The Random Hypergraph Assignment Problem

Ralf Borndörfer
Zuse Institute Berlin, Takustr. 7, 14195 Berlin, Germany
E-mail: borndoerfer@zib.de

Olga Heismann
Zuse Institute Berlin, Takustr. 7, 14195 Berlin, Germany
E-mail: heismann@zib.de

Keywords: assignment, hyperassignment, random, expected value

Received: July 13, 2014

Parisi's famous (proven) conjecture states that the expected cost of an optimal assignment in a complete bipartite graph on $n + n$ nodes with i. i. d. exponential edge costs with mean 1 is $\sum_{i=1}^{n} 1/i^2$, which converges to an asymptotic limit of $\pi^2/6$ as n tends to infinity. We consider a generalization of this question to complete "partitioned" bipartite hypergraphs $G_{2,n}$ that contain edges of size two and proper hyperedges of size four. We conjecture that for i. i. d. uniform hyperedge costs on $[0, 1]$ and i. i. d. exponential hyperedge costs with mean 1, optimal assignments expectedly contain half of the maximum possible number of proper hyperedges. We prove that under the assumption of this number of proper hyperedges the asymptotic expected minimum cost of a hyperassignment lies between 0.3718 and 1.8310 if hyperedge costs are i. i. d. exponential random variables with mean 1. We also consider an application-inspired cost function which favors proper hyperedges over edges by means of an edge penalty parameter p. We show how results for an arbitrary p can be deduced from results for $p = 0$.

Povzetek: V članku je opisana analiza komplektnosti dvojno povezanih grafov na osnovi razširitve Parisijevega izreka.

1 Introduction

A way to gain a better understanding of the structure of a combinatorial optimization problem is to analyze the optimal values of random instances. For the assignment problem, such results were conjectured after extensive computational experiments and then proven theoretically. In particular, the famous (proven) Conjectures of Mézard and Parisi [Mézard and Parisi, 1985] state that the expected optimal cost value of an assignment problem on a complete bipartite graph with i. i. d. uniform edge costs on $[0, 1]$ or i. i. d. exponential edge costs with mean 1 converges to $\frac{\pi^2}{6} = 1.6449\ldots$ if the number of vertices tends to infinity. The limit is equal for both distributions since it can be proven that only the density at 0 is relevant, which coincides for both distributions [Aldous, 1992]. For a survey on the *random assignment problem* and several of its variants, see [Krokhmal and Pardalos, 2009].

We consider a generalization of this setting to a class of *bipartite hypergraphs* in terms of what we call the random *hypergraph assignment problem* (HAP). This problem is an idealized version of vehicle rotation planning problems in long-distance passenger rail transport, see [Borndörfer et al., 2011] for further details and [Maróti, 2006] for a survey on railway vehicle rotation planning.

We will deal with HAPs in a special well-structured type of bipartite hypergraphs $G_{2,n}$, that contain on each side n "parts" of size 2 each. In this case, the HAP is already NP-hard [Borndörfer and Heismann, 2012] and therefore interesting to analyze. The hyperedge set of such a partitioned hypergraph $G_{2,n}$ consists only of edges of size 2 and proper hyperedges of size 4, and it has a structure that makes it easy to view a hyperassignment as a combination of two assignments, one consisting only of edges, and the other one consisting only of proper hyperedges (that can also be viewed as edges). Despite this simple general idea, however, combining the two assignments involves a choice over an exponential number of possibilities which is quite difficult to analyze. We will explain this in more detail in Section 2 after introducing the problem.

In Section 3, we conjecture that the expected number of proper hyperedges in an optimal solution of the random HAP on partitioned hypergraphs $G_{2,2n}$ with i. i. d. uniform random edge costs on $[0, 1]$ or i. i. d. exponential random edge costs with mean 1 is n. This conjecture is based on extensive computational results. Assuming that this conjecture holds, we can prove a lower bound of 0.3718 and an upper bound of 1.8310 for the expected value of a minimum cost hyperassignment in $G_{2,2n}$ for the exponential edge cost distribution and for vertex numbers tending to infinity. To achieve this, we first use a combinatorial argument to represent the bounds in terms of bounds for random assignments. Then, we compute these bounds using results for the random assignment problem.

