



**HAL**  
open science

## On the average complexity of Berge Algorithm

Julien David, Mostafa Gholami, Loïck Lhote

► **To cite this version:**

Julien David, Mostafa Gholami, Loïck Lhote. On the average complexity of Berge Algorithm. 2025. ⟨hal-04637413⟩

**HAL Id: hal-04637413**

**<https://hal.science/hal-04637413v1>**

Preprint submitted on 21 May 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# 1 On the average complexity of Berge Algorithm

2 Julien David ✉

3 CNRS NormaSTIC FR 3638, France

4 Normandie University, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France

5 Mostafa Gholami ✉

6 CNRS NormaSTIC FR 3638, France

7 Normandie University, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France

8 Loick Lhote ✉

9 CNRS NormaSTIC FR 3638, France

10 Normandie University, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France

## 11 — Abstract —

---

12 Given a hypergraph, computing its transversal hypergraph is a classical problem whose exact  
13 worst-case complexity is still unknown today. In this article, for random hypergraph models à la  
14 Erdos and Renyi, we compute the average number of minimal transversals. We also study the  
15 average complexity of Berge’s algorithm, which sequentially enumerates the minimal transversals.  
16 We prove that the average complexity of Berge is quasi-linear in the average number of minimal  
17 traverses. Our contribution is also methodological: we use a wide range of analytic combinatorics  
18 techniques to extract the coefficients of probabilistic generating functions.

19 **2012 ACM Subject Classification** Design and analysis of algorithms

20 **Keywords and phrases** Average Complexity, Minimal Transversal, Berge Algorithm, Hypergraphs,  
21 Analytic Combinatorics

22 **Digital Object Identifier** 10.4230/LIPIcs...

## 23 **1** Introduction

24 A *hypergraph*  $\mathcal{H}$  is a pair  $(\mathcal{V}, \mathcal{E})$  where  $\mathcal{V}$  is a set of vertices and  $\mathcal{E}$  is a set of hyperedges,  
25 each hyperedge being a subset of  $\mathcal{V}$ . A *transversal* of  $\mathcal{H}$  is a subset of vertices  $T \subseteq \mathcal{V}$  that  
26 intersects all the hyperedges. A transversal is said *minimal* if no proper subset is also a  
27 transversal. The *Transversal Hypergraph Generation* problem (THG problem) consists in  
28 computing, for a given hypergraph  $\mathcal{H}$ , the transversal hypergraph  $Tr(\mathcal{H})$  formed with all the  
29 minimal transversals of  $\mathcal{H}$ .

30 This problem has gained considerable attention in recent years due to its generality  
31 and wide applicability in a variety of fields, including artificial intelligence and logic [16],  
32 computational biology [10, 36, 26], cryptography [32], database theory [37], data mining  
33 [23, 1, 7], distributed computing [21], e-commerce [3], machine learning [24], mathematical  
34 programming [35, 33], mobile communication systems [39], semantic web [30], topology [14],  
35 combinatorics [38], and game theory [8].

36 Several problems are equivalent to the THG problem. For instance, the enumeration of  
37 minimal dominating sets or total dominating sets in graphs can be polynomially reduced to  
38 the THG problem [9, 29].

39 The size of  $Tr(\mathcal{H})$ , or equivalently, the number of minimal transversals, may be exponential  
40 in the size of  $\mathcal{H}$ . Efforts have been done to determine an upper bound for the size of  $Tr(\mathcal{H})$   
41 such as in [4] (and the references it contains) where both upper and lower bounds have been  
42 computed for tripartite 3-uniform hypergraphs. Since the size of  $Tr(\mathcal{H})$  may be exponential  
43 in the size of  $\mathcal{H}$ , it is not possible to find an algorithm that runs in polynomial-time. Johnson,  
44 Yannakakis and Papadimitriou [28] introduced various classes of complexity involving the



© anonymous;  
licensed under Creative Commons License CC-BY 4.0



Leibniz International Proceedings in Informatics  
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

45 size of the intermediate results or the size of the output. They defined the class of *output-*  
 46 *polynomial time* problems that can be solved in polynomial time with respect to the size of  
 47 the input and the output.

48 Finding an *output-polynomial* algorithm for Transversal Hypergraph Generation is a  
 49 long-standing problem [12]. While the question remains open for arbitrary hypergraphs,  
 50 algorithms achieving output polynomial time complexity exist for several classes of hyper-  
 51 graphs: hypergraphs with bounded minimal transversals [15], with bounded edge sizes [6, 15],  
 52 of bounded degree [17], of bounded-edge intersections, of bounded conformality [6].

53 The Berge algorithm, also known as the sequential method, was the first proposed  
 54 algorithm to solve the THG problem [5]. It computes the transversal hypergraph by adding  
 55 one by one the hyperedges and by updating the set of minimal transversals. Takata proved  
 56 that its worst-case complexity is not output-polynomial [40] and exhibited specific hypergraphs  
 57 where the quantity of intermediate minimal transversals could exponentially exceed the final  
 58 solution. In recent years, several improvements of the Berge algorithm have been conducted.  
 59 As instances, we can cite the algorithms proposed by Dong and Li [13], Bailey, Manoukian,  
 60 and Ramamohanarao [2], Kavvadias and Stavropoulos [31]. Using a method akin to Takata's  
 61 in [40], Hagen proved that most of the improvements of Berge algorithms that mentioned  
 62 are not output polynomial [25].

63 The currently fastest known algorithm for solving the Transversal Hypergraph Generation  
 64 problem is due to Fredman and Khachiyan [20, 34] and runs in quasipolynomial time  $N^{o(\log N)}$ ,  
 65 where  $N$  is the size of the input plus the output.

