EXACT EXPECTATIONS FOR RANDOM GRAPHS

HENRIK ERIKSSON, KIMMO ERIKSSON, AND JONAS SJÖSTRAND

Abstract. For a random graph on n vertices where the edges appear with individual rates, we give exact formulas for the expected time at which the number of components has gone down to k and the expected length of the corresponding minimal spanning forest.

For a random bipartite graph we give a formula for the expected time at which a k-assignment appears. This result has bearing upon the random assignment problem.

Résumé. Pour un graphe aléatoire avec n sommets où les arcs apparaissent avec des intensités individuelles nous donnons des formules exactes pour le temps espéré quand le nombre de composants se réduit à k et pour la longueur espérée de l'arborescence recouvrante minimale correspondante.

Pour un graphe bipartie aléatoire nous donnons une formule pour calculer le temps expecté quand une k-affectation apparait. Ce résultat est lié au problème d'affectation aléatoire.

1. Introduction

The study of random graphs, initiated by Erdös forty years ago, has been preoccupied by asymptotic results. In this paper, we will derive exact enumerative formulas in some cases where previously only asymptotics were known.

By a random graph $G(t)$ on n vertices we mean a stochastic process starting with an edgeless graph at $t = 0$ and where edge after edge appears at random times. We take these times to be independent exponentially distributed variables, so the probability of edge e_{ij} not having appeared at time t is $\exp(-a_{ij}t)$. The rates a_{ij} may be viewed as labelling the edges of K_n and any nonnegative edge labelling specifies such a random graph process.

The number of graph components at time t is a thoroughly studied statistic and asymptotics for its expectation and distribution are known in many cases (see Janson [8]). Of particular importance is the time T_n at which the graph becomes one component, i.e. gets a spanning tree, and also the minimal length L_n of such a tree, i.e. the sum of its edge times. (By Kruskal's algorithm the first tree has indeed minimal length.) For the simplest case when all rates are 1, there is the following beautiful result by Frieze.

$$\lim_{n\to\infty} \mathrm{E}(L_n) = \zeta(3) = 1 + \frac{1}{2^3} + \frac{1}{3^3} + \frac{1}{4^3} + \cdots = 1.202\ldots$$

The first published version of the above result seems to be by Fenner and Frieze [3], where the distribution used was the uniform distribution on $[0, 1]$. Later Frieze [4] showed that the same result holds for any distribution function F such that $F'(0)$ exists and equals 1. The Exp(1) distribution is an obvious example.

In the first theorem of this paper, we will give simple exact formulas for the following more general statistics.

$$\begin{aligned} \mathrm{E}(T_n(k)) &= \mathrm{E}(t \,|\, G(t) \text{ becomes a } k\text{-component graph}) \\ \mathrm{E}(L_n(k)) &= \mathrm{E}(\text{length of the minimal spanning } k\text{-forest}) \end{aligned}$$

Note that our formulas work for arbitrary rates a_{ij}. The special case where all $a_{ij} = 1$ turns out to be a sum over all integer partitions of n (Corollary 1.2). In the general case we will use the following notation. Let the rates a_{ij} label the edges $\mathcal{E}_n$ of K_n. For an edge $e_{ij} \in \mathcal{E}_n$, let $[e_{ij}] = a_{ij}$ and for

Date: November 12, 2000.
1991 *Mathematics Subject Classification.* Primary: 05C80; Secondary: 05C40, 60K99.

a subset $A \subseteq \mathcal{E}_n$ of edges, let $[A]$ denote the rate sum $\sum_{e_{ij}\in A} a_{ij}$. A subset $A \subseteq \mathcal{E}_n$ is a *clique* if it spans a complete subgraph. Any partition of the vertices of K_n into j parts defines a subgraph consisting of j cliques. If B is the edge set of such a subgraph, we write $B \in \text{Clique}_j$.

