LU TP 97-11
Revised
May 7, 1998

Statistical Properties of Unrestricted Crew Scheduling Problems

Martin Lagerholm[1], Carsten Peterson[2] and Bo Söderberg[3]

Complex Systems Group, Department of Theoretical Physics
University of Lund, Sölvegatan 14A, S-223 62 Lund, Sweden

Submitted to *European Journal of Operational Research*

Abstract:

A statistical analysis is performed for a random unrestricted local crew scheduling problem, expressed in terms of pairing arrivals with departures. The analysis is aimed at understanding the structure of similar problems with global restrictions, and estimating their difficulty. The methods developed are of a general nature and can be of use in other problems with a similar structure. For large random problems, the ground-state energy scales like $\sqrt{N}$ and the average excitation like N, where N is the number of arrivals/departures. The average ground-state degeneracy is such that the probability of hitting an optimal pairing by chance scales like $2N2^{-N}$ for large N. By insisting on the local ground-state energy for a restricted problem, airports can be split into smaller parts, and the state space reduced by typically a factor $\sim 2^{N_a}$, with N_a the total number of airports.

[1]martin@thep.lu.se
[2]carsten@thep.lu.se
[3]bs@thep.lu.se

1 Introduction

Airline crew scheduling represents an important class of optimization problems where the topological structure is important. In ref. [1], a novel Potts artificial neural network approach is developed for attacking semi-realistic airline crew scheduling problems of the following type: A weekly flight schedule is given in terms of a set of flights, each with a specified airport and time for departure and arrival. The object is to assign a crew to each flight, while minimizing a cost-function defined by the total required crew time (including the waiting-time at airports). The solution is subject to a set of global constraints: The crews are required to travel along closed tours, starting and ending at a certain airport, the home-base. These tours are subject to limitations as to duration and leg-count.

As shown in ref. [1], a great deal of simplification is gained by reformulating the problem as that of mapping arrivals on departures at each airport, implying an implicit representation of the crews. In fact, without the global restrictions, the problem is reduced to a set of independent local subproblems, one at each airport. Each local problem amounts to minimizing the local waiting-time, and is simply solvable in polynomial time.

In this paper, we focus on this kind of unrestricted local problems, in particular their statistical properties. These are quite interesting, and in no way trivial, in spite of the triviality of the problems. In particular, we consider the ensemble of *random* local problems of a fixed size N, as defined by the number of arrivals/departures. In addition, we analyze the properties of random solutions to such problems.

Such a statistical analysis of this type of problem does not exist in the literature, and we feel it is interesting for the following reasons: The results illuminate the structure of the corresponding restricted problems, and as a by-product, useful tools are provided for probing their difficulty, and for simplifying their solution. Some of the methods used are novel, and may be used also in other contexts, where a similar structure occurs. In addition, a lower bound to the waiting-time is provided by the solutions to the unrestricted problem. This bound is often saturated [1].

The methodology we use contains the following steps. First, the analysis of a local problem is simplified by considering its topology (defined by the relative ordering in time between arrivals and departures) separately from its geometry (defined by the lengths of the time intervals between consecutive events). The ensemble of problems is thus factorized into the direct product of the *ensemble of topologies* and the *ensemble of geometries.*

After introducing a notation for the topology, we consider problems with a fixed topology, and evaluate averages over the geometry of entities related to the waiting-time. These are simple, since the effect of the geometry on the waiting-time spectrum is a mere shift.

The apparent difficulty of a problem is probed by analyzing the distribution of waiting-times of random solutions. A nice feature is that the waiting-time spectrum for each problem is quantized in steps of the basic period of the schedule.

Subsequently, all variables of interest are averaged also over the topology. This is a more difficult task, and requires the use of a subtle recursive method.

An alternative measure of the difficulty of a problem is the ground-state degeneracy, i.e. the number of solutions with minimal waiting-time, as compared to the total number of solutions, given by $N!$. This is independent of geometry. Due to the character of the dependence on topology, non-standard methods are required to compute the average over topology.

This paper is organized as follows: In Section 2, the problem ensemble is defined. In Section 3, a formalism is introduced for characterizing the topology. The degeneracy structure as a function of topology is analyzed, and various energy moments for fixed topology are computed, by averaging over geometry and/or pairing. In Section 4, a statistical analysis of the detailed degeneracy structure is considered. In particular, the average ground-state degeneracy is computed. Section 5 contains our conclusions.

2 Unrestricted Crew Scheduling

2.1 The Local Problem

A local problem of size N is defined by specifying the times for N arrivals (*arr*'s) and N departures (*dep*'s), denoted respectively by t_i^a and t_i^d, $i = 1 \ldots N$. The object is simply to find a one-to-one mapping (a *pairing*) between the *arr*'s and *dep*'s, such that the *energy* (or objective function) E, given by the total waiting-time, is minimal. In general this can be done in more than one way, implying a degeneracy of the ground-state.

The pairing of an *arr* A with a *dep* D implies that the crew assigned to A should next be assigned to D. The periodicity of the schedule implies that any *arr* may be mapped on any *dep*: if the *dep* is earlier, it is taken as the same *dep* in the next period.