66 In this paper, we adopt a probabilistic point of view on the THG problem and we are  
 67 interested in computing the average size of the output and the average case complexity of the  
 68 algorithms. In [22], the authors studied the average-case complexity of the decision version  
 69 of the THG problem on uniform instances. However, the uniform distribution leads with high  
 70 probability to very specific hypergraphs. An upper bound on the average complexity of  
 71 the `MTMiner` algorithm [27] for the THG problem, is also given in [11]. The authors study  
 72 various random models in which they give bounds on the number of transversals or minimal  
 73 transversals. Then they use those results to obtain bounds on the average complexity of the  
 74 `MTMiner` algorithm.

75 In this article, we use the probabilistic models introduced in [11] (see Section 2). However,  
 76 our methodology based on Analytic Combinatorics [19] is for the first time used in the  
 77 analysis of minimal transversals. The methodology allows us obtain more precise results, in  
 78 particular an equivalence on the asymptotic average number of minimal transversals. We  
 79 also study the average-case behaviour of Berge algorithm instead of `MTMiner`.

80 Analytic Combinatorics is a general method developed by Flajolet and Sedgewick in the  
 81 80's [19]. The method mixes tools from the Symbolic Method and complex analysis in order  
 82 to describe the asymptotic behavior of algorithms or large random data structures.

83 The main tools are generating functions whose coefficients describe the probabilistic  
 84 behavior of the parameters of interest. Explicit expressions of the generating functions are  
 85 usually obtained using the Symbolic Method that translates the combinatorial structure of  
 86 the problem into operations on the generating functions. Then, tools from complex analysis  
 87 such as transfer theorems or saddle point method are used to compute the asymptotic of the  
 88 coefficients.

89 Paper is organized as follows. Section 2 details random models we consider on hypergraphs.  
 90 Section 3 recall basic definitions and notations on hypergraphs and Berge Algorithm. We  
 91 also show the relation between the probability of a given set to be a minimal transversal and  
 92 the average case complexity of Berge algorithm. Section 4 describe probabilistic generating

93 functions for a set with  $\ell$  vertices to be a minimal transversal in the general model. As we  
 94 will see, both an exponential generating function and an ordinary one will be necessarily to  
 95 obtain the average complexity of Berge Algorithm. Section 5 contains the main results of  
 96 the paper. We present the probability for a given set of size  $\ell$  to be a minimal transversal in  
 97 an hypergraph with  $m$  hyperedges. From this we deduce the average number of minimal  
 98 transversal and the average complexity of Berge algorithm. Finally, we show that the  
 99 Poisson-Mellin-Newton cycle can be used on the exponential generating function to obtain a  
 100 closed formula to compute the probability for a set to be a minimal transversal, that did not  
 101 appear in [11]. Section 6 contains our perspectives and how we intend to improve our results  
 102 using an even more general random model. Section 7 contains the full detailed proofs.

## 103 2 Random models

104 The two following random models we present in this section already appear in [11]. The  
 105 hypergraph  $\mathbf{H}$  can be seen as a binary matrix  $M(\mathbf{H}) = (\chi_{i,j}(\mathbf{H}))_{i=1..m,j=1..n}$ , called the  
 106 incidence matrix of  $\mathbf{H}$ , with  $\chi_{i,j}(\mathbf{H}) = 1$  if and only if the vertex  $v_j$  belongs to the hyperedge  
 107  $e_i$ . In the first random model, the columns and the rows of the matrix are independent. It  
 108 can be seen as the adaptation of the Erdős-Renyi model [18] on hypergraphs.

109 ► **Definition 1** ( $\mathcal{HG}(m, n, p)$  random model). *Let  $p \in [0, 1]$  be a parameter. In the  $\mathcal{HG}(m, n, p)$*   
 110 *model, the family of random variables  $(\chi_{i,j}(\mathbf{H}))_{i=1..m,j=1..n}$  forms an independent family of*  
 111 *random variables where for all  $(i, j)$ ,  $\chi_{i,j}$  follows a Bernoulli law of parameter  $p$ .*

112 The second model generalizes the previous one.

113 ► **Definition 2** ( $\mathcal{HG}(m, n, \pi)$  random model). *Fix a sequence  $\pi = (\pi_1, \dots, \pi_n)$  in  $[0, 1]^n$  of real*  
 114 *parameters. In the  $\mathcal{HG}(m, n, \pi)$  model, the family of random variables  $(\chi_{i,j}(\mathbf{H}))_{i=1..m,j=1..n}$*   
 115 *forms an independent family of random variables where for all  $(i, j)$ ,  $\chi_{i,j}$  follows a Bernoulli*  
 116 *law of parameter  $p_j$ .*

117 All the rows (hyperedges) follow the same distribution conversely to the columns where the  
 118 degree of the vertex  $v_j$  follows a binomial law of parameter  $(m, p_j)$ . Then, the average degree  
 119 of  $v_j$  is controlled and is  $mp_j$ .

120 In [11], the authors obtain precise bounds on the average number of transversals and  
 121 minimal transversals in the  $\mathcal{HG}(m, n, p)$  model, and some less precise results in the  $\mathcal{HG}(m, n, \pi)$   
 122 model. In this paper, we obtain more precise results than in [11] in the  $\mathcal{HG}(m, n, p)$  model  
 123 and some versions of the  $\mathcal{HG}(m, n, \pi)$  model.