Theorem 1.1. *For a random graph $G(t)$ on n vertices with rates a_{ij}*

$$\begin{aligned} \mathrm{E}(T_n(k)) &= \sum_{j=k+1}^{n} \sum_{B\in \text{Clique}_j} \frac{t(j,k)}{[\mathcal{E}_n - B]} \\ \mathrm{E}(L_n(k)) &= \sum_{j=k+1}^{n} \sum_{B\in \text{Clique}_j} \frac{l(j,k)}{[\mathcal{E}_n - B]} \end{aligned}$$

where $t(j,k) = \sum_{i=k+1}^{j} s(j,i)$ and $l(j,k) = \sum_{i=k+1}^{j}(i-k)s(j,i)$ are sums of signed Stirling numbers of the first kind.

For example, K_3 has three subgraphs in Clique_2 and one in Clique_3, so

$$\mathrm{E}(T_3(1)) = \frac{1}{a_{12}+a_{13}} + \frac{1}{a_{12}+a_{23}} + \frac{1}{a_{13}+a_{23}} - \frac{2}{a_{12}+a_{13}+a_{23}}.$$

If we put all $a_{ij} = 1$ and also $k = 1$, we should get an expression that tends to Frieze's result when n goes to infinity. We use the notation $\binom{n}{1^{e_1},2^{e_2},\ldots}$ for the multinomial number $\frac{n!}{(1!)^{e_1}(2!)^{e_2}\ldots}$.

Corollary 1.2. *For a random graph $G(t)$ on n vertices with rates 1*

$$\begin{aligned} \mathrm{E}(T_n(1)) = & \sum \tbinom{n}{1^{e_1},2^{e_2},\ldots}\tbinom{e}{e_1,e_2,\ldots} \tfrac{(-1)^e}{\frac{e}{2}(n^2-(1^2e_1+2^2e_2+\ldots))} \\ \mathrm{E}(L_n(1)) = & \sum \tbinom{n}{1^{e_1},2^{e_2},\ldots}\tbinom{e}{e_1,e_2,\ldots} \tfrac{(-1)^e}{\binom{e}{2}(n^2-(1^2e_1+2^2e_2+\ldots))} \end{aligned}$$

where the sums are over integer partitions of n into two or more parts, i.e. nonnegative integer sequences (e_i) such that $e = e_1 + e_2 + \cdots \geq 2$ and $1e_1 + 2e_2 + \cdots = n$.

For example, $\mathrm{E}(L_3(1)) = \binom{3}{1,2}\binom{2}{1,1}/\binom{2}{2}(9-1-4) - \binom{3}{1,1,1}\binom{3}{3}/\binom{3}{2}(9-3) = \frac{7}{6}$. Clearly, the terms of the formulas grow exponentially with n. It is an intriguing consequence of Frieze's theorem that, nevertheless, the last sum must tend to $\zeta(3)$ as $n \to \infty$.

1.1. **Random assignment.**

Our reason for studying this problem was to get a grip on the *random assignment problem* which is the study of minimal assignments in a random bipartite graph $G(t) \subseteq K_{m,n}$. As before, the time of appearance of edge e_{ij} is assumed to be exponentially distributed with rate a_{ij} and we consider these rates as labelling the edges of the complete bipartite graph $K_{m,n}$.

A k-assignment is a disjoint set of k edges. The following two statistics are of special interest.

$$\begin{aligned} \mathrm{E}(T_{mn}(k)) &= \mathrm{E}(t \,|\, G(t)\,\text{obtains a } k\text{-assignment}) \\ \mathrm{E}(L_{mn}(k)) &= \mathrm{E}(\text{length of the minimal } k\text{-assignment}) \end{aligned}$$

The analogy with the previous definitions is deceptive, for here the k-assignment of $T_{mn}(k)$ is not the k-assignment of $L_{mn}(k)$ in general! Since the length of an assignment is the sum of the appearance times of the edges involved, the first k-assignment to be completed need not have the minimal length. Because of this, the computation of $\mathrm{E}(L_{mn}(k))$ seems to be quite difficult and we have obtained formulas only for the simplest cases.

For $k = m = n$, we have the original random assignment problem. The following conjecture, proposed by Parisi [7] in 1998, has been verified up to $n = 7$.

Conjecture 1.3. $\mathrm{E}(L_{nn}(n)) = 1 + 1/4 + 1/9 + \cdots + 1/n^2$.