In what follows, we will use the period as the unit of time (and thus energy). Then the times for the *arr*'s and *dep*'s can be restricted to the unit interval. If an *arr* i is mapped on a *dep* j, their contribution to the total waiting-time is given by

$$t_{ij}^w = \left(t_j^d - t_i^a\right) \mod 1 \in [0,1]. \tag{1}$$

Thus, whatever the mapping, the energy E is restricted to the interval $[0, N]$, and between different pairings it can only change by an integer amount. Pairings yielding the lowest possible energy E_0 are said to belong to the ground-state, while a pairing with $E = E_0 + k$, $k > 0$ is said to belong to k:th excited state. In figure 1, an example of a flight schedule is depicted.

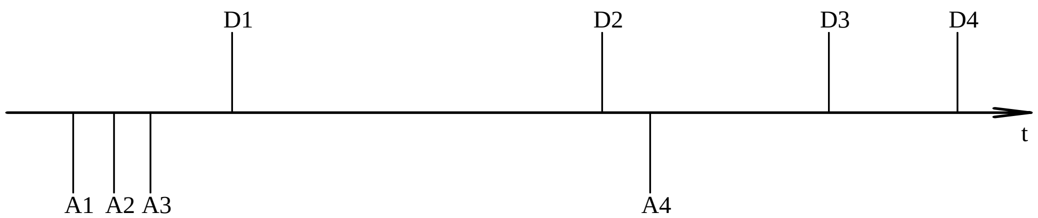

Figure 1: An example of a local flight schedule.

2.2 The Random Local Problem

A random local problem is defined by independently choosing the times for N *arr*'s and N *dep*'s randomly on the unit period. It is useful to divide the characterization of a given problem in two parts: its **topology** and its **geometry**. The topology is defined by the relative ordering in the combined (cyclic) sequence of *arr*'s and *dep*'s. The probability is the same for every distinct topology. The geometry is defined by the sizes of the $2N$ inter-spaces $x_i, i = 1, \ldots, 2N$, into which the period is divided.

For a given problem, the solution space is defined by the set of **pairings**, i.e. the $N!$ possible mappings between *arr*'s and *dep*'s.

We will be interested mainly in the statistical properties of the following entities: the ground-state energy (minimal waiting-time) E_0, the energy E of a random pairing, and the (integer) difference $D = E - E_0$ defining the excitation energy. We will also be interested in the degeneracies of the ground-state and the excited states. To that end, we will consider three kinds of averages: over respectively the *topology*, the *geometry*, and the *pairing*.

The degeneracies of the ground-state and the excited states depend only on the topology, i.e. the combined ordering of *arr*'s and *dep*'s.

For a fixed topology, the excitation energy D is independent of the geometry, and depends entirely on the pairing. Conversely, the ground-state energy E_0 obviously is independent of the pairing, and depends only on the geometry. Thus, for a fixed topology, D and E_O are completely *uncorrelated*, in the combined ensemble of random geometries and random pairings.

3 Analysis for Fixed Topology

3.1 Characterization of the Topology

A simple way to achieve a ground-state pairing (i.e. solve the problem) for a given topology is as follows:

1. An *arr* immediately followed by a *dep* is paired with that *dep*, and both are removed from the sequence. Note that the *dep* could be in the next period.
2. The process is continued until all *arr*'s and *dep*'s are used.

As an example, consider the sequence $[A_1 A_2 A_3 D_1 D_2 A_4 D_3 D_4]$, corresponding to the topology of the schedule in figure 1.

- Pairing A_3 with D_1 leaves $[A_1 A_2 D_2 A_4 D_3 D_4]$.
- Pairing A_2 with D_2 leaves $[A_1 A_4 D_3 D_4]$.

- Pairing A_4 with D_3 leaves $[A_1 D_4]$.
- Pairing A_1 with D_4 finishes the process.

A graphical representation of the topology is now defined as follows.

- If necessary, rotate the sequence such that no pairing crosses the interval border.
- Represent the *arr*'s and *dep*'s by equally spaced points on the interval.
- For each pairing in turn, draw a line from the *arr* to the *dep* on the lowest level (one). All previously drawn lines that overlap with the new line are lifted one level.

Thus a set of lines at different levels are obtained, each line starting at an *arr* and ending on a *dep*. For the example above, the result is shown in fig. 2. Each line represents a crew waiting for a *dep*.

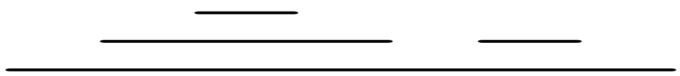

Figure 2: Graphical representation of the topology $[AAADDADD]$

It turns out that for the properties of the total energy spectrum, full knowledge of the topology is not necessary; it suffices to know how many lines there are at the different levels. Thus let P_k be the number of lines at level k ($k = 1, 2, 3, \ldots$). For the example above, the P sequence is $[1, 2, 1]$, i.e. $P_1 = 1, P_2 = 2, P_3 = 1$, while $P_k = 0$ for $k > 3$.

3.2 Degeneracy Structure

The ground-state degeneracy is now given by the product of the line-levels, i.e.

$$g_0 = \prod_{k \geq 1} k^{P_k}. \tag{2}$$

This is because each *dep* must terminate a line, and the number of available crews is equal to the number of lines alive at that point, which is given by the line's level. For the example we get $3 \times 2 \times 2 \times 1 = 12$, corresponding to half of the 24 possible pairings.