## 124 3 Berge Algorithm and parameters of interest

125 Given two simple hypergraphs  $\mathcal{H} = (\mathcal{V}, \mathcal{E})$  and  $\mathcal{H}' = (\mathcal{V}', \mathcal{E}')$  where  $\mathcal{E} = \{e_1, \dots, e_m\}$  and  
 126  $\mathcal{E}' = \{e'_1, \dots, e'_{m'}\}$ . There are two different “unions”, namely  $\mathcal{H} \cup \mathcal{H}'$  and  $\mathcal{H} \vee \mathcal{H}'$  which are  
 127 defined as follows.

$$128 \quad \mathcal{H} \cup \mathcal{H}' = \{e_1, e_2, \dots, e_m, e'_1, e'_2, \dots, e'_{m'}\}$$

$$129 \quad \mathcal{H} \vee \mathcal{H}' = \{e_i \cup e'_j; i = 1, 2, \dots, m, \quad j = 1, 2, \dots, m'\}$$

► **Proposition 1.** [5] *Let  $\mathcal{H}$  and  $\mathcal{H}'$  be two simple hypergraphs. Then*

$$Tr(\mathcal{H} \cup \mathcal{H}') = \min(Tr(\mathcal{H}) \vee Tr(\mathcal{H}'))$$

## XX:4 On the average complexity of Berge Algorithm

130 The Berge algorithm, as described in [5], uses Proposition 1 to generate the transversal  
 131 hypergraph through the following procedure. Given a hypergraph  $\mathcal{H}$  with edges  $\mathcal{E} =$   
 132  $\{e_1, e_2, \dots, e_m\}$ , let  $H_i = \{e_1, e_2, \dots, e_i\}$  for  $i = 1, 2, \dots, m$ . The transversal hypergraph of  
 133  $\mathcal{H}_i$  is determined as

$$134 \quad Tr(H_i) = \min[Tr(\mathcal{H}_{i-1}) \vee Tr(\{e_i\})] = \min[Tr(\mathcal{H}_{i-1}) \vee \{v \mid v \in e_i\}].$$

135 This result is an iterative computation process, known as Berge algorithm, eventually yielding  
 $Tr(\mathcal{H}) = Tr(\mathcal{H}_m)$ . The given pseudocode for the Berge algorithm is as follows.

### ■ Algorithm 1 Berge Algorithm

---

1 Initialize  $Tr(\mathcal{H}_1)$  with the vertices  $v$  for  $v \in e_1$ ;  
 2 **for**  $i = 2$  to  $m$  **do**  
 3     |  $Tr(\mathcal{H}_i) = \min[Tr(\mathcal{H}_{i-1}) \vee \{v \mid v \in e_i\}]$ ;  
 4 Output  $Tr(\mathcal{H}_m)$ ;

---

136 The order in which hyperedges are considered has an impact on the algorithm's efficiency.  
 137 A natural question is therefore to establish an optimal order for the hyperedges and check  
 138 whether it has an impact on the worst-case complexity. This question has been addressed  
 139 by Takata[40], who demonstrated that the Berge algorithm is at least super-polynomial  
 140 for any order. He gives a characterization of hypergraphs for which this lower bound is reached:

- 143 ■  $\mathcal{G}_0$  is the hypergraph consisting of one hyperedge, containing a single vertex, that is  
 144  $\mathcal{V}(\mathcal{G}_0) = \{v_0\}$  and  $\mathcal{E}(\mathcal{G}_0) = \{\{v_0\}\}$ .
- 145 ■  $\mathcal{G}_i = (\mathcal{A} \cup \mathcal{B}) \vee (\mathcal{C} \cup \mathcal{D})$ , where  $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$  are all vertex-disjoint copies of  $\mathcal{G}_{i-1}$ .

146 ► **Theorem 1.** [40] *The running time of the Berge algorithm on the family of hypergraphs*  
 147  $\mathcal{G}_i (i \in \mathbb{N})$  *is superpolynomial in the size of the input and output no matter how the hyperedges*  
 148 *are ordered.*

149 Let's now consider the complexity of Berge algorithm. For each minimal transversal  
 150  $M \in Tr(\mathcal{H}_{i-1})$ , the algorithm adds a vertex  $v \in e_i$  and check minimality, which can be done  
 151 in time  $\Theta(|M| \times i)$  and space  $\Theta(|M|)$ , using an hash table. Since  $|M| \leq \min\{n, i\}$ , the cost  
 152 of Berge algorithm over a hypergraph  $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ , noted  $B(\mathcal{H})$ , is in the following range:

$$153 \quad \sum_{i=2}^m |Tr(\mathcal{H}_{i-1})| \cdot |e_i| \leq B(\mathcal{H}) \leq \sum_{i=2}^m |Tr(\mathcal{H}_{i-1})| \cdot |e_i| \cdot \min\{n, i\} \cdot i$$

154 In the  $\mathcal{HG}(m, n, \pi)$  model, let  $\mathbb{E}_\pi[Tr_{m,n}]$  be the average number of minimal transversal in  
 155 a hypergraph with  $m$  hyperedges and  $n$  vertices and  $\mathbb{E}_\pi[B(\mathcal{H})]$  the average complexity of  
 156 Berge algorithm. We have:

$$157 \quad \sum_{i=1}^{m-1} \mathbb{E}_\pi[Tr_{i,n}] \leq \mathbb{E}_\pi[B(\mathcal{H})] \leq n \cdot m \cdot \min\{n, m\} \sum_{i=1}^{m-1} \mathbb{E}_\pi[Tr_{i,n}]$$

Let  $c_{m,X}$  be the probability for a set  $X$  of size  $\ell$  to be a minimal transversal in a hypergraph  
 with  $m$  hyperedges. We have :

$$\mathbb{E}_\pi[Tr_{m,n}] = \sum_{\ell=1}^{\min\{n,m\}} \sum_{\substack{X \subseteq \{1, \dots, n\} \\ |X|=\ell}} c_{m,X}$$