Note that this expression tends to $\zeta(2)$ as $n \to \infty$. This asymptotic conjecture, stated by Mézard and Parisi in 1985 [6], is also unproved.

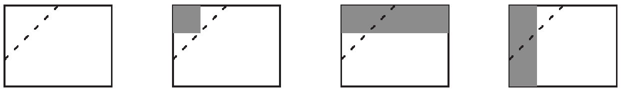

FIGURE 1. The four tableau shapes in Tabloid_2.

Generalizations of the conjecture to general k, m, n have been proposed by Linusson and Wästlund [5] and by Buck, Chan and Robbins [1], who also consider more general rates a_{ij}. As far as we know, however, arbitrary rates have not been studied before.

For the expected time until the first k-assignment appears, our method gives the following formula.

Theorem 1.4.

$$\mathrm{E}(T_{mn}(k)) = \sum_{B \in \text{Tabloid}_k} \frac{S_{m,n,k}(B)}{[\mathcal{E}_n - B]}$$

The numerators $S_{m,n,k}(B)$ are certain products of binomial coefficients to be defined in Section 3. An edge set B is called tabloidal if the matrix entries $e_{ij} \in B$ fill a tableau shape (Young diagram), possibly after permutation of rows and columns. It is in Tabloid_k if the inner corners of this tableau satisfy $i + j < k$ and the outer corners $i + j \geq k$.

With $k = 2$, for example, four tableau shapes are involved and the expression becomes

$$\mathrm{E}(T_{mn}(2)) = \frac{\binom{m-1}{1}\binom{n-1}{1}}{[\mathcal{E}_n]} - \sum_{i,j} \frac{1}{[\mathcal{E}_n - e_{ij}]} + \sum_i \frac{1}{[\mathcal{E}_n - \text{row}_i]} + \sum_j \frac{1}{[\mathcal{E}_n - \text{col}_j]}.$$

If we try to compute $\mathrm{E}(L_{mn}(k))$ as $T_{mn}(1) + T_{mn}(2) + \cdots + T_{mn}(k)$ we get a value that is too small, since the subassignments of the assignment of minimal length need not be minimal for T_{mn}. For small k we have been able to calculate $\mathrm{E}(L_{mn}(k))$, for example

$$\mathrm{E}(L_{mn}(2)) = T_{mn}(1) + T_{mn}(2) + \sum_{i,j} \frac{a_{ij}[\text{row}_i - e_{ij}][\text{col}_j - e_{ij}]}{[\mathcal{E}_n][\mathcal{E}_n - e_{ij}][\mathcal{E}_n - \text{row}_i][\mathcal{E}_n - \text{col}_j]}.$$

The derivation of this expression and similar ones will appear in our forthcoming paper [2].

2. EXPECTATIONS ON SPANNING k-FORESTS

This section is devoted to the proof of Theorem 1.1. We will start by proving the formula for $\mathrm{E}(T_n(k))$, which will take a series of steps.

In order to express the expected time until the random graph $G(t)$ contains a spanning k-forest (i.e. k components), we will study the Boolean lattice of all subsets of edges. Each subset $A \subseteq \mathcal{E}_n$ of edges determines a subgraph $(\mathcal{V}_n, A)$ of K_n, where all n original vertices are still present. Let $\kappa(A)$ denote the number of connected components of this subgraph.

When edges appear, one at a time, we follow a path upwards in the Boolean lattice from the empty set to the complete edge set. At some time $T_n(k)$, the path reaches a set A where $\kappa(A) = k$. We want to determine the expected time until this happens. The following lemma is an immediate consequence of the independence of the edge appearances.

Lemma 2.1.

$$\mathrm{E}(T_n(k)) = \sum_{\kappa(A) > k} \mathrm{P}(A)\mathrm{E}(T_A), \tag{1}$$

where $\mathrm{P}(A)$ *denotes the probability that at some point the edge set of the random graph equals* A*, and* $\mathrm{E}(T_A)$ *is the expected time spent in configuration* A *until another edge appears.*

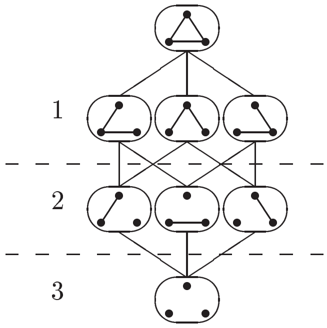

FIGURE 2. The Boolean subgraph lattice of K_3 partitioned according to the number of components.