In hypergraph assignment problems that arise from practical applications, proper hyperedges represent unions of edges. Such hyperedges have costs that are smaller than the sum of the costs of the edges that they contain; these edges are considered to be similar and a solution with much similarity is desirable [Borndörfer et al., 2011]. We consider a setting with *regularity-rewarding* cost functions, in which the number of proper hyperedges in a solution and the optimal value of a random HAP in $G_{2,n}$ do not only depend on the number of vertices n but also on an edge *penalty parameter* p. We will show how the number of proper hyperedges and the value of an optimal solution for every p can be deduced from results for $p = 0$ in Section 4.

The paper ends in Section 5 with a discussion of the results.

A short conference version of this paper has already been published as [Heismann and Borndörfer, 2013].

2 The hypergraph assignment problem

We consider in this paper hypergraph assignment problems on a special type of bipartite hypergraphs.

Definition 2.1. Let $G_{2,n} = (U, V, E)$ be the *bipartite hypergraph* with *vertex sets*

$$U = \bigcup_{i=1}^{n} U_i, \qquad V = \bigcup_{i=1}^{n} V_i$$

with

$$U_i = \{u_i, u_i'\}, \qquad V_i = \{v_i, v_i'\}$$

and *hyperedge set* $E = E_1 \cup E_2$ where

$$E_1 = \{\{u, v\} : u \in U, v \in V\}$$

are the *edges* and

$$E_2 = \{U_i \cup V_j : i, j \in \{1, \ldots, n\}\}$$

are the *proper hyperedges* of size 4. The sets U_i and V_i, $i \in \{1, \ldots, n\}$ are called the *parts* on the U- and V-side, respectively.

For a visualization of such a hypergraph, see Figure 1.

Note that every hyperedge in $G_{2,n}$ connects a part on the U- and a part on the V-side. We remark that the HAP can be formulated in the same way for more general bipartite hypergraphs, with less structure and possibly containing hyperedges which contain more than four vertices, see [Borndörfer and Heismann, 2012].

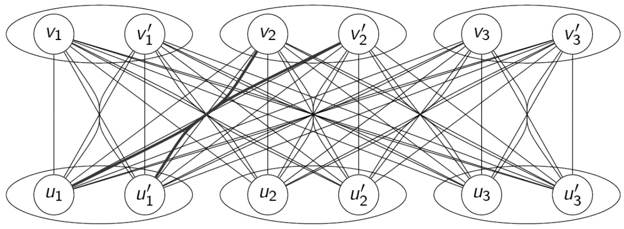

Figure 1: Visualization of the bipartite hypergraph $G_{2,3}$. The thick hyperedge is the proper hyperedge $U_1 \cup V_2 = \{u_1, u_1', v_2, v_2'\}$.

Definition 2.2. For a vertex subset $W \subseteq U \cup V$ we define the *incident hyperedges*

$$\delta(W) := \{e \in E : e \cap W \neq \emptyset, e \setminus W \neq \emptyset\}$$

to be the set of all hyperedges having at least one vertex in both W and $(U \cup V) \setminus W$.

A *hyperassignment* in $G_{2,n}$ is a subset $H \subseteq E$ of pairwise disjoint hyperedges that cover U and V, i. e., for all $e_1, e_2 \in H$, $e_1 \cap e_2 = \emptyset$, and $\bigcup H = U \cup V$. Given a cost function $c_E : E \to \mathbb{R}$, the cost of a hyperassignment is $\sum_{e \in H} c_E(e)$. The *hypergraph assignment problem* with input $(G_{2,n}, c_E)$ consists of finding a hyperassignment in $G_{2,n}$ of minimum cost w. r. t. c_E.

For bipartite hypergraphs $G_{2,n}$, the hypergraph assignment problem can be seen as a combination of two assignment problems. Namely, observe that for every hyperassignment H and every part U_i and V_i, $i \in \{1, \ldots, n\}$, the set of incident hyperedges $\delta(U_i) \cap H$ and $\delta(V_i) \cap H$ consists either of one proper hyperedge or of two edges. If we decide for every part U_i and V_i whether the hyperassignment to be constructed is incident to one proper hyperedge or to two edges, we can restrict the hyperedge set of $G_{2,n}$ to

- the set of edges connecting pairs of vertices from the parts U_i, V_i that will be incident to edges—this is the first assignment problem, and
- the proper hyperedges $\{U_i \cup V_j\}$ for U_i and V_j that will be incident to proper hyperedges—viewing U_i and V_j as composite vertices and the hyperedges as edges connecting them—this is the second assignment problem.