In the  $\mathcal{HG}(m, n, p)$  model, the formula simplifies. Let  $\mathbb{E}_p[Tr_{m,n}]$  be the average number of minimal transversal and  $c_{m,\ell}$  be the probability for a set of size  $\ell$  to be a minimal transversal. We have:

$$\mathbb{E}_p[Tr_{m,n}] = \sum_{\ell=1}^{\min\{n,m\}} \binom{n}{\ell} c_{m,\ell}$$

158 Regarding the average complexity of Berge algorithm we have:

$$159 \quad \sum_{i=1}^{m-1} \mathbb{E}_p[Tr_{i,n}] = \sum_{i=1}^{m-1} \sum_{\ell=1}^{\min\{n,m\}} \binom{n}{\ell} c_{i,\ell} = \sum_{\ell=1}^{\min\{n,m\}} \binom{n}{\ell} \sum_{i=1}^{m-1} c_{i,\ell}. \quad (1)$$

160 We call *cumulative sum*, noted  $\bar{c}_{m,\ell}$ , the value  $\sum_{i=1}^{m-1} c_{i,\ell}$ . We have  $\sum_{i=1}^{m-1} \mathbb{E}_\pi[Tr_{i,n}] =$   
 161  $\sum_{\ell=1}^{\min\{n,m\}} \binom{n}{\ell} \bar{c}_{m,\ell}$ . In Section 4, we give an exponential and an ordinary probabilistic  
 162 generating function whose coefficient is  $c_{m,X}$  and an o.p.g.f whose coefficient is  $\bar{c}_{m,\ell}$ .

## 163 4 Generating functions

### 164 4.1 Exponential generating function

165 Given a probability distribution  $\pi = (\pi_1, \dots, \pi_n)$ , let's consider  $\ell$  columns of the incidence  
 166 matrix  $M(\mathbf{H})$ . Let  $X = \{x_1, \dots, x_\ell\} \subseteq \{1, \dots, n\}$  be the set of selected columns. We have  
 167  $(p_1, \dots, p_\ell) = (\pi_{x_1}, \dots, \pi_{x_\ell})$  the restriction of  $\pi$  over the  $\ell$  columns. The probability that a  
 168 row contains only zeros over those  $\ell$  columns is given by  $z_\ell(\pi) = (1 - p_1)(1 - p_2) \cdots (1 - p_\ell)$ .  
 169 From now on, if there is no ambiguity, we will use  $z_\ell$  instead of  $z_\ell(\pi)$ . The probability that a  
 170 row contains exactly one 1 (at column  $i$ ) over the  $\ell$  columns is given by  $o_{\ell,i} = \frac{p_i}{1-p_i} z_\ell$ . Finally  
 171 the probability that a row contains at least two 1 is  $1 - z_\ell - \sum_{i=1}^{\ell} o_{\ell,i}$ .

► **Proposition 2.** *In the  $\mathcal{HG}(m, n, \pi)$  model, the probabilistic generating function of a set  $C$  of  $\ell$  vertices to be a minimal transversal is*

$$S_\ell(u) = e^{(1-z_\ell) \cdot u} \prod_{i=1}^{\ell} (1 - e^{-o_{\ell,i} \cdot u})$$

**Proof.** The proof in the  $\mathcal{HG}(m, n, \mu)$  model consist in an integration over  $\mathcal{HG}(m, n, \pi)$  models. Assume that we have a distribution  $\pi = (p_1, \dots, p_n)$ . In the p.g.f.  $S_\ell(u)$ , the variable  $u$  marks each line of the matrix  $(\chi_{i,j}(\mathbf{H}))_{i=1..m, j=1..n}$ . For fixed  $n$  and  $\ell$ , the probability that a set of size  $\ell$  of vertices is a minimal transversal of a random hypergraph with  $m$  hyperedges is given by  $m![u^m]S_\ell(u)$ . First let's note that everything that happens on the  $n - \ell$  columns that are not related to our set of vertices is irrelevant to determine if the set is a minimal transversal. Therefore we can restrict our distribution  $\pi$  to  $(p_1, \dots, p_\ell)$  (up to a relabeling of the  $p_i$ ). The formula can then be obtained by using the symbolic method [19]. The p.g.f. of a non-empty set of lines that contain exactly one 1 at position  $i$  is  $e^{o_{\ell,i} \cdot u} - 1$ . Such lines exists for each position between 1 and  $\ell$ . Every remaining lines has to contain at least two 1. The associated p.g.f. is given by  $e^{(1-z_\ell - \sum_{i=1}^{\ell} o_{\ell,i}) \cdot u}$ . Combining all those parameters we obtain the generating function:

$$\left[ \prod_{i=1}^{\ell} (e^{o_{\ell,i} \cdot u} - 1) \right] e^{(1-z_\ell - \sum_{i=1}^{\ell} o_{\ell,i}) \cdot u}$$

172 after a step of simplification where  $e^{(-\sum_{i=1}^{\ell} o_{\ell,i}) \cdot u}$  is added in the product, we obtain the  
 173 announced results. ◀

174 **4.2 Ordinary generating function**

The Borel transform of an exponential generating function  $S_\ell(u)$  allows to obtain an ordinary generating function  $\hat{S}_\ell(u)$  sharing the same coefficients. It is defined as follows.