Naively, the degeneracy of the mth excited state can be obtained by adding m dummy lines, covering the entire interval; they represent extra crews. Then each *dep* has m additional crews to choose between, and we get for the naive degeneracy[4]

$$a_m = \prod_{k \geq 1} (k + m)^{P_k}, \tag{3}$$

[4] For the example above, this precisely corresponds to the spectrum $\mathbf{O}|n> = a_n|n>$ of the quantum-mechanical operator $\mathbf{O} = \mathbf{a}\mathbf{a}\mathbf{a}\mathbf{a}^\dagger\mathbf{a}^\dagger\mathbf{a}\mathbf{a}^\dagger\mathbf{a}^\dagger$, where $\mathbf{a}^\dagger, \mathbf{a}$ are harmonic oscillator creation and annihilation operators, satisfying $[\mathbf{a}, \mathbf{a}^\dagger] = 1$, and $|n> \propto \left(\mathbf{a}^\dagger\right)^n |0>$ is the n:th excited state.

where we have assumed indistinguishable dummy lines; otherwise we would have an extra factor $m!$. This defines an infinite sequence.

However, some of the possible pairings will contain permanently grounded crews (closed lines), or lines extending over more than one period. The contribution to the naive multiplicity from solutions with n grounded crews and/or excessive periods depends on the proper degeneracy n steps down. Denoting the **proper degeneracies** by g_m, the relation is

$$a_m = \sum_{n=0}^{m} g_{m-n} \left(\begin{array}{c} N+n \\ n \end{array} \right), \tag{4}$$

where the last factor is a binomial coefficient. This represents a kind of renormalization, and can be inverted to yield the proper degeneracies

$$g_m = \sum_{n=0}^{m} a_{m-n} (-)^n \left(\begin{array}{c} N+1 \\ n \end{array} \right). \tag{5}$$

This must define a finite sequence, since $g_m \geq 0$ and the total number of pairings, $\sum_m g_m = N!$, is finite.

From eq. (3), it is obvious that the naive degeneracy a_m is an Nth degree polynomial in m; hence it can be written as

$$a_m = \sum_{k=0}^{N} c_k \left(\begin{array}{c} m+k \\ k \end{array} \right), \tag{6}$$

with some coefficients c_k. Define the *generating functions*

$$A(x) = \sum_m a_m x^m \ , \ \ G(x) = \sum_m g_m x^m. \tag{7}$$

Due to eq. (5), these are related by $G(x) = (1-x)^{N+1} A(x)$. Then, in terms of c_k, we have

$$A(x) \ = \ \sum_{k=0}^{N} c_k (1-x)^{-k-1}, \tag{8}$$

$$G(x) \ = \ \sum_{k=0}^{N} c_k (1-x)^{N-k}. \tag{9}$$

From this, we see that the g sequence is indeed finite: $g(m) = 0$ for all $m > N$. The individual degeneracies g_m can be obtained from G and its derivatives at $x = 0$,

$$\begin{aligned} g_0 \ &= \ G(0) \ = \ \sum_k c_k, \\ g_1 \ &= \ G'(0) \ = \ -\sum_{k=0}^{N} (N-k) c_k, \end{aligned} \tag{10}$$

etc.

3.3 The Excitation Energy for a Random Pairing

Conversely, the moments over g_m are readily obtained from G and its derivatives at $x = 1$,

$$\begin{aligned} \sum_m g_m &= G(1) = c_N, \\ \sum_m m g_m &= G'(1) = -c_{N-1}, \\ \sum_m m(m-1) g_m &= G''(1) = 2c_{N-2}, \end{aligned} \tag{11}$$

etc. In order to relate c_k to P_k, we can express the coefficients of the polynomial a_m in two different ways. From eq. (3) we get for the leading coefficients

$$a_m = m^N + m^{N-1} \sum_k k P_k + \frac{m^{N-2}}{2} \left\{ \left(\sum_k k P_k \right)^2 - \sum_k k^2 P_k \right\} + \ldots, \tag{12}$$

while eq. (6) gives, upon expanding the binomial coefficients,

$$\begin{aligned} a_m &= \frac{m^N}{N!} c_N + \frac{m^{N-1}}{N!} \left\{ \frac{1}{2} N(N+1) c_N + N c_{N-1} \right\} \\ &+ \frac{m^{N-2}}{N!} \left\{ \frac{1}{24} N(N-1)(N+1)(3N+2) c_N + \frac{1}{2} N^2 (N-1) c_{N-1} + N(N-1) c_{N-2} \right\} + \ldots \end{aligned} \tag{13}$$

From this we obtain the relations

$$\begin{aligned} c_N &= N!, \\ c_{N-1} &= N! \left\{ \frac{1}{N} \sum_k k P_k - \frac{N+1}{2} \right\}, \\ c_{N-2} &= N! \left\{ \frac{1}{2N(N-1)} \left(\sum_k k P_k \right)^2 - \frac{1}{2N(N-1)} \sum_k k^2 P_k - \frac{1}{2} \sum_k k P_k + \frac{(N+1)(3N-2)}{24} \right\}, \end{aligned} \tag{14}$$

etc. This gives (as it should) $\sum_m g_m = N!$, the number of possible pairings. In a given topology, the fraction of pairings having an energy $D = m$ steps above E_0 is given by $g_m/N!$. Thus, the first few moments of the excitation energy D for a random pairing are:

$$\begin{aligned} < D >_p &= \frac{1}{N!} \sum_m m g_m = \frac{N+1}{2} - \frac{1}{N} \sum_k k P_k, \\ < D^2 >_p &= \frac{1}{N!} \sum_m m^2 g_m \\ &= \frac{1}{N(N-1)} \left(\sum_k k P_k \right)^2 - \frac{1}{N(N-1)} \sum_k k^2 P_k - \frac{N+1}{N} \sum_k k P_k + \frac{(N+1)(3N+4)}{12}, \end{aligned} \tag{15}$$

where $<>_p$ denotes the average over pairings for a fixed topology.