Solving these two assignment problems independently produces the minimum cost hyperassignment subject to the fixed edge and hyperedge incidences.

The HAP in $G_{2,n}$ can thus be solved in two steps. The first step decides which parts U_i and V_i will be incident to proper hyperedges. Of course, we must chose the same number of parts on the U- and the V-side, equal to the number of proper hyperedges in the hyperassignment to be constructed; the other parts will be incident to edges. The second steps consists of solving the resulting two assignment problems stated above.

3 Expected optimal values for the random HAP with exponential or uniform costs

Predicting the optimal value of a random hypergraph assignment problem in $G_{2,n}$ involves a prediction of the number of proper hyperedges in an optimal solution. This number depends on how advantageous it is to choose a proper hyperedge instead of two edges (so that one has just one number adding to the cost instead of two) compared to the disadvantage of having less freedom (there are fewer possibilities to cover a single vertex with a proper hyperedge than with an edge) when searching for a hyperassignment with the least possible cost. We conjecture that one can expect that an optimal hypergraph assignment in $G_{2,n}$ contains half of the possible number of proper hyperedges.

Conjecture 3.1. The expected number of proper hyperedges in a minimum cost hyperassignment in $G_{2,2n}$ with cost function c_E such that all $c_E(e)$, $e \in E$ are i. i. d. exponential random variables with mean 1 or i. i. d. uniform random variables on $[0,1]$ is n.

Table 1 backs this conjecture. It gives computational results for the random hypergraph assignment problem in the bipartite hypergraph $G_{2,n}$ with i. i. d. uniform random variables on $[0,1]$ and i. i. d. exponential random variables with mean 1 as hyperedge costs. For every n, we report the mean value and the standard deviation of the optimal cost value and the number of proper hyperedges in the optimal solution for 1000 computations. The HAPs were solved as integer programs using CPLEX 12.5.

The first column (n) of Table 1 shows the number of parts on the U- and V-side. Columns 2 and 6 (opt. val.) give the mean optimal values. Their standard deviations can be seen in columns 3 and 7 (s. d.) for the two cost function distributions, respectively. Columns 4 and 8 (# pr. hy.) show the number of proper hyperedges in the optimal solutions found, columns 5 and 9 (s. d.) show their standard deviations. The important finding w. r. t. Conjecture 3.1 is that the values in columns 4 and 8 are about half the values in column 1 in each row.

The computational results also suggest that the expected optimal cost converges to a value around 1.05 for both distributions. Although for larger n more hyperedges are contained in a hyperassignment, the optimal value does not increase much. This can be intuitively explained by noting that for larger n there are also more possible hyperassignments to select from, and the chances to find a hyperassignment that has a low cost are therefore still good even if it will contain more hyperedges.

We will now compute a lower and upper bound on the expected value of a minimum cost hyperassignment in $G_{2,2n}$ with n proper hyperedges for the exponential distribution. To this end, we will use the following result: For a complete bipartite graph with vertex sets of size m and n and with i. i. d. exponential random variables with mean 1 as edge costs, the expected minimum value of the sum of k pairwise disjoint edges (this is called a partial assignment) is

$$E(m,n,k) := \sum_{\substack{i,j\geq 0\\ i+j\leq k-1}} \frac{1}{(n-i)(m-j)}.$$

This result was conjectured in [Coppersmith and Sorkin, 1999] and first proved in [Linusson and Wästlund, 2004]. The latter paper also shows that for $m = n = k$ this term can be written as

$$E(n,n,n) = \sum_{i=1}^{n} \frac{1}{i^2}.$$

That this formula gives the expected value of a random assignment is Parisi's Conjecture.