$$\hat{S}_\ell(u) = \int_0^\infty S_\ell(tu)e^{-t}dt$$

Applying the Borel transform to our exponential generating function ultimately yields the following ordinary generating function:

$$\hat{S}_\ell(u) = u^\ell \ell! \frac{1}{1-u(1-z_\ell)} \prod_{i=1}^\ell \frac{o_{\ell,i}}{1-u(1-z_\ell - io_{\ell,i})}$$

This generating function actually has a nice combinatorial interpretation: recall that the variable  $u$  mark each line. The  $u^\ell \prod_{j=1}^\ell o_{\ell,j}$  therefore denotes  $\ell$  lines that contains exactly one 1, one for each  $\ell$  columns. The  $\ell!$  allows to choose the order in which they appear. We'll call those lines *separators*. They guarantee the "minimality" property. After the  $\ell - i$ -th separator has been added in the matrix and until the next separator is added, each row cannot contain only zeros and cannot contain only one 1 in one of the  $i$  remaining positions. This probability is given by  $(1 - z_\ell - io_\ell)$ . A sequence of such rows is given by the quasi inverse  $\frac{1}{1-u(1-z_\ell - io_\ell)}$ . The product is obtained to deal with each block between each separators. Note that in the  $\mathcal{HG}(m, n, p)$  model, the function simplifies into  $\hat{S}_\ell(u) = u^\ell \ell! o_\ell^\ell \prod_{i=0}^\ell \frac{1}{1-u(1-z_\ell - io_\ell)}$ . Given the formal expression for  $\hat{S}_\ell(u)$ , we have:

$$\hat{S}_\ell(u) = \sum_{m \geq 0} c_{m,\ell} u^m$$

The generating function for the sequence  $\{\bar{c}_{i,\ell}\}_1^\infty$  is denoted by  $\bar{S}_\ell(u)$ , which can be written as  $\bar{S}_\ell(u) = \sum_{m \geq 0} \bar{c}_{m,\ell} u^m$ . This function,  $\bar{S}_\ell(u)$ , is referred to as the cumulative generating function:

$$\bar{S}_\ell(u) = \frac{1}{1-u} \hat{S}_\ell(u)$$

175 We will use the coefficients of  $\bar{S}_\ell(u)$  in calculating the average-case complexity of the Berge  
176 algorithm.

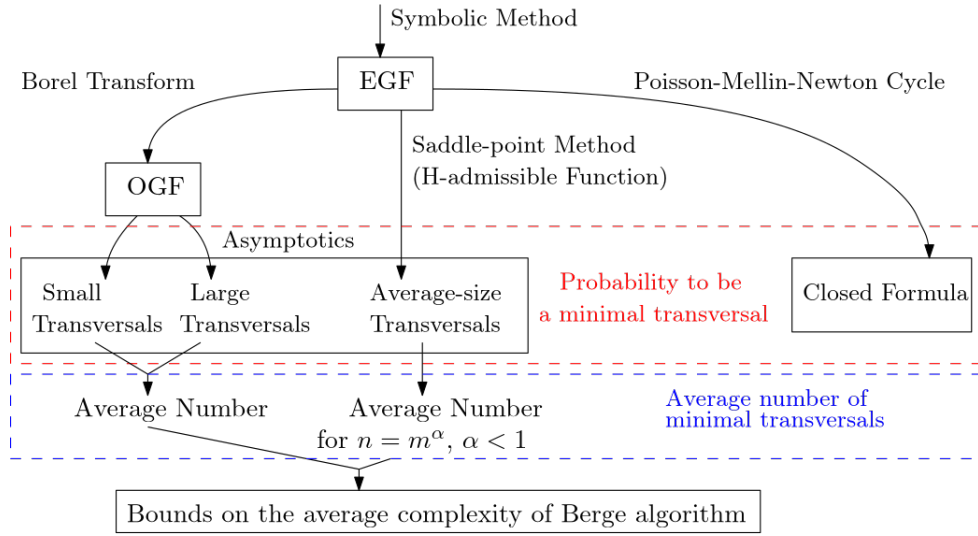
177 **5 Main results**

178 Figure 1 summarizes our contributions and the main results of this paper. In the  $HG(m, n, \pi)$   
179 model, we recall that  $z_\ell = \prod_{i=1}^\ell (1 - p_i)$ , where  $(p_1, \dots, p_\ell)$  is the restriction of  $\pi$  over the  
180 vertices of a set  $X$  of cardinal  $\ell$ .

181 **► Theorem 2.** *In the  $\mathcal{HG}(m, n, \pi)$  model, the probability for a set  $X$  of cardinal  $\ell$  to be a*  
182 *minimal transversal, or  $c_{m,X}$  is as follows:*

$$183 \quad c_{m,X} = \begin{cases} (1 - z_\ell)^m \left(1 + \mathcal{O}\left(\frac{\log m}{m^{1-\epsilon}}\right)\right) & \text{for } mz_\ell \geq m^\epsilon, \\ \binom{m}{\ell} \ell! \left(\prod_{i=1}^\ell o_{\ell,i}\right) \left(1 + \mathcal{O}\left(\frac{\log m}{m^{1-\epsilon}}\right)\right) & \text{for } mz_\ell \leq m^{-\epsilon}, \\ \frac{m^m}{e^m} \frac{S_\ell(r)}{r^m} \left(1 + \mathcal{O}\left(m^{-1/3} + \frac{1}{m}\right)\right) & \text{for } m^{-\epsilon} \leq mz_\ell \leq m^\epsilon \end{cases}$$

184 The results from Theorem 2 are based on the value of  $mz_\ell$ . For  $mz_\ell \geq m^\epsilon$  and  $mz_\ell \leq m^{-\epsilon}$ ,  
185 the results are the asymptotics of the coefficients of the ordinary generating function, and



■ **Figure 1** Summary of our main results and methods.