3.4 The Ground-State Energy in a Random Geometry

For a given topology, the ground-state energy depends on the geometry, and is given by

$$E_0 = \sum_{i=1}^{2N} k_i x_i , \tag{16}$$

where x_i is the length of the ith sub-interval, and k_i the number of lines in that interval.

To perform averages over a random geometry, we need to analyze the distribution of the intervals x_i. Independently of the topology, they obey the distribution

$$dP = (2N-1)!\,\delta\left(\sum_i x_i - 1\right) \prod_{i=1}^{2N} \left(\Theta(x_i) dx_i\right). \tag{17}$$

For a single x_i, this implies the distribution

$$f(x_i) = (2N-1)\,(1-x_i)^{2N-2}\ ,\ 0 \le x_i \le 1. \tag{18}$$

Using the identity $\sum_i x_i = 1$, and the permutation symmetry between different x_i, we have

$$< x_i > \ = \ \frac{1}{2N}, \tag{19}$$

$$< x_i x_j > \ = \ \frac{1+\delta_{ij}}{2N(2N+1)}. \tag{20}$$

In what follows, we will also need the number of intervals with k lines, to be denoted by Q_k; in terms of P_k (defining $P_0 = 0$) it is simply

$$Q_k = P_k + P_{k+1}\,,\ k \ge 0, \tag{21}$$

since adding a line on level k implies adding two intervals, with respectively k and $k-1$ lines. Obviously, we have $\sum_k Q_k = 2N$.

Now we are ready to compute the first few moments of E_0, yielding

$$\begin{aligned}
< E_0 >_g &= \sum_{i=1}^{2N} k_i < x_i >_g = \frac{1}{2N}\sum_i k_i = \frac{1}{2N}\sum_k kQ_k = \frac{1}{N}\sum_k kP_k - \frac{1}{2}, \\
< E_0^2 >_g &= \sum_{ij} k_i k_j < x_i x_j > = \frac{1}{2N(2N+1)} \sum_{ij} k_i k_j (1+\delta_{ij}) \\
&= \frac{1}{2N(2N+1)} \left(\sum_{kl} klQ_kQ_l + \sum_k k^2 Q_k \right) \\
&= \frac{1}{2N(2N+1)} \left(4\left(\sum_k kP_k\right)^2 + 2\sum_k k^2 P_k - 2(2N+1)\sum_k kP_k + N(N+1) \right),
\end{aligned} \tag{22}$$

where $<>_g$ denotes average over the geometry for fixed topology.

3.5 The Total Energy

The total energy E is given by the ground-state and excitation energies, $E = E_0 + D$. Averaging over both geometry and pairing, we obtain the surprisingly simple result

$$< E >_{gp} = < E_0 >_g + < D >_p = \frac{N}{2}, \tag{23}$$

independently of topology.

Since we already have the second moments of E_0 and D, we only need $< E_0 D >$ to compute the second moment of E. Using the fact that E_0 and D are statistically independent for a fixed topology, we get

$$< E_0 D >_{gp} = < E_0 >_g < D >_p = -\frac{1}{N^2}\left(\sum_k kP_k\right)^2 + \frac{N+2}{2N}\sum_k kP_k - \frac{N+1}{4}. \tag{24}$$

This implies

$$< E^2 >_{gp} = \frac{(N+1)(6N^2-N+4)}{12(2N+1)} - \frac{(N+2)}{N^2(N-1)(2N+1)}\left(N\sum_k k^2 P_k - \left(\sum_k kP_k\right)^2\right) \tag{25}$$

4 Topology Statistics

4.1 The Topology Ensemble

We now want to perform averages also over the topology, for the moments of the ground-state energy E_0, the excitation energy D, and the total energy E. In table 1 a complete list of topologies for N up to four is given, along with various characteristics. Note the similarity in symmetry-properties between the P, Q and g sequences.

To compute $< P_k >$, we must analyze the number of ways $m[P]$ to obtain a given P sequence.

- The lowest level lines define P_1 non-overlapping subgroups; these are cyclically indistinguishable, so the naive multiplicity must be divided by P_1.
- For the P_k indistinguishable lines on a higher level k, there are P_{k-1} possible lines on the previous level to put them above. This can be done in $\begin{pmatrix} P_k + P_{k-1} - 1 \\ P_k \end{pmatrix}$ ways.
- In addition, there are $2N$ cyclic rotations of the AD sequence that yield the same topology.
- The P sequence must sum up to N, and P_1 must not vanish.