Theorem 3.2. *Let* $\mathbf{E}$ *be the expected value of the minimum cost of a hyperassignment in* $G_{2,2n} = (U,V,E)$ *with exactly* n *proper hyperedges and cost function* c_E *with i. i. d. exponential random variables* $c_E(e)$ *with mean 1 for all* $e \in E$. *The following holds for* $n \to \infty$:

$$0.3718 < \mathbf{E} < 1.8310.$$

Proof. By definition,

$$E(n) := E(2n,2n,n) = \sum_{\substack{i,j\geq 0\\ i+j\leq n-1}} \frac{1}{(2n-i)(2n-j)}.$$

Using $E(n)$, we can bound the expected value of a hyperassignment in $G_{2,2n}$ with i. i. d. exponential random variables with mean 1 as hyperedge costs restricted to the hyperassignments with n proper hyperedges as follows.

For the lower bound, observe that in the best possible hyperassignment the selected n proper hyperedges can be only as good as the n pairwise disjoint proper hyperedges with the least possible cost sum in $G_{2,2n}$. Also, the selected $2n$ edges can be only as good as the $2n$ pairwise disjoint edges with the least possible cost sum in $G_{2,2n}$. Thus, $E(n) + E(2n)$ is a lower bound for $\mathbf{E}$.

On the other hand, choosing the n pairwise disjoint proper hyperedges with the least possible cost sum in $G_{2,2n}$ and finding the best possible edges for the remaining "unused" vertices leads to an upper bound of $E(n) + E(2n,2n,2n)$ for $\mathbf{E}$.

To transform the two-indexed sum describing $E(n)$ to a sum with only one index, we calculate the difference $D(n) := E(n+1) - E(n)$ and use the recursive formula

$$E(n) = E(1) + \sum_{i=1}^{n-1} D(i) = \frac{1}{4} + \sum_{i=1}^{n-1} D(i). \qquad (1)$$

We get

$$\begin{aligned} D(n) &= E(n+1) - E(n) \\ &= E(2n+2, 2n+2, n+1) - E(2n,2n,n) \end{aligned}$$

Table 1: Computational results for random hypergraph assignment problems in $G_{2,n}$ for i. i. d. uniform random variables on $[0, 1]$ or i. i. d. exponential random variables with mean 1 as hyperedge costs. The mean optimal values (column 2 and 6) and their standard deviations (column 3 and 7) are rounded to the third decimal place. The number of proper hyperedges in the optimal hyperassignments (column 4 and 8) and their standard deviations (column 5 and 9) are rounded to one decimal place. 1000 computations were done for each value of n and each distribution. The values in columns 4 and 8 are about half the value of column 1 in each row. This supports Conjecture 3.1.

	uniform on $[0,1]$				exponential with mean 1			
n	opt. val.	s. d.	# pr. hy.	s. d.	opt. val.	s. d.	# pr. hy.	s. d.
10	0.943	0.177	5.5	2.0	1.019	0.206	5.3	2.0
20	1.006	0.136	10.4	2.8	1.039	0.141	10.4	2.8
30	1.018	0.109	15.5	3.4	1.049	0.117	15.3	3.4
40	1.037	0.096	20.7	4.0	1.045	0.097	20.5	3.9
50	1.036	0.085	25.8	4.4	1.054	0.085	25.4	4.3
60	1.044	0.078	31.0	4.8	1.050	0.080	30.6	4.7
70	1.041	0.074	35.8	4.9	1.053	0.079	35.6	5.1
80	1.044	0.070	40.9	5.4	1.054	0.069	40.6	5.4
90	1.044	0.066	45.9	5.8	1.053	0.066	45.9	5.8
100	1.047	0.061	50.9	6.3	1.057	0.063	50.6	6.3
110	1.047	0.058	56.3	6.3	1.054	0.060	56.1	6.4
120	1.048	0.057	61.1	6.6	1.052	0.056	61.1	6.7
130	1.051	0.055	66.4	7.1	1.054	0.053	66.3	6.9
140	1.053	0.054	71.6	7.4	1.053	0.051	71.3	7.1
150	1.051	0.053	76.0	7.7	1.051	0.050	76.2	7.5
160	1.048	0.049	81.6	7.4	1.054	0.048	81.2	7.6

$$= \sum_{\substack{i,j\geq 0\\ i+j\leq n}} \frac{1}{(2n+2-i)(2n+2-j)} - \sum_{\substack{i,j\geq 0\\ i+j\leq n-1}} \frac{1}{(2n-i)(2n-j)}.$$

