{\lambda m}{\binom{m}{2} + O(\lambda)} = \frac{2\lambda}{m-1} + O(\lambda^2).$$

For convenience we suppose that the vertices are revealed one at a time, with a coin toss to decide which vertex to be revealed first. Then the probability that v_{n+1} is exposed is

$$\frac{\lambda}{m-1} + O(\lambda^2)$$

for both the first and the second vertex.

Secondly, suppose that the unexposed endpoint of an edge from an exposed vertex is going to be revealed. If at one stage of the process there are two such edges, then again we reveal the endpoints one at a time, flipping a coin to decide the order. In case there is a collision, this will be apparent when the first new vertex is exposed. If there are m ordinary remaining unexposed vertices, then the total rate of the edges from a particular exposed vertex v to them is $m + \lambda$, and consequently the probability that v_{n+1} is exposed is

$$\frac{\lambda}{m} + O(\lambda^2).$$

This will hold also for the second edge of two to be exposed at one stage of the process, provided m denotes the number of remaining unexposed vertices at that point.

If v_{n+1} is among the first $2k$ vertices to be exposed, then it will participate in σ_k. We have neglected the possibility that there is a collision at v_{n+1}, since this is an event of probability $O(\lambda^2)$. When m ordinary unexposed vertices remain, the probability that v_{n+1} is the next vertex to be exposed is at least λ/m. Hence

$$P_k(n) \geq \frac{1}{n}\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-2k+1}\right).$$

We can improve slightly on this inequality by noting that the normalized probability that v_{n+1} is one of the two first vertices to be exposed is exactly

$$\frac{n}{\binom{n}{2}} = \frac{2}{n-1}.$$

Taking this into account, we get

$$P_k(n) \geq \frac{1}{n}\left(\frac{2}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{n-2k+1}\right). \tag{6}$$

3.3 A lower bound on $E\left(C_{k,n}\right)$

From (4) and (6) it follows that

$$\begin{aligned} E\left(C_{k,n}\right) \geq \frac{1}{n}&\left(\frac{2}{n-1}+\frac{1}{n-2}+\cdots+\frac{1}{n-2k+1}\right) \\ &+\frac{1}{n-1}\left(\frac{2}{n-2}+\frac{1}{n-3}+\cdots+\frac{1}{n-2k+2}\right) \\ &\vdots \\ &+\frac{1}{n-k+1}\left(\frac{2}{n-k}\right). \end{aligned} \tag{7}$$

It is straightforward to prove by induction on n that when n is even and $k=n/2$, (7) becomes

$$E\left(C_{n/2,n}\right) \geq \frac{1}{2}\left(1+\frac{1}{4}+\frac{1}{9}+\cdots+\frac{1}{(n/2)^2}\right)+\frac{1}{n}.$$

By a simple integral estimate,

$$E\left(C_{n/2,n}\right) \geq \frac{1}{2}\left(\frac{\pi^2}{6}-\int_{n/2}^{\infty}\frac{dx}{x^2}\right)+\frac{1}{n}=\frac{\pi^2}{12}. \tag{8}$$

4 The upper bound

By modifying the argument given in the previous section, we can also establish an upper bound on $E\left(C_{k,n}\right)$. For this purpose, we are going to design the process differently.

4.1 The process

We modify the process described in Section 3. This time, only the vertices that participate in σ_r will be exposed. At each stage, we have the following information:

1. The costs of all edges between exposed vertices. In particular, σ_r is known.
2. For each exposed vertex v, we know the minimum cost of the edges from v to unexposed vertices.
3. For some exposed vertices, we may also know to which vertex this minimum cost edge goes, and the cost of the second cheapest edge to an unexposed vertex.
4. We also know the minimum cost of the edges connecting two unexposed vertices.

As in the previous section, we assume that the information we have is sufficient to verify that the given r-matching is indeed of minimum cost. We also assume that for the exposed vertices for which the minimum cost edge to an unexposed vertex is known, this edge never goes to v_{n+1}. As will become clear below, this assumption is justified by the fact that such an edge is revealed only in case of a collision. The event of a collision at v_{n+1} has probability $O(\lambda^2)$, which is negligible since we are estimating a probability of order λ.

We of course assume that $2r+2 \leq n$, so that there are at least two ordinary unexposed vertices. Given the information we have, we compute a proposed minimum cost $(r+1)$-matching under the assumption that no collision takes place. Then we ask whether or not the proposed matching is valid. If it is invalid, that is, if there is a collision, then this must be between the minimum cost edges to unexposed vertices from two of the exposed vertices. This endpoint is revealed, and we can assume that it is not v_{n+1}, since the probability for this is negligible. We repeat this until we find that there is no collision, and that therefore the proposed matching is valid. It must then be the minimum $(r+1)$-matching.

4.2 An upper bound on $P_k(n)$

We analyze a particular stage of the process and we wish to obtain an upper bound on the probability that v_{n+1} is one of the two new vertices that are used in σ_{r+1}. We let $m = n - 2r$ be the number of ordinary unexposed vertices.

If σ_{r+1} has an edge between two unexposed vertices, then by the analysis of Section 3, the probability that it uses the vertex v_{n+1} is

$$\frac{\lambda m}{\binom{m}{2}} = \frac{2\lambda}{m-1},$$

neglecting a term of order λ^2.