186 obtained by Cauchy's formula (Lemmas 13 and 14). It should be noted that for simplicity,  
 187 we have proved Lemmas 13 and 14 for  $\mathcal{HG}(m, n, p)$  model. The same argument works for  
 188  $\mathcal{HG}(m, n, \pi)$  model and gives the desired results. For  $m^{-\epsilon} \leq mz_\ell \leq m^\epsilon$ , the result is derived  
 189 from the asymptotics of the coefficients of the exponential generating function using the  
 190 saddle point method, in which  $r$  is the saddle point (Lemma 7).

191 Restricting our focus to the  $HG(m, n, p)$  model and defining  $q = 1 - p$ , the following  
 192 corollary is a consequence of Theorem 2.

193 ► **Corollary 3.** *In the  $\mathcal{HG}(m, n, p)$  model, the probability of  $\ell$  given vertices to be a minimal*  
 194 *transversal, is as follows:*

$$195 \quad c_{m,\ell} = \begin{cases} (1 - z_\ell)^m \left(1 + O\left(\frac{\log m}{m^{1-\epsilon}}\right)\right) & \text{for } \ell \leq \log_{\frac{1}{q}} m^{1-\epsilon}, \\ \binom{m}{\ell} o_\ell^\ell \ell! \left(1 + O\left(\frac{\log m}{m^{1-\epsilon}}\right)\right) & \text{for } \ell \geq \log_{\frac{1}{q}} m^{1+\epsilon}, \\ e^{-mz_\ell} (1 - e^{-o_\ell r})^\ell \left(1 + O\left(\frac{1}{m^{1-2\epsilon}}\right)\right) & \text{for } \log_{\frac{1}{q}} m^{1-\epsilon} \leq \ell \leq \log_{\frac{1}{q}} m^{1+\epsilon} \end{cases}$$

196 Regarding the *cumulative sum*, we obtained results using the ordinary generating function.

197 ► **Proposition 3.** *In the  $\mathcal{HG}(m, n, p)$  model, we have*

$$198 \quad \bar{c}_{m,\ell} = \begin{cases} \Gamma\left(\frac{q}{p}\right) \frac{\ell!}{\Gamma\left(\frac{q}{p} + \ell + 1\right)} \frac{1}{o_\ell} \left(1 + O\left(\frac{\log m}{m^\epsilon}\right)\right) & \text{for } \ell \leq \log_{\frac{1}{q}} m^{1-\epsilon}, \\ \binom{m+1}{\ell+1} o_\ell^\ell \ell! \left(1 + O\left(\frac{\log m}{m^{1-\epsilon}}\right)\right) & \text{for } \ell \geq \log_{\frac{1}{q}} m^{1+\epsilon} \end{cases}$$

199 and for  $\log_{\frac{1}{q}} m^{1-\epsilon} \leq \ell \leq \log_{\frac{1}{q}} m^{1+\epsilon}$

$$200 \quad \ell! o_\ell^\ell \binom{j_0}{\ell} (1 - z_\ell - \ell o_\ell)^{j_0 - \ell} \leq \bar{c}_{m,\ell} \leq (m - \ell)! o_\ell^\ell \binom{j_1}{\ell} (1 - z_\ell)^{j_1 - \ell}$$

$$201 \quad \text{with } j_0 = \min\left(m, \left\lfloor \frac{\ell}{z_\ell + \ell o_\ell} \right\rfloor\right), \quad j_1 = \min\left(m, \left\lfloor \frac{\ell}{z_\ell} \right\rfloor\right).$$

## XX:8 On the average complexity of Berge Algorithm

202 The method to obtain the first two asymptotics is similar to the one used to extract  $c_{m,\ell}$   
 203 and is described in Lemmas 13 and 14. The proof of the last inequalities, given in Lemma 15  
 204 is a direct consequence of intermediates bounds used in the proof of Lemmas 13.

Let  $av(n, m, \ell)$  be the average number of minimal transversals of size  $\ell$  in the  $\mathcal{HG}(m, n, p)$  model. The average number of minimal transversal is equal to

$$\sum_{\ell=1}^{\min\{m,n\}} av(n, m, \ell) = \sum_{\ell=1}^{\min\{m,n\}} \binom{n}{\ell} m! [u^m] S_\ell(u) = \sum_{\ell=1}^{\min\{m,n\}} \binom{n}{\ell} c_{m,\ell}$$

205 We obtain the asymptotics when  $m$  is greater than  $n$ .

► **Proposition 4.** *In the  $\mathcal{HG}(m, n, p)$  model, assume that there exists a constant  $0 < \alpha < 1$  such that  $n = m^\alpha$ . The average number of minimal transversals in the  $\mathcal{HG}(m, n, p)$  model is asymptotically equivalent to*

$$av(n, m, \lfloor j \rfloor) + av(n, m, \lfloor j + 1 \rfloor) + av(n, m, \lfloor j + 2 \rfloor)$$

206 with  $j = \log_{\frac{1}{1-p}}(\frac{mp}{\kappa}) + 1$  where  $\kappa$  is the unique solution to  $\alpha + \log_{\frac{1}{1-p}}(\frac{1-e^{-(1-p)\kappa}}{1-e^{-\kappa}}) = 0$

207 Note that the asymptotic is concentrated around only three positions. First we prove  
 208 that  $av(n, m, \lfloor j + 1 \rfloor)$  is maximal by studying the ratio of two consecutive elements. This  
 209 ratio is decreasing and is close to 1 for  $\ell$  around  $j$ . It also tends to 0 for  $\ell > j + 1$  or to  $+\infty$   
 210 for  $\ell < j - 1$  which implies that the other elements are negligible. These last two points use  
 211 the fact that  $n = m^\alpha$  with  $\alpha < 1$ . For  $\alpha > 1$ , the asymptotic is smoother but it is out of  
 212 scope of this article. Also, in practice (the *Frequent Itemset Mining Dataset Repository* for  
 213 instance), the assumption that  $\alpha$  is smaller than 1, that is  $m > n$ , seems reasonable. The  
 214 detailed proof is given in Section 7.3.