Shifting the index of the first sum to get the same summand type in both sums yields

$$= \sum_{\substack{i,j\geq -2\\ i+j\leq n-4}} \frac{1}{(2n-i)(2n-j)} - \sum_{\substack{i,j\geq 0\\ i+j\leq n-1}} \frac{1}{(2n-i)(2n-j)}.$$

We now split the sums to sums with index range $i, j \geq 0$, $i + j \leq n - 4$ so that they can cancel. The remainder is as follows. For the first sum, it is used that it is symmetric in i and j. The term $\frac{(4n+3)^2}{4(n+1)^2(2n+1)^2}$ is the sum of the values where $-2 \leq i, j \leq -1$. This has to be subtracted from the first term as otherwise these values would be counted twice.

$$D(n) = 2 \cdot \sum_{\substack{-2\leq i\leq -1, j\geq -2\\ i+j\leq n-4}} \frac{1}{(2n-i)(2n-j)} - \frac{(4n+3)^2}{4(n+1)^2(2n+1)^2}$$

$$- \sum_{\substack{i,j\geq 0\\ i+j=n-1}} \frac{1}{(2n-i)(2n-j)} - \sum_{\substack{i,j\geq 0\\ i+j=n-2}} \frac{1}{(2n-i)(2n-j)} - \sum_{\substack{i,j\geq 0\\ i+j=n-3}} \frac{1}{(2n-i)(2n-j)}.$$

Splitting the first sum into two parts with $i = -1$ and $i = -2$ and substituting j by $a - i$ where $i + j = a$ yields

$$\begin{aligned} D(n) = &\sum_{j=-2}^{n-3} \frac{2}{(2n+1)(2n-j)} \\ &+ \sum_{j=-2}^{n-2} \frac{2}{(2n+2)(2n-j)} \\ &- \frac{(4n+3)^2}{4(n+1)^2(2n+1)^2} \\ &- \sum_{i=0}^{n-1} \frac{1}{(2n-i)(n+1+i)} \\ &- \sum_{i=0}^{n-2} \frac{1}{(2n-i)(n+2+i)} \\ &- \sum_{i=0}^{n-3} \frac{1}{(2n-i)(n+3+i)}. \end{aligned}$$

Using the notation $H_n = \sum_{i=1}^{n} \frac{1}{i}$ for the n-th harmonic number and partial fraction decomposition to get denominators linear in n for the last two summations, we get

$$\begin{aligned}
D(n) &= \frac{2H_{2n+2} - 2H_{n+2}}{2n+1} + \frac{2H_{2n+2} - 2H_{n+1}}{2n+2} \\
&\quad - \frac{(4n+3)^2}{4(n+1)^2(2n+1)^2} - \frac{2H_{2n} - 2H_n}{3n+1} \\
&\quad - \frac{2H_{2n} - 2H_{n+1}}{3n+2} - \frac{2H_{2n} - 2H_{n+2}}{3n+3} \\
&= \frac{2H_{2n} + \frac{2}{2n+1} + \frac{2}{2n+2} - 2H_n - \frac{2}{n+1} - \frac{2}{n+2}}{2n+1} \\
&\quad + \frac{2H_{2n} + \frac{2}{2n+1} + \frac{2}{2n+2} - 2H_n - \frac{2}{n+1}}{2n+2} \\
&\quad - \frac{(4n+3)^2}{4(n+1)^2(2n+1)^2} - \frac{2H_{2n} - 2H_n}{3n+1} \\
&\quad - \frac{2H_{2n} - 2H_n - \frac{2}{n+1}}{3n+2} \\
&\quad - \frac{2H_{2n} - 2H_n - \frac{2}{n+1} - \frac{2}{n+2}}{3n+3}.
\end{aligned}$$

Finally, simplification yields

$$\begin{aligned}
D(n) = &-(H_{2n} - H_n)\cdot \\
&\frac{9n^2 + 11n + 4}{3(n+1)(2n+1)(3n+1)(3n+2)} \\
&+ \frac{8n^2 + 13n + 6}{12(n+1)^2(2n+1)^2(3n+2)}.
\end{aligned}$$