This gives for the multiplicity $m[P]$

$$m[P] = \frac{2N}{P_1}\delta_{N-\sum P_k} \prod_{k=1}^{\infty} \begin{pmatrix} P_k + P_{k+1} - 1 \\ P_{k+1} \end{pmatrix}. \tag{26}$$

The total number of possible topologies for a fixed N is simply the number of possible AD sequences, i.e. the number of distinct orderings of a sequence of N A's and N D's. This is given by $\begin{pmatrix} 2N \\ N \end{pmatrix}$. This should match the total multiplicity $M_N \equiv \sum_{[P]} m([P])$, which can be recursively obtained from $m[P]$ by expanding the Kronecker delta in terms of a complex integral along a *small* contour C around the origin:

$$\delta_{N-\sum P_k} = \oint_C \frac{dz}{2\pi i z^{N+1}} \prod z^{P_k}. \tag{27}$$

By iteratively using the identity

$$z^a \sum_{b=0}^{\infty} \begin{pmatrix} a+b-1 \\ a-1 \end{pmatrix} w^b = z^a \sum_{b=0}^{\infty} \begin{pmatrix} -a \\ b \end{pmatrix} (-w)^b = \left(\frac{z}{1-w}\right)^a. \tag{28}$$

when summing $m[P]\prod z^{P_k}$ over the P_k in order of decreasing k, we obtain convergence of the power factor to ω^{P_k}, with ω given by

$$\omega = \frac{z}{1-\omega} \;\Rightarrow\; \omega = \frac{1}{2}\left(1 - \sqrt{1-4z}\right). \tag{29}$$

N	Pattern	Prob.	$\{P_k\}$	$\{Q_k\}$	$\{a_i\}$	$\{g_i\}$	$\langle E_0\rangle$	$\langle D\rangle$
1	AD	1	[1]	[1,1]	1,2,3,...	(1)	1/2	0
2	AADD	2/3	[1,1]	[1,2,1]	2,6,12,...	(2)	2/2	0
2	ADAD	1/3	[2]	[2,2]	1,4,9,...	(1,1)	1/2	1/2
3	AAADDD	3/10	[1,1,1]	[1,2,2,1]	6,24,60,...	(6)	9/6	0
3	AADADD	3/10	[1,2]	[1,3,2]	4,18,48,...	(4,2)	7/6	1/3
3	AADDAD	3/10	[2,1]	[2,3,1]	2,12,36,...	(2,4)	5/6	2/3
3	ADADAD	1/10	[3]	[3,3]	1,8,27,...	(1,4,1)	3/6	3/3
4	AAAADDDD	4/35	[1,1,1,1]	[1,2,2,2,1]	24,120,360,...	(24)	8/4	0
4	AAADADDD	4/35	[1,1,2]	[1,2,3,2]	18,96,300,...	(18,6)	7/4	1/4
4	AAADDADD	4/35	[1,2,1]	[1,3,3,1]	12,72,240,...	(12,12)	6/4	2/4
4	AADAADDD	4/35						
4	AAADDDAD	4/35	[2,1,1]	[2,3,2,1]	6,48,180,...	(6,18)	5/4	3/4
4	AADADADD	4/35	[1,3]	[1,4,3]	8,54,192,...	(8,14,2)	5/4	3/4
4	AADDAADD	2/35	[2,2]	[2,4,2]	4,36,144,...	(4,16,4)	4/4	4/4
4	AADADDAD	4/35						
4	AADDADAD	4/35	[3,1]	[3,4.1]	2,24,108,...	(2,14,8)	3/4	5/4
4	ADADADAD	1/35	[4]	[4,4]	1,16,81,256,...	(1,11,11,1)	2/4	6/4

Table 1: Inequivalent topologies for $N \leq 4$. For each topology, the third column gives the probability of occurrence, while the next two give its P_k and Q_k sequence, respectively. In the following two columns the a_i and g_i sequences are given, whereas the last two columns give the average ground-state energy and excitation, respectively. Note that different topologies might have identical characteristics in terms of P_k, etc.

Thus, the total multiplicity is given by

$$M_N = 2N \oint_C \frac{dz\ z^{-N}}{2\pi i z} \sum_{P_1} \frac{\omega(z)^{P_1}}{P_1} = 2N \oint_C \frac{dz\ z^{-N}}{2\pi i z} (-\log(1-\omega(z))) = \begin{pmatrix} 2N \\ N \end{pmatrix}, \tag{30}$$

obtained by extracting the z^N coefficient of $-\log(1-\omega(z))$ as $\frac{1}{2N}\begin{pmatrix} 2N \\ N \end{pmatrix}$, by doing the integral in terms of w using $z = w(1-w)$.

4.2 Distribution of P_1

The number of non-overlapping groups of lines is given by the number of lines at level 1, i.e. P_1. In fact, $w(z)$ can be seen as a generating function,

$$w(z) = \sum_{n=1}^{\infty} m_n^{(1)} z^n = -\frac{1}{2}\sum_{n=1}^{\infty}(4z)^n \begin{pmatrix} n-\frac{3}{2} \\ n \end{pmatrix} = \frac{1}{2}\sum_{n=1}^{\infty}(2z)^n \frac{(2n-3)!!}{n!}, \tag{31}$$

for the number of ways ($m_n^{(1)}$) to arrange n lines in such a group. The distribution of P_1 is easy to compute by slightly modifying eq. (30): skipping the summation over P_1 and dividing by M_N yields

$$\mathrm{Prob}(P_1) = \frac{2N}{\begin{pmatrix} 2N \\ N \end{pmatrix}} \oint_C \frac{dz\ z^{-N}}{2\pi i z} \frac{\omega(z)^{P_1}}{P_1} = \frac{N!(2N-1-P_1)!}{(2N-1)!(N-P_1)!} \tag{32}$$

for $P_1 > 0$. For large $P_1 \ll N$, this is close to 2^{-P_1}.