The complexity of Berge algorithm is lower and upper bounded, up to a polynomial term, by the double sum given in Equation (1) which involves the cumulative sums  $\bar{c}_{m,\ell}$ . Basing ourselves on Proposition 3, we consider the following decomposition:

$$\sum_{i=1}^{m-1} \mathbb{E}_\pi [Tr_{i,n}] = \sum_{\ell=1}^{\log_{\frac{1}{q}} m^{1-\epsilon}} \binom{n}{\ell} \bar{c}_{m,\ell} + \sum_{\ell=\log_{\frac{1}{q}} m^{1-\epsilon}}^{\log_{\frac{1}{q}} m^{1+\epsilon}} \binom{n}{\ell} \bar{c}_{m,\ell} + \sum_{\ell=\log_{\frac{1}{q}} m^{1+\epsilon}}^{\min\{n,m\}} \binom{n}{\ell} \bar{c}_{m,\ell}$$

215 By proving the first and the third sum are equivalent to the first and the last term of the  
 216 second sum, we conclude that the second sum is the one that matter asymptotically. Then,  
 217 by identifying the  $\ell$  that maximizes the coefficient  $\bar{c}_{m,\ell}$ , we deduce an upper bound on the  
 218 average complexity of the Berge algorithm.

► **Proposition 5.** *In the  $\mathcal{HG}(m, n, p)$  model, assume that there exists a constant  $1 - \frac{p}{\ln(\frac{1}{1-p})} < \alpha < 1$  such that  $n = m^\alpha$ . For all  $\epsilon > 0$ , the average complexity of the Berge algorithm satisfies*

$$\mathbb{E}_p[B(\mathcal{H})] \leq \epsilon n^2 m^2 (\log_{\frac{1}{q}} m) \ell! o_\ell^\ell \binom{n}{\ell} \binom{m}{\ell} (1 - z_\ell)^{m-\ell} (1 + o(1))$$

219 with  $\ell = \log_{\frac{1}{q}}(m) - \log_{\frac{1}{q}}(\ln m) - \log_{\frac{1}{q}}(\frac{1-\alpha}{p}) + o(1)$ .

220 The complete proof is given in Section 7.4. The condition on  $\alpha$  is mainly technical and other  
 221 results can be obtained if  $1 - \frac{p}{\ln(\frac{1}{1-p})} \geq \alpha$ .

222 To evaluate the result of Proposition 5 in  $\ell = \log_{\frac{1}{q}}(m) - \log_{\frac{1}{q}}(\ln m) - \log_{\frac{1}{q}}(\frac{1-\alpha}{p}) + o(1)$   
 223 helps clarifying the asymptotics.

224 ► **Theorem 4.** *In the  $\mathcal{HG}(m, n, p)$  model, assume that there exists a constant  $1 - \frac{p}{\ln(\frac{1}{1-p})} <$   
 225  $\alpha < 1$  such that  $n = m^\alpha$ . The average complexity of the Berge algorithm is*

$$\mathbb{E}_p[B(\mathcal{H})] = \mathcal{O}\left(n^{\ell+2} m^4 \log_{\frac{1}{q}}(m) (1-\alpha)^\ell \frac{1}{m^{\frac{1-\alpha}{p}}}\right)$$

226 with  $\ell = \log_{\frac{1}{q}}(m) - \log_{\frac{1}{q}}(\ln m) - \log_{\frac{1}{q}}(\frac{1-\alpha}{p}) + o(1)$ .

227 Note that a more precise upper bound could be obtained: we upper bounded a ratio  $\frac{\ln(m)^{\ell-1}}{l!}$   
 228 by  $m$  in the formula to make it easier to read. Indeed, we can now clearly see that the  
 229 dominant term is a quasi-polynomial in  $n$ . It can be proven using Proposition 4 that the  
 230 number of solutions computed by the Berge algorithm is also quasi-polynomial. Therefore  
 231 we can conclude with the following corollary.

232 ► **Corollary 5.** *In the  $\mathcal{HG}(m, n, p)$  model, assume that there exists a constant  $1 - \frac{p}{\ln(\frac{1}{1-p})} <$   
 233  $\alpha < 1$  such that  $n = m^\alpha$ . The average output complexity of the Berge algorithm is quasi-linear.*

## 234 5.1 Another Point of view

235 In this section, we show how the Poisson-Mellin-Newton cycle can be used as a transformation  
 236 between the analytic combinatorics point of view and the probabilistic point of view.