To get bounds on $E(n)$ using Equation (1), we first use that

$$\begin{aligned}
&\sum_{n=1}^{\infty} \frac{8n^2 + 13n + 6}{12(n+1)^2(2n+1)^2(3n+2)} \\
&\qquad = -\frac{1}{4} - \frac{\pi}{\sqrt{3}} + \frac{\pi^2}{9} - \frac{10\ln(2)}{3} + \ln(27). \quad (2)
\end{aligned}$$

Then, observe that $H_{2n} - H_n$ is a non-negative number monotonically increasing with n. Also, this is an alternating harmonic number that for $n \to \infty$ converges to $\ln(2)$. For $n = 80$, $H_{2n} - H_n$ can be calculated and results in a fraction, which is > 0.69. Therefore, for $n \geq 80$,

$$0.69 < H_{2n} - H_n < \ln(2) \quad (3)$$

Now, computing the partial sum

$$\sum_{n=1}^{79} -(H_{2n} - H_n) \frac{9n^2 + 11n + 4}{3(n+1)(2n+1)(3n+1)(3n+2)}$$

exactly and the limes

$$\sum_{n=80}^{\infty} -(H_{2n} - H_n) \frac{9n^2 + 11n + 4}{3(n+1)(2n+1)(3n+1)(3n+2)}$$

after substituting for $H_{2n} - H_n$ the lower and upper bounds given by (3), Equations (1) and (2) yield

$$0.1859 < \lim_{n\to\infty} E(n) < 0.1860.$$

Thus, we get for the lower bound

$$\begin{aligned}
\lim_{n\to\infty} (E(n) + E(2n)) &= 2 \cdot \lim_{n\to\infty} E(n) \\
&> 2 \cdot 0.1859 \\
&= 0.3718
\end{aligned}$$

and for the upper bound

$$\begin{aligned}
\lim_{n\to\infty} (E(n) + E(2n, 2n, 2n)) &= \lim_{n\to\infty} E(n) \\
&\quad + \lim_{n\to\infty} E(2n, 2n, 2n) \\
&< 0.1860 + \frac{\pi^2}{6} \\
&< 1.8310.
\end{aligned}$$

□

We remark that the upper bound computed in Theorem 3.2 is greater than the expected optimal value of the random assignment problem $\frac{\pi^2}{6} = 1.6449\ldots$. We believe that it must be possible to reduce it, because moving from an assignment problem in a complete bipartite graph with $4n$ vertices on each side to a HAP in $G_{2,2n}$ adds more possibilities (still all assignments are feasible solutions but using hyperassignments with proper hyperedges gives additional ones). Indeed, it is clear that if we do not prescribe the number of proper hyperedges in an optimal solution, the expected optimal value of a hyperassignment in $G_{2,2n}$ will tend to some number $\leq \frac{\pi^2}{6}$. As already discussed, the computational results shown in Table 1 suggest that the correct number is some value around 1.05, much smaller than $\frac{\pi^2}{6}$.

4 Regularity rewarding costs

Hypergraph assignment problems arising from practical applications feature costs for proper hyperedges that depend on the costs of the edges that they contain. Indeed, proper hyperedges model a "reward" for choosing combinations of edges; in this way, one can model a so-called regularity of the solution [Borndörfer et al., 2011]. More precisely, one considers partitioned bipartite hypergraphs and wants to favor the simultaneous choice of a set of edges that connects all nodes in a certain part in U to all nodes in a certain part in V. To this purpose, one introduces a proper hyperedge that represents the union of such pairwise disjoint edges and that has a cost that is smaller than the sum of the edge costs. If different edge combinations result in the same hyperedge, the cost is inferred from the edge set with the minimum cost sum. Here is a more formal statement.

Definition 4.1. Let $G = (U, V, E)$ be a partitioned hypergraph. For $e \in E$, let

$$\mathcal{E}(e) := \{E' \subseteq E_1 : \ e_1 \cap e_2 = \emptyset \ \forall e_1, e_2 \in E' \text{ with } e_1 \neq e_2, \bigcup E' = e\}$$

be the set of all pairwise disjoint edge sets with union e.