For exponential distributions, the expected time in A is the reciprocal of the total rate of the as yet unappeared edges:

$$\mathrm{E}(T_n(k)) = 1/[\mathcal{E}_n - A],$$

where $[X]$ is defined as $\sum_{e_{ij}\in X} a_{ij}$. When all edges have appeared the denominator is zero, so we can expect to stay in that state forever.

Furthermore, the probability of the lattice path reaching A is the sum of the probabilities for each single path from the empty set to A, that is, for each permutation of the edges of A. In every step, the probability of a given edge appearing next is its share of the total rate of the unappeared edges.

Let S_A denote the set of permutations of the edges of A, and let $\pi = (\pi_1, \pi_2, \dots) \in S_A$ denote a generic such permutation. Thus,

$$\mathrm{P}(A) = \sum_{\pi\in S_A} \frac{a_{\pi_1}}{[\mathcal{E}_n]} \frac{a_{\pi_2}}{[\mathcal{E}_n - \pi_1]} \frac{a_{\pi_3}}{[\mathcal{E}_n - \pi_1 - \pi_2]} \cdots,$$

and consequently

$$\mathrm{P}(A)\mathrm{E}(T_A) = \left(\prod_{e\in A} a_e\right) \sum_{\pi} \prod_{i=0}^{|A|} \frac{1}{[\mathcal{E}_n - \pi_1 - \pi_2 - \cdots - \pi_i]}. \tag{2}$$

This expression can be simplified into a single sum as follows:

Lemma 2.2.

$$\mathrm{P}(A)\mathrm{E}(T_A) = \sum_{B\subseteq A} \frac{(-1)^{|A-B|}}{[\mathcal{E}_n - B]} \tag{3}$$

Proof. We use $\subseteq$ for arbitrary subsets and $\subset$ for proper subsets.

We argue by induction on the size of A. The identity is true for $A = \emptyset$. Fix an A with $|A| > 0$ and suppose the equality holds for all smaller sets. By the induction hypothesis we can write the right-hand side of (2) as

$$\sum_{e\in A} \sum_{B\subseteq A-e} \frac{(-1)^{|A-e-B|}[e]}{[\mathcal{E}_n - B][\mathcal{E}_n - A]}.$$

Changing the order of summation we obtain

$$\sum_{B\subset A} \sum_{e\in A-B} \frac{(-1)^{|A-B|-1}[e]}{[\mathcal{E}_n - B][\mathcal{E}_n - A]} = \sum_{B\subset A} \frac{(-1)^{|A-B|-1}[A-B]}{[\mathcal{E}_n - B][\mathcal{E}_n - A]}.$$

By partial fraction decompostition, the last sum can be rewritten as

$$\sum_{B\subset A} \frac{(-1)^{|A-B|}}{[\mathcal{E}_n - B]} - \frac{1}{[\mathcal{E}_n - A]} \sum_{B\subset A} (-1)^{|A-B|}.$$

Finally, we note that $\sum_{B\subseteq A}(-1)^{|A-B|}=0$ implies $-\sum_{B\subset A}(-1)^{|A-B|}=1$. □

Combination of Lemma 2.1 and Lemma 2.2 yields the following expression for the expected time until the random graph $G(t)$ has k components:

$$\mathrm{E}(T_n(k)) = \sum_{\kappa(A)>k}\sum_{B\subseteq A}\frac{(-1)^{|A-B|}}{[\mathcal{E}_n - B]} = \sum_{B\subseteq\mathcal{E}_n}\frac{1}{[\mathcal{E}_n - B]}\sum_{B\subseteq A,\ \kappa(A)>k}(-1)^{|A-B|} \tag{4}$$

by changing the order of summation. We next observe that the sum

$$\sum_{B\subseteq A,\ \kappa(A)>k}(-1)^{|A-B|} \tag{5}$$

is zero for many sets B. For suppose that B has some component that is not a clique (a complete subgraph). In other words there exists some edge e that we can add to B without changing the number of components of B. This property is obviously shared also by all supersets A of B where e is not already present. Pair every such A with $A+e$. The pair's contribution to the sum is $(-1)^{|A-B|}+(-1)^{|A+e-B|}=0$. The entire sum is over a disjoint collection of such pairs, so it is zero too.