Suppose instead that σ_{r+1} contains two edges from exposed vertices v_i and v_j to unexposed vertices. First assume that we do not know the minimum cost edge to an unexposed vertex for any of them (that is, none of them falls under (3) above). Then the probability that v_{n+1} participates in σ_{r+1} is

$$\frac{\lambda m + \lambda m}{m(m-1)} = \frac{2\lambda}{m-1},$$

since we are conditioning on no collision occurring.

If on the other hand for at least one of v_i and v_j the minimum cost edge to an unexposed vertex is known according to (3), then the probability that v_{n+1} participates in σ_{r+1} is even smaller, at most $\lambda/(m-1)$.

This gives the following upper bound on the probability that v_{n+1} participates in σ_k:

$$P_k(n) \leq \frac{2}{n-1} + \frac{2}{n-3} + \cdots + \frac{2}{n-2k+1}.$$

4.3 An upper bound on $E(C_{k,n})$

Inductively we obtain the following upper bound on the expected cost of σ_k:

$$\begin{aligned} E(C_{k,n}) \leq \frac{1}{n}\left(\frac{2}{n-1} + \frac{2}{n-3} + \cdots + \frac{2}{n-2k+1}\right) \\ + \frac{1}{n-1}\left(\frac{2}{n-2} + \cdots + \frac{2}{n-2k+2}\right) \\ \vdots \\ + \frac{1}{n-k+1}\left(\frac{2}{n-k}\right). \end{aligned}$$

We give a slightly weaker but simpler upper bound, valid for even n and $k = n/2$, in order to establish (2). If we replace all the terms of the form $2/(n-i)$ except the first one in each pair of parentheses by $1/(n-i) + 1/(n-i-1)$, we obtain

$$\begin{aligned}
E\left(C_{n/2,n}\right) \leq \frac{1}{n} + \frac{1}{2(n-1)} + \cdots + \frac{1}{(n/2)(n/2+1)} \\
+ \frac{1}{n}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n-1}\right) + \frac{1}{n-1}\left(\frac{1}{2} + \cdots + \frac{1}{n-2}\right) + \\
\vdots \\
+ \frac{1}{(n/2)((n/2)+1)} \\
= \frac{1}{2}\left(1 + \frac{1}{4} + \cdots + \frac{1}{(n/2)^2}\right) + \frac{1}{n+1}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) \\
\leq \frac{1}{2}\left(\frac{\pi^2}{6} - \int_{n/2+1}^{\infty} \frac{dx}{x^2}\right) + \frac{1 + \log n}{n+1} = \frac{\pi^2}{12} - \frac{1}{n+2} + \frac{1}{n+1} + \frac{\log n}{n+1} \\
= \frac{\pi^2}{12} + \frac{\log n}{n} - \frac{\log n - \frac{n}{n+2}}{n(n+1)} \leq \frac{\pi^2}{12} + \frac{\log n}{n}, \quad (9)
\end{aligned}$$

since $\log n \geq n/(n+2)$ for $n \geq 2$. Together with (8), this establishes (2).

5 Concluding remarks

We have inductively established lower and upper bounds on $C_{k,n}$ from lower and upper bounds on the normalized probability $P_{k,n}$ that an extra vertex participates in the minimum matching. For the corresponding problem on the complete m by n bipartite graph (the so called *assignment problem*), the expected cost of the minimum k-matching is known [4, 7], and is given by the simple formula

$$\sum_{\substack{i,j \geq 0 \\ i+j<k}} \frac{1}{(m-i)(n-j)}, \qquad (10)$$

conjectured in [3] as a generalization of the formula

$$\sum_{i=1}^{n} \frac{1}{i^2}$$

suggested in [8] for the special case $k = m = n$. For the complete graph, no such formula has even been conjectured.

We can now see why finding the exact value is harder for the complete graph. The probability that the extra vertex is included when we pass from σ_r to σ_{r+1} depends on whether σ_{r+1} is obtained by adding an edge between two vertices that do not participate in σ_r, or by an alternating path that replaces edges in σ_r by other edges. For the bipartite graph, the probability that an extra vertex is included is the same in the two cases, and thereby known. A proof of (10) based on this method is given in [9].

References

[1] Aldous, David, *Asymptotics in the random assignment problem*, Probab. Theory Relat. Fields **93** (1992), 507–534. MR1183889

[2] Aldous, David, *The $\zeta(2)$ limit in the random assignment problem*, Random Structures & Algorithms **18** no. 4 (2001), 381–418. MR1839499

[3] Coppersmith, Don and Sorkin, Gregory B., *Constructive Bounds and Exact Expectations For the Random Assignment Problem*, Random Structures & Algorithms **15** (1999), 133–144. MR1704340

[4] Linusson, Svante and Wästlund, Johan, *A proof of Parisi's conjecture on the random assignment problem*, Probab. Theory Relat. Fields **128** (2004), 419–440.

[5] Mézard, Marc and Parisi, Giorgio, *Replicas and optimization*, Journal de Physique Lettres **46** (1985), 771–778.

[6] Mézard, Marc and Parisi, Giorgio, *On the solution of the random link matching problems*, Journal de Physique Lettres **48** (1987), 1451–1459.

[7] Nair, Chandra, Prabhakar, Balaji and Sharma, Mayank, *Proofs of the Parisi and Coppersmith-Sorkin random assignment conjectures*, Random Structures & Algorithms **27** No. 4 (2005), 413–444. MR2178256

[8] Parisi, Giorgio, *A conjecture on random bipartite matching*, arXiv:cond-mat/9801176 (1998).

[9] Wästlund, Johan, *An easy proof of the $\zeta(2)$ limit in the random assignment problem*, manuscript available at the author's webpage.