For some *penalty* $p \geq 0$, we call a cost function $c_E^p : E \to \mathbb{R}$ *regularity-rewarding* if for all proper hyperedges $e \in E_2$,

$$c_E(e) = \min_{E' \in \mathcal{E}(e)} \left(\sum_{e' \in E'} c_E(e') - p \cdot |E'| \right).$$

The greater p, the more irregularity is punished and regularity rewarded. We remark that the cost of a hyperedge in a vehicle rotation planning model depends on several other parameters such as an additional irregularity penalty for hyperedges that are not inclusion-wise maximal [Borndörfer et al., 2011]. This is the reason why we call p a penalty and not a bonus or a reward.

A way to define a regularity-rewarding random cost function c_E^p is to draw a *random basic cost* r_e for each edge $e \in E_1$, e. g., from a uniform distribution on $[0, 1]$ or an exponential distribution with mean 1, and then to set

$$c_E^p(e) := \begin{cases} r_e + p & \text{if } e \text{ is an edge,} \\ \min_{E' \in \mathcal{E}(e)} \sum_{e' \in E'} r_{e'} & \text{if } e \text{ is a proper hyperedge.} \end{cases}$$

In the following, we will assume that c_E^p is structured in this way with arbitrary r_e.

For a given bipartite hypergraph $G_{2,n} = (U, V, E)$ and random basic costs r_e for the edges $e \in E_1$, we denote by $z(h, p)$ the minimal cost value of a hyperassignment with penalty p that contains exactly $0 \leq h \leq n$ proper hyperedges.

Obviously, the number of proper hyperedges and the value of an optimal solution will depend on p. If $p = 0$, there is no reward for choosing a proper hyperedge. For every solution using proper hyperedges, we can find a solution with the same value that contains only edges by replacing each proper hyperedge $\{u_i, u_i', v_j, v_j'\}$ by either the two edges $\{u_i, v_j\}, \{u_i', v_j'\}$ or the two edges $\{u_i, v_j'\}, \{u_i', v_j\}$ depending on which two edges have the lower cost sum. On the other hand, if p is very large, choosing edges for a solution becomes so disadvantageous that the number of proper hyperedges in an optimal solution will become very high.

Fortunately, knowledge about the case $p = 0$ gives information about all other penalties as the following theorem shows. Thus, we do not need to analyze random HAPs for regularity-rewarding cost functions separately for each penalty p.

For some random basic cost distribution, we denote by $Z(h)$ the expected value of $z(h, 0)$ with respect to this distribution. Although $z(h, 0)$ is defined only for integral h, we will view $Z(h)$ as a continuous, monotonically increasing, differentiable function on $[0, n]$. This will allow us to formulate our result in a much easier way than if we would have to replace the derivative by its discretization. We can require $Z(h)$ to be monotonically increasing, because $z(h, 0)$ is monotonically increasing with increasing h. The reason is that, as described above, using proper hyperedges in the solution cannot lead to smaller optimal values than using only edges in the case $p = 0$.

Theorem 4.2. *Consider the complete bipartite hypergraph $G_{2,n} = (U, V, E)$ and let r_e, $e \in E_1$ be random basic costs chosen from some random distribution. Denote by $h_d^1, \ldots h_d^k$ the solutions to the equation $Z'(h) = 2p$ and let*

$$h^* = \arg \min_{h \in \{0, h_d^1, \ldots, h_d^k, n\}} (Z(h) - (2n - 2h)p)$$

Then the expected number of proper hyperedges in an optimal solution to the HAP in $G_{2,n}$ w. r. t. c_E^p with basic random costs r_e is h^* and the expected optimal value of the random HAP is