We have now reduced the sum to sets B where all components are cliques. If there are j components, we write $B\in\mathrm{Clique}_j$. It turns out that each clique can be contracted into a node of $K_j=(\mathcal{V}_j,\mathcal{E}_j)$:

Lemma 2.3. *If $B\in\mathrm{Clique}_j$, the following equality of generating functions holds:*

$$\sum_{B\subseteq A\subseteq\mathcal{E}_n}(-1)^{|A-B|}x^{\kappa(A)} = \sum_{A'\subseteq\mathcal{E}_j}(-1)^{|A'|}x^{\kappa(A')}. \tag{6}$$

Proof. Fix j and argue by induction over n. The equality is trivially true for $n=j$ (in which case B must be the empty set). Suppose it holds for $n=i\geq j$. Now study $n=i+1$. Some clique has at least two vertices, say x and y. We shall see that omitting vertex y does not affect the sum.

If A is an extension of B including an edge from y to a vertex z outside the clique, then toggle edge xz to obtain $\bar{A}$ (that is, the only difference between A and $\bar{A}$ is whether the edge xz is included or not). Obviously $\kappa(A)=\kappa(\bar{A})$ so they both contribute to the same term in the generating function. Their joint contribution is zero. Hence, the existence of vertex y can be disregarded in the sum over edge sets A that are extensions of B, so we are down to i vertices again. □

The right-hand generating function of (6) can be computed explicitly. It can be derived as a special case of the prime example of the exponential method in Stanley's book [9], but we prefer to give a direct proof.

Lemma 2.4.

$$\sum_{A'\subseteq\mathcal{E}_j}(-1)^{|A'|}x^{\kappa(A')} = x(x-1)(x-2)\cdots(x-j+1).$$

(This is the generating function $\sum_{i=1}^{j}s(j,i)x^i$ for the signed Stirling numbers of the first kind.)

Proof. We argue by induction on j. The formula is true for $j=1$. For $j>1$ vertices, all A' such that at least two edges involve the jth vertex can be combined into zero-contributing pairs by toggling the edge forming a triangle with the lexically first pair of such edges. Hence we need only consider A' where the jth vertex is involved in zero edges (one possibility, decreasing the number of components by one if removed), or one edge ($j-1$ possibilities, changing sign if removed), thus resulting in the factor $(x-j+1)$ times the generating function for $j-1$ vertices. □

By the previous two lemmas we can compute the sum (5) for $B \in \text{Clique}_j$ as a sum of Stirling numbers:

$$\sum_{B\subseteq A\subseteq \mathcal{E}_n,\ \kappa(A)>k} (-1)^{|A-B|} = \sum_{A'\subseteq \mathcal{E}_j,\ \kappa(A')>k} (-1)^{|A'|} = \sum_{i=k+1}^{j} s(j,i).$$

We can now plug this into (4) and obtain

$$\mathrm{E}(T_n(k)) = \sum_{j=k+1}^{n} \sum_{i=k+1}^{j} s(j,i) \sum_{B\in \text{Clique}_j} \frac{1}{[\mathcal{E}_n - B]},$$

thereby proving the first formula of Theorem 1.1.

2.1. **Special cases.** The spanning tree special case (i.e. $k = 1$) can be simplified further. Since the sum of all Stirling numbers $s(j,i)$ for a fixed $j > 1$ is zero, we have

$$\sum_{i=2}^{j} s(j,i) = -s(j,1) = (-1)^j (j-1)!.$$

(Recall that the signed Stirling numbers of the first kind have the combinatorial interpretation that $s(j,i)$ is $(-1)^{j+i}$ times the number of permutations in S_j with i cycles.)