The grouping of lines corresponds to a grouping of the flights, with a matching number of *arr*'s and *dep*'s in each group. In a ground-state pairing, flights in different groups are never paired, and within a group, the *arr* has to precede its paired *dep*. This fact can be used to simplify also restricted problems.

4.3 Moments of P_k

Similarly, $< P_k >$ can be obtained from realizing that inserting a factor P_k in the sum over P_k gives a factor $P_{k-1}\omega/(1-\omega)$. This yields in the end an extra factor $P_1(\omega/(1-\omega))^{k-1}$ in the P_1 sum:

$$\begin{aligned} M_N < P_k > &= 2N \oint_C \frac{dz}{2\pi i z^{N+1}} \left(\frac{\omega}{1-\omega}\right)^{k-1} \sum_{P_1>1} \omega^{P_1} && (33) \\ &= 2N \oint_C \frac{dz}{2\pi i z^{N+1}} \left(\frac{\omega}{1-\omega}\right)^{k} && (34) \\ &= 2N \oint_C \frac{d\omega(1-2\omega)}{2\pi i \omega^{N-k+1}(1-\omega)^{N+k+1}}, && (35) \end{aligned}$$

where the last expression is a reformulation in terms of a loop integral around $\omega = 0$. In a similar way $< P_k P_l >$ etc. can be computed. We obtain

$$< P_k > \;=\; \frac{2N}{M_N}\left\{\left(\begin{array}{c} 2N-1 \\ N-k \end{array}\right) - \left(\begin{array}{c} 2N-1 \\ N-k-1 \end{array}\right)\right\} = \frac{2k}{M_N}\left(\begin{array}{c} 2N \\ N-k \end{array}\right), \tag{36}$$

$$< P_k P_m > \;=\; \frac{2N}{M_N}\left\{\left(\begin{array}{c} 2N \\ N-m \end{array}\right) - \left(\begin{array}{c} 2N \\ N-k-m \end{array}\right)\right\},\; k \leq m. \tag{37}$$

From these expressions, we can derive the following particular averages, needed to compute the various energy moments over the topology:

$$\begin{aligned} \left\langle \sum_k k P_k \right\rangle &= \frac{N}{2} 4^N \left(\begin{array}{c} 2N \\ N \end{array}\right)^{-1} \approx \frac{N}{2}\sqrt{\pi N}\left(1 + \frac{1}{8N} + \ldots\right), \\ \left\langle \sum_k k^2 P_k \right\rangle &= N^2, \\ \left\langle \left(\sum_k k P_k\right)^2 \right\rangle &= \frac{N^2(5N+1)}{6}, \end{aligned} \tag{38}$$

where the approximate form in the top equation is valid for large N.

4.4 Full Energy Averages

We are now ready to compute the final averages also over the topology. Inserting the results of eqs. (38) into eqs. (22), we obtain for the moments of the ground-state energy E_0 of a random problem:

$$\begin{aligned} < E_0 > &= \frac{1}{2} 4^N \left(\begin{array}{c} 2N \\ N \end{array}\right)^{-1} - \frac{1}{2} \approx \frac{1}{2}\sqrt{\pi N} - \frac{1}{2}, \\ < E_0^2 > &= \frac{5N}{6} - \frac{1}{2} 4^N \left(\begin{array}{c} 2N \\ N \end{array}\right)^{-1} + \frac{1}{2} \approx \frac{5N}{6}, \\ < E_0^2 >_c &\approx \frac{(10-3\pi)N}{12}, \end{aligned} \tag{39}$$

where $\langle ab \rangle_c = \langle ab \rangle - \langle a \rangle \langle b \rangle$ is the connected moment; the approximate forms are valid for large N.

Similarly, from eqs. (15) we get for the moments of the excitation energy D, for a random pairing in a random topology,

$$< D > \;=\; \frac{N+1}{2} - \frac{1}{2} 4^N \left(\begin{array}{c} 2N \\ N \end{array}\right)^{-1} \approx \frac{N}{2} - \frac{1}{2}\sqrt{\pi N} + \frac{1}{2}, \tag{40}$$

$$\begin{aligned} < D^2 > &= \frac{N^2}{4} + \frac{17N}{12} + \frac{1}{3} - \frac{N+1}{2} 4^N \left(\begin{array}{c} 2N \\ N \end{array}\right)^{-1} \\ &\approx \frac{N^2}{4} - \frac{N}{2}\sqrt{\pi N} + \frac{17N}{12}, \end{aligned} \tag{41}$$

$$< D^2 >_c \;\approx\; \frac{(11-3\pi)N}{12}. \tag{42}$$

Averaging the combined moment, eq. (24), over topology yields

$$< E_0 D > \;=\; \frac{N+2}{4} 4^N \begin{pmatrix} 2N \\ N \end{pmatrix}^{-1} - \frac{13N+5}{12} \approx \frac{N}{4}\sqrt{\pi N} - \frac{13N}{12}, \tag{43}$$

$$< E_0 D >_c \;=\; -\frac{5N+1}{6} + \frac{1}{4} 16^N \begin{pmatrix} 2N \\ N \end{pmatrix}^{-2} \approx -\frac{(10-3\pi)N}{12}. \tag{44}$$

Combining this with the moments of E_0 and D, we get for the moments of the total energy, $E = E0 + D$, the following simple results

$$< E > \;=\; \frac{N}{2}, \tag{45}$$

$$< E^2 > \;=\; \frac{N^2}{4} + \frac{N}{12}, \tag{46}$$

$$< E^2 >_c \;=\; \frac{N}{12}, \tag{47}$$

which can be understood by noting that a random pairing in a random topology corresponds to a set of N lines with random endpoints. Then each line has a uniform length distribution between 0 and 1, and E is their total length.