$$Z(h^*) - (2n - 2h^*)p.$$

Proof. First, observe that

$$z(h, p) = z(h, 0) + (2n - 2h)p$$

holds since the cost of each hyperassignment H w. r. t. c_E^p is

$$\begin{aligned} c_E^p(H) &= \sum_{e \in E} c_E^p(e) \\ &= \sum_{e \in E_1} c_E^p(e) + \sum_{e \in E_2} c_E^p(e) \\ &= \sum_{e \in E_1} (r_e + p) + \sum_{e \in E_2} \min_{E' \in \mathcal{E}(e)} \sum_{e' \in E'} r_{e'} \\ &= \sum_{e \in E_1} r_e + |E_1 \cap H| p \\ &\quad + \sum_{e \in E_2} \min_{E' \in \mathcal{E}(e)} \sum_{e' \in E'} r_{e'} \\ &= \sum_{e \in E_1} r_e + (2n - 2|E_2 \cap H|)\, p \\ &\quad + \sum_{e \in E_2} \min_{E' \in \mathcal{E}(e)} \sum_{e' \in E'} r_{e'} \\ &= \sum_{e \in E} c_E^0(e) + (2n - 2|E_2 \cap H|)\, p \\ &= c_E^0(H) + (2n - 2|E_2 \cap H|)\, p. \end{aligned}$$

Since this holds for all random basic costs, it also holds for the expected value of all random basic cost distributions and we get

$$\mathbf{E}(z(h, p)) = Z(h) + (2n - 2h)p.$$

Its derivative is $Z'(h) - 2p$. A minimum of a differentiable function is attained either at the bounds or where the derivative is equal to zero, which proves the theorem. $\square$

5 Discussion

In this paper, we have presented results on the expected minimum cost of the random hypergraph assignment problem for two types of cost functions.

For the first type, i. i. d. exponential random variables with mean 1 or i. i. d. uniform random variables on $[0, 1]$, we conjectured that the number of proper hyperedges in an optimal solution is expected to be n for the hypergraph $G_{2,2n}$, and showed computational results supporting this conjecture. Assuming this number of proper hyperedges in an optimal solution, we proved bounds on the expected optimal value for a vertex number tending to infinity. A proof of our conjecture as well as convergence results and either sharper bounds or an exact limit would be a natural continuation of our work towards a generalization of Mézard and Parisi's Conjecture. A first step is to extend the proof of our bounds to fixed numbers of hyperedges other than n by altering the computation.

For the second type of regularity-rewarding cost functions, we established a connection between results for different penalty values. This result could be extended by an analysis similar to that for the first cost function type in future.

All our results hold for complete partitioned hypergraphs $G_{2,n}$. A further line of research could try to extend these results to bipartite hypergraphs with larger part sizes or even bipartite hypergraphs that are not partitioned or/and not complete.

Our results show how to approach the random HAP using results for the random assignment problem. Probably approaches using more sophisticated probability-theoretical results are needed to understand more about the problem.

References

[Aldous, 1992] Aldous, D. (1992). Asymptotics in the random assignment problem. *Probability Theory and Related Fields*, 93:507–534.

[Borndörfer and Heismann, 2012] Borndörfer, R. and Heismann, O. (2012). The hypergraph assignment problem. Technical Report 12-14, ZIB.

[Borndörfer et al., 2011] Borndörfer, R., Reuther, M., Schlechte, T., and Weider, S. (2011). A Hypergraph Model for Railway Vehicle Rotation Planning. In Caprara, A. and Kontogiannis, S., editors, *11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2011)*, volume 20 of *OpenAccess Series in Informatics (OASIcs)*, pages 146–155, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ZIB Report 11-36.

[Coppersmith and Sorkin, 1999] Coppersmith, D. and Sorkin, G. B. (1999). Constructive bounds and exact expectations for the random assignment problem. *Random Structures & Algorithms*, 15(2):113–144.

[Heismann and Borndörfer, 2013] Heismann, O. and Borndörfer, R. (2013). The random hypergraph assignment problem. In *Proceedings of the 16th International Multiconference INFORMATION SOCIETY – IS 2013*.

[Krokhmal and Pardalos, 2009] Krokhmal, P. A. and Pardalos, P. M. (2009). Random assignment problems. *European Journal of Operational Research*, 194(1):1–17.

[Linusson and Wästlund, 2004] Linusson, S. and Wästlund, J. (2004). A proof of Parisi's conjecture on the random assignment problem. *Probability Theory and Related Fields*, 128(3):419–440.

[Maróti, 2006] Maróti, G. (2006). *Operations research models for railway rolling stock planning*. PhD thesis, Technische Universiteit Eindhoven.

[Mézard and Parisi, 1985] Mézard, M. and Parisi, G. (1985). Replicas and optimization. *Journal de Physique Lettres*, 46:771–778.