Specializing to all rates a_{ij} equaling 1, every partition B of K_n into cliques of given sizes contributes equally. With e_1 cliques of size 1, e_2 cliques of size 2, etc., the number of possibilities for B is

$$\frac{n!}{(1!)^{e_1} e_1! (2!)^{e_2} e_2! \ldots},$$

and the rate reciprocal $1/[\mathcal{E}_n - B]$ is $2/(n^2 - (1^2 e_1 + 2^2 e_2 + \ldots))$. Let $e = e_1 + e_2 + \ldots$ denote the number of cliques. We must multiply everything by $(-1)^e (e-1)! = (-1)^e e!/e$. Thus we obtain

$$\sum \binom{n}{1^{e_1}, 2^{e_2}, \ldots} \binom{e}{e_1, e_2, \ldots} \frac{(-1)^e}{\frac{e}{2}(n^2 - (1^2 e_1 + 2^2 e_2 + \ldots))},$$

where the sum is taken over all $e_1, e_2, \ldots$ such that $1e_1 + 2e_2 + \cdots = n$. This proves the first statement of Corollary 1.2.

2.2. **Expected length of the minimal spanning k-forest.** By Kruskal's algorithm for spanning trees, the minimal spanning k-forest in K_n consists of the longest edges of the first spanning r-forests for $r = n, n-1, n-2, \ldots, k$.

Thus, the expected length of the minimal spanning k-forest in K_n is

$$\mathrm{E}(L_n(k)) = \sum_{r=k}^{n} \mathrm{E}(T_n(r)) = \sum_{r=k}^{n} \sum_{j=r+1}^{n} \sum_{i=r+1}^{j} s(j,i) \sum_{B\in \text{Clique}_j} \frac{1}{[\mathcal{E}_n - B]}$$

$$= \sum_{j=k+1}^{n} \sum_{i=k+1}^{j} \sum_{r=k}^{i-1} s(j,i) \sum_{B\in \text{Clique}_j} \frac{1}{[\mathcal{E}_n - B]}$$

proving the second half of Theorem 1.1. The second half of Corollary 1.2 is proved in the same way as the first half.

3. Expected time to a k-assignment in $K_{m,n}$

In this section we study a random bipartite graph $G(t) \subseteq K_{m,n}$. Instead of trees, we are waiting for a k-assignment (a collection of k disjoint edges) to appear. Let $\mu(A)$ denote the size of the maximal subset of disjoint edges in an edge set $A \subseteq \mathcal{E}_{m,n}$.

As edges appear in $G(t)$ we follow a path from the empty set and upwards in the Boolean lattice, until we reach a set A where $\mu(A) = k$. Let $T_{mn}(k)$ be the waiting time. Precisely as in the previous section, we have

$$\mathrm{E}(T_{mn}(k)) = \sum_{\mu(A)<k} \mathrm{P}(A)\mathrm{E}(T_A). \tag{7}$$

As before, we have $\mathrm{E}(T_A) = 1/[\mathcal{E}_{m,n} - A]$ and obtain

$$\mathrm{P}(A)\mathrm{E}(T_A) = \sum_{B \subseteq A} \frac{(-1)^{|A-B|}}{[\mathcal{E}_{m,n} - B]}, \tag{8}$$

and finally

$$\mathrm{E}(T_{mn}(k)) = \sum_{B \subseteq \mathcal{E}_{m,n}} \frac{1}{[\mathcal{E}_{m,n} - B]} \sum_{B \subseteq A,\ \mu(A)<k} (-1)^{|A-B|}. \tag{9}$$

Hence, we have reduced the problem of computing $\mathrm{E}(T_{mn}(k))$ to a sum over edge sets B where each term is a fraction. The denominators are simple, but the numerators are sums over supersets $A \supseteq B$ that do not contain a k-assignment:

$$S_{m,n,k}(B) := \sum_{B \subseteq A,\ \mu(A)<k} (-1)^{|A-B|}, \tag{10}$$

which turns out to be much more complicated than in the previous problem.