Computing the corresponding standard deviations, we have for a typical random problem (and a random pairing, for D and E) at large N,

$$E_0 \;\sim\; \frac{1}{2}\sqrt{\pi N} \;\pm\; \frac{1}{2}\sqrt{\left(\frac{10}{3} - \pi\right) N}, \tag{48}$$

$$D \;\sim\; \frac{N}{2} - \frac{1}{2}\sqrt{\pi N} \;\pm\; \frac{1}{2}\sqrt{\left(\frac{11}{3} - \pi\right) N}, \tag{49}$$

$$E \;\sim\; \frac{N}{2} \;\pm\; \frac{1}{2}\sqrt{\frac{N}{3}}, \tag{50}$$

and we see that E scales as $\sqrt{N}$, while E and D scale as N, while the standard deviation scales as $\sqrt{N}$ in each case.

Of interest are also the correlations at large N, given to order N by

$$\begin{aligned} < E_0 D >_c \;&\sim\; -\frac{N}{12}(10 - 3\pi), \\ < E_0 E >_c \;&\sim\; 0, \\ < D E >_c \;&\sim\; \frac{N}{12}, \end{aligned} \tag{51}$$

indicating that E and E_0 become uncorrelated for a random pairing of a large random problem.

4.5 Statistics for Individual Degeneracies

An interesting but more difficult thing to compute is the average fraction $\gamma_n =< g_n > /N!$ of pairings having a given excitation energy $D = n$, in particular the ground-state fraction γ_0, which gives the average probability of hitting a ground-state by chance. Since g_n is simply related to a_n, we will start by considering

$$\alpha_n = \frac{\begin{pmatrix} 2N \\ N \end{pmatrix}}{2N} < a_n >, \tag{52}$$

in terms of which γ_n can be expressed as

$$\gamma_n = \frac{N!}{(2N-1)!} \sum_m (-)^m \begin{pmatrix} N+1 \\ m \end{pmatrix} \alpha_{n-m}. \tag{53}$$

We then have

$$\alpha_n = \frac{1}{2N} \sum_{[P]} m[P] \prod_k (n+k)^{P_k} = \sum_{[P]} \frac{1}{P_1} \delta_{N-\sum P_k} \prod_{k\geq 1} (n+k)^{P_k} \begin{pmatrix} P_k + P_{k+1} - 1 \\ P_{k+1} \end{pmatrix}. \tag{54}$$

A generating function for the N-dependence of α_n is then

$$A_n(z) \equiv \sum_N z^N \alpha_n(N) = \sum_{[P]} \frac{1}{P_1} \prod_{k\geq 1} [z(n+k)]^{P_k} \begin{pmatrix} P_k + P_{k+1} - 1 \\ P_{k+1} \end{pmatrix}. \tag{55}$$

Again, starting the P_k summation at a large k_0 and proceeding in order of decreasing k, gives convergence of the full sum in the limit $k_0 \to \infty$. Denoting by $\omega_{k+n-1}^{P_k}$ the result of summing above a certain k, we have by eq. (28),

$$\omega_{k-1} = \frac{kz}{1-w_k}, \tag{56}$$

and the final result is

$$A_n(z) = \sum_{P_1} \frac{1}{P_1} \omega_n^{P_1} = -\log\left(1 - \omega_n(z)\right). \tag{57}$$

The recursion relation (56) can now be linearized by assuming $\omega_k = p_k/q_k$, which can be solved e.g. by

$$q_{k-1} = q_k - p_k, \tag{58}$$

$$p_{k-1} = kzq_k, \tag{59}$$

which, upon eliminating p_k gives

$$kzq_k - q_{k-1} + q_{k-2} = 0. \tag{60}$$

By partial integration, it is simple to prove that the following sequence of integrals solves the recursion relation (60) for q_k,

$$q_k = \mathrm{Im} \int_0^\infty \frac{(-iy)^k}{k!} \exp\left(-z\frac{y^2}{2} + iy\right) dy \ , \ k \geq 0, \ z > 0, \tag{61}$$

which can be extended to negative k by recursion. The q_k can be expanded in z as

$$q_k = \begin{cases} \sum_{m=0}^{\infty} \begin{pmatrix} 2m+k \\ 2m \end{pmatrix} (2m-1)!!z^m \ , \ k \geq 0, \\ \sum_{m=0}^{[-(k+1)/2]} \begin{pmatrix} -k-1 \\ 2m \end{pmatrix} (2m-1)!!z^m \ , \ k < 0, \end{cases} \tag{62}$$

where [] denotes integer part. Note that for non-negative k the series is an asymptotic one, while for negative k it is finite. In particular, we have $q_{-1} = 1$. An independent solution to (60) is given by

$$\hat{q}_k(z) = \frac{z^{-k}}{k!} q_{-k-1}(-z), \tag{63}$$

but this solution is irrelevant, having the wrong large-k behaviour.