An edge set B can be identified with a subset of the entries of an $m \times n$ matrix. We observe that the existence of a k-assignment is independent of the order of rows and columns. We say that B is *tabloidal* if the rows and columns can be reordered to make B tableau shaped, i.e. consisting of all entries above a lattice path. It turns out that only tabloidal B give nonzero $S_{m,n,k}(B)$.

Lemma 3.1. *$S_{m,n,k}(B) \neq 0$ only for nontabloidal edge sets B.*

Proof. It is easy to see that B is nontabloidal if and only if its set of matrix entries contains the pattern . For some of the supersets A in (10), the upper right square is filled, and these cancel the terms where this square is removed (the toggling trick). We are left with A such that the inclusion of the upper right square would complete a k-assignment. Since $\mu(A) = k-1$, König's theorem tells us that it can be covered by $k-1$ rows and columns. The upper right square isn't covered so the lower left square must be covered. Therefore it can be toggled and so everything cancels out. $\square$

Not even all tabloidal edge sets B contribute to the sum, only those in $\mathrm{Tabloid}_k$, that is such that the inner corners of the tableau lie above the k-diagonal and the outer corners lie below or on it (see Fig. 3). For such a tableau the numerator $S_{m,n,k}(B)$ is determined by the empty mid rectangles at the inner corners, defined by extending the vertical and horizontal lines. Let the size of the ith rectangle be $m_i \times n_i$ and note that $B \in \mathrm{Tabloid}_k$ means that the k-diagonal intersects the rectangle, defining a local k_i-diagonal.

It can now be proved that

$$S_{m,n,k}(B) = (-1)^{I+1} \prod_{i=1}^{I} S_{m_i,n_i,k_i}(\emptyset)$$

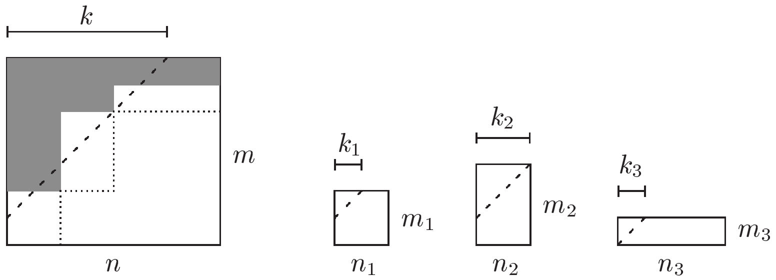

FIGURE 3. The decomposition into mid rectangles.

so we only have to find the correct expression for a mid rectangle, which turns out to be

$$S_{m,n,k}(\emptyset) = \binom{m-1}{k-1}\binom{n-1}{k-1}.$$

Now, we finally can complete Theorem 1.4.

Proposition 3.2. *In the formula*

$$\mathrm{E}(T_{mn}(k)) = \sum_{B \in \mathrm{Tabloid}_k} \frac{S_{m,n,k}(B)}{[\mathcal{E}_n - B]}$$

the numerators are given by

$$S_{m,n,k}(B) = (-1)^{I+1} \prod_{i=1}^{I} \binom{m_i - 1}{k_i - 1}\binom{n_i - 1}{k_i - 1}$$

where m_i, n_i, k_i refer to the I mid rectangles at the inner corners of the tableau B (see Fig. 3).

Proof. We must count extensions of B not containing a k-assignment. $S_{m,n,k}(B)$ is the number of even legal extensions minus the number of odd legal extensions. First we show that it is enough to consider extensions obtained by filling in entries in the mid rectangles only. Let U be the set of entries not in these rectangles nor in B. To see that the contribution of all extensions with some fixed entry in U is zero we invoke Lemma 3.1. This argument can be repeated for all entries in U and, by the inclusion-exclusion principle, the total contribution of extensions with at least one entry in U is

$$-\sum_{\emptyset \neq V \subseteq U} S_{m,n,k}(B+V) = 0.$$

Now we are ready for a proof by induction. The statement is trivial for $m = n = 1$ and for $k = 1$. Our induction assumption will be that the statement is true for all values of m, n, k smaller than the current ones (with at least one strict inequality).

Case 1. $B \neq \emptyset$. If B is the whole matrix there is nothing to prove. Otherwise we have the rectangle decomposition in Fig 3. We now show that an extension contains a k-assignment if and only if the ith mid rectangle contains a k_i-assignment for all i. Suppose that the latter statement is true. Then rows and columns can be permuted such that the dashed k-diagonal is covered, thus producing the k-assignment. Conversely, if the filled entries of the ith mid rectangle can be covered by less than k_i rows and columns, then together with the rows above and the columns to the left we have a covering of all filled entries by less than k rows and columns.

Finally we note that the signed sum of extensions not containing a k-assignment is equal to minus the corresponding signed sum of extensions containing a k-assignment, since $\sum(-1)^{|A|} = 0$ where the sum is taken over all subsets A of the union of the mid rectangles. Thus we get that

$$S_{m,n,k}(B) = (-1)^{I+1} \prod_{i=1}^{I} S_{m_i,n_i,k_i}(\emptyset)$$

Case 2. $B = \emptyset$. We regroup the terms in $S_{m,n,k}(\emptyset)$ according to the number of filled entries in the first row. Let S_0 denote the contribution from the legal extensions where the first row is completely empty. By inclusion-exclusion we get

$$S_0 = \sum_{V \subseteq \mathrm{row}_1} S_{m,n,k}(V),$$

and by our induction assumption, all terms in the sum vanish except for $V = \emptyset$ (the term we want to compute), $|V| = n$ (the whole row filled) and $|V| = k-1$ (and there are $\binom{n}{k-1}$ such terms). It is obvious how to reduce the cases of an empty row and of a filled row, so we get

$$\begin{aligned} S_{m,n,k}(\emptyset) &= S_{m-1,n,k}(\emptyset) - \binom{n}{k-1}(-1)S_{m-1,k-1,k-1}(\emptyset)S_{1,n-k+1,1}(\emptyset) - S_{m-1,n,k-1}(\emptyset) \\ &= \binom{m-2}{k-1}\binom{n-1}{k-1} + \binom{n}{k-1}\binom{m-2}{k-2}\binom{k-2}{k-2}\binom{0}{0}\binom{n-k}{0} - \binom{m-2}{k-2}\binom{n-1}{k-2} \\ &= \binom{m-1}{k-1}\binom{n-1}{k-1}. \end{aligned}$$

This completes the proof. $\square$

References

[1] M. W. Buch, C. S. Chan and D. P. Robbins, On the expected value of the minimum assignment, preprint, 2000, `arXiv.math.CO/0004175`.

[2] H. Eriksson, K. Eriksson and J. Sjöstrand, Random assignment with arbitrary rates, in preparation.

[3] T. I. Fenner and A. M. Frieze, On the connectivity of random m-orientable graphs and digraphs, *Combinatorica* **2** (1982), no. 4, 347–359.

[4] A. M. Frieze, On the value of a random minimum spanning tree problem, *Discrete Appl. Math.* **10** (1985), no. 1, 47–56.

[5] S. Linusson and J. Wästlund, A generalization of the random assignment problem, preprint, 2000, `arXiv.math.CO/0006146`.

[6] Mézard and G. Parisi, Replicas and optimization, *J. Physiques Lettres* **46** (1985), 771–778.

[7] G. Parisi, A conjecture on random bipartite matching, preprint, 1998, `xxx.lanl.gov/ps/cond-mat/9801176`.

[8] S. Janson, Tomasz Luczak and Andrzej Rucinski, *Random Graphs*, Wiley, New York, 2000.

[9] R. P. Stanley, *Enumerative Combinatorics* vol. 2, Cambridge University Press, Cambridge, 1999.

NADA, KTH, SE-100 44 Stockholm, Sweden
E-mail address: `henrik@nada.kth.se`

IMA, Mälardalens högskola, Box 883, SE-721 23 Västerås, Sweden
E-mail address: `kimmo.eriksson@mdh.se`

NADA, KTH, SE-100 44 Stockholm, Sweden
E-mail address: `jonass@nada.kth.se`