In terms of q_k we now have

$$1 - \omega_k = q_{k-1}/q_k, \tag{64}$$
$$A_n = \log(q_n) - \log(q_{n-1}), \tag{65}$$

which should then be expanded in powers of z to yield $\alpha_n(N)$ as the coefficient of z^N. In particular, we have $A_0 = \log(q_0)$, where $q_0 = 1 + z + 3z^2 + 15z^3 + \ldots$.

In table 2, results are displayed for the average degeneracy of the two lowest states, based on an expansion of A_0 and A_1 in powers of z. It is easy to check for small N, using table 1, that $g_m / \sum g_k$

N	$K_N = \frac{1}{2}\binom{2N}{N}$	$K_N < g_0 >$	$K_N < g_1 >$	$< g_0 >$	$< g_1 >$	γ_0	γ_1
1	1	1	0	1.00000	0.00000	1.000000	0.000000
2	3	5	1	1.66667	0.33333	0.833333	0.166667
3	10	37	22	3.70000	2.20000	0.616667	0.366667
4	35	353	411	10.0857	11.7429	0.420238	0.489286
5	126	4081	7676	32.3889	60.9206	0.269907	0.507672
6	462	55205	149741	119.491	324.115	0.165960	0.450159
7	1716	854197	3.09875×10^6	497.784	1805.80	0.098767	0.358294
8	6435	1.4876×10^7	6.84187×10^7	2311.74	10632.3	0.057335	0.263697
9	24310	2.88019×10^8	1.61447×10^9	11847.7	66411.7	0.032649	0.183013
10	92378	6.13891×10^9	4.07031×10^{10}	66454.3	440614.	0.018313	0.121421
11	352716	1.42882×10^{11}	1.09496×10^{12}	405092.	3.10436×10^6	0.010148	0.077771
12	1.35208×10^6	3.60668×10^{12}	3.13708×10^{13}	2.66751×10^6	2.32019×10^7	0.005569	0.048438
13	5.20030×10^6	9.81584×10^{13}	9.55147×10^{14}	1.88755×10^7	1.83672×10^8	0.003031	0.029496
14	2.00583×10^7	2.86562×10^{15}	3.08337×10^{16}	1.42865×10^8	1.5372×10^9	0.001639	0.017633

Table 2: Results for the average degeneracy of the lowest energy-states for various system sizes N. The second column gives an integer normalization factor K_N. Dividing the integers in the next two columns by K_N yields the average number of ground-states $< g_0 >$ and first excited states $< g_1 >$, respectively. Dividing these by $N!$ yields γ_0 and γ_1.

for $m = 0, 1$, averaged over topologies with the proper probabilities, indeed agrees with γ_m of table 2, obtained from the expansion of A_k.

The result for γ_0 strongly indicates an asymptotic behaviour of $\gamma_0 \sim 2N2^{-N}$. This corresponds to an exponential decrease with N of the average probability for a random pairing to hit a ground-state. However, the average number of ground-states grows faster than exponentially: $< g_0 > \sim 2N\, N!\, 2^{-N}$.

This abundance of ground-states indicates that a corresponding restricted problem might well have a solution with a locally minimal waiting-time, if the restrictions are not too severe; this is used in ref. [1] to simplify the solution of a set of restricted crew scheduling problems.

By reducing the state-space of a restricted problem to the set of ground-states of the corresponding unrestricted problem, the average information gained at each airport is given by $\log \gamma_0$, which for a large airport roughly yields $N \log 2$. Summing this over several airports yields a total information gain scaling as $N_f \log 2$, with N_f the total number of flights. Partly, this is due to a grouping of flights, which contributes an average information gain of $N_a \log 2$, with N_a the number of airports.

5 Conclusions

We have performed a statistical analysis of an ensemble of random unrestricted local crew scheduling problems, formulated in terms of mapping arrivals onto departures at a single airport so as to minimize waiting-time.

For the ground-state energy E_0 of a large random problem, we find that both the average and the fluctuations scale like $\sqrt{N}$. For a random pairing of such a problem, on the other hand, the excitation D and the total energy $E = E_0 + D$ both grow linearly with N, with fluctuations scaling like $\sqrt{N}$.

The individual degeneracies of the lowest energy states for random problems are such that the average probability for hitting an optimal pairing by chance decreases like $2N2^{-N}$ for large N. Since the total number of pairings grows like $N!$, the average number of ground-states grows very fast with system size.

The results and the methods of analysis described in this paper are useful when designing efficient algorithms for the crew scheduling problems with global restrictions, by providing means for estimating the difficulty of a given problem, and for understanding and simplifying its structure.

The optimal crew waiting-time for a restricted problem is bounded from below by the ground-state energy of the corresponding unrestricted problem, which is useful for gauging algorithmic performance for problems of realistic size. Due to a faster than exponential growth of the number of ground-states with problem size, this bound is often saturated. This can be used to simplify a restricted problem: By insisting on the local ground-state energy, airports can be split into several parts. For a large random problem, this results on the average in a reduction of state-space size by a factor of two for each airport.

Some of the calculations in this paper are based on novel methods of a general nature, that may have applications also in other contexts with a similar topological structure.

References

[1] M. Lagerholm, C. Peterson and B. Söderberg, "Airline Crew Scheduling using Potts Mean Field Techniques", *LU TP 97-10* (submitted to *European Journal of Operational Research*).