237 Restricting ourselves  $\mathcal{HG}(m, n, p)$  model, we recall that the probabilistic generating  
 238 function that a given set of  $\ell$  vertices is a minimal transversal is:

$$239 S_\ell(u) = e^{(1-(1-p)^\ell)u} \left(1 - e^{-p(1-p)^{\ell-1}u}\right)^\ell$$

240 The Poisson generating function is  $P_\ell(u) = e^{-u} S_\ell(u)$ .

241 It can be shown that, after applying a Mellin transform, we obtain:

$$242 M_\ell(s) = \sum_{k=0}^{\ell} (-1)^{\ell-k} \binom{\ell}{k} \frac{\Gamma(s)}{(1-p)^{(l-1)s} (1-p+lp-kp)^s}$$

243 The coefficient of the original function are now given by the Rice integral:

$$244 c_{m,\ell} = \frac{1}{2\pi i} \int_{\Lambda} \frac{\Gamma(m+1)}{\Gamma(m+s+1)} M_\ell(s) ds$$

After several steps of simplification we obtain:

$$c_{m,\ell} = \sum_{i=\ell}^m \binom{m}{i} l! \left\{ \begin{matrix} i \\ l \end{matrix} \right\} (1 - (1-p)^l - l \cdot p(1-p)^{l-1})^{m-i} \cdot (p(1-p)^{l-1})^i$$

245 This formula has a combinatorial interpretation: given a set of size  $\ell$ , its probability to be a  
 246 minimal transversal of a random hypergraph with  $m$  hyperedges is given by:

247 ■ there is at least  $\ell$  lines that contains only one 1, that is at least one per vertex. Assume  
 248 there are  $i$  such lines. Then there is  $\binom{m}{i}$  possibilities to select such lines.

249 ■ then, the  $i$  lines have to be distributed amongst the  $\ell$  vertices, that is a set composition  
 250 of  $\{1, \dots, i\}$  into  $\ell$  subsets. There are  $l! \left\{ \begin{matrix} i \\ l \end{matrix} \right\}$  ways to do so, where  $\left\{ \begin{matrix} i \\ l \end{matrix} \right\}$  are the Stirling  
 251 number of the second kind.

252 ■ The probability that each of the  $i$  lines contains exactly one 1 is given by  $(p(1-p)^{l-1})^i$ .

253 ■ The probability that the rest of the  $m-i$  lines contain at least two 1 is given by  
 254  $(1 - (1-p)^l - l \cdot p(1-p)^{l-1})^{m-i}$ .

255 Note that such closed formula did not appear in [11].

256 **6 Perspectives**

257 In the future, we would like to consider a third model that we consider powerful enough to  
 258 capture the complexity of real datasets but is in a sense simpler since all the columns have  
 259 the same probabilistic behaviour.

260 ► **Definition 3** ( $\mathcal{HG}(m, n, \mu)$  random model). *Fix  $\mu$  a probability measure over  $[0, 1]$ . In  
 261 the  $\mathcal{HG}(m, n, \mu)$  model, the incidence matrix  $M(\mathbf{H})$  is generated according to the following  
 262 procedure:*

- 263 1. *Compute a random vector  $\pi = (\pi_1, \dots, \pi_n)$  where  $(\pi_1, \dots, \pi_n)$  forms a family of i.i.d*  
 264 *random variables with probability measure  $\mu$ ;*
- 265 2. *Compute  $M(\mathbf{H})$  according to the multiparametric model  $\mathcal{HG}(m, n, \pi)$ .*

266 With the  $\mathcal{HG}(m, n, \mu)$  random model, the distribution of the degrees is controlled and it  
 267 is possible to describe the influence of the degrees on the number of minimal transversals.  
 268 The columns of the matrix remain independent but the rows are no more independent since  
 269 they share the same random variable  $\pi$ . Models are closely related:

$$270 \quad \Pr_{m,n,\mu}[\mathbf{H}] = \int_{[0,1]^n} \Pr_{m,n,\pi}[\mathbf{H}] d\mu_{\otimes n}(\pi) \quad (2)$$

271 with  $\mu_{\otimes n}$  the product measure  $\mu \otimes \dots \otimes \mu$  ( $n$  times). Also, if we consider the Dirac Measure  
 272  $\mu = \delta_p$ , then  $\mathcal{HG}(m, n, \delta_p)$  is exactly the  $\mathcal{HG}(m, n, p)$  model. To the best of our knowledge,  
 273 this idea is new in the field of algorithm analysis and so is our approach to obtain results  
 274 in this random model. It is general enough to capture the behaviour of algorithm in both  
 275 classical models like the Erdős-Renyi and on real-world hypergraphs. Though we only gave  
 276 the average complexity of Berge algorithm in the  $\mathcal{HG}(m, n, p)$  model, the results we presented  
 277 already allow to numerically compute the average cost of the algorithm in both  $\mathcal{HG}(m, n, \pi)$   
 278 and  $\mathcal{HG}(m, n, \mu)$  models. In our perspectives, we would like to obtain the average number of  
 279 minimal transversals and complexity of Berge algorithm for other measures  $\mu$  like

- 280 ■ the uniform distribution between two values  $a$  and  $b$ , that is  $\mu = \mathcal{U}_{[a, b]}$ ;
- 281 ■ a measure that combines two Dirac measures;
- 282 ■ the Dirac measure (or one of the two previously mentioned), when the probability  $p$  is  
 283 close to 0 and/or when  $n$  is not polynomial in  $m$ .

284 Note that, in the  $\mathcal{HG}(m, n, p)$  model, we claim to be able to extract the average number  
 285 of minimal transversals with  $n$  vertices and  $m$  hyperedges when  $m < n$ , that is  $n = m^\alpha$   
 286 with  $\alpha > 1$ . We omitted the proof for the clarity of the paper and since we are convinced  
 287 that, in practice, it is safe to assume that  $\alpha < 1$ . The case where  $n$  is almost equal to  $m$   
 288 remains a challenge though. Our results could also be applied to study algorithms which are  
 289 improvements of Berge Algorithm, as mentioned in the introduction. Finally, in order to fully  
 290 understand why Berge algorithm is more efficient in practice than the one by Fredman and  
 291 Khachiyan [20], one would also require to analyse this algorithm in various random models.

## 7 Details of the proofs

### 7.1 Details for the EGF

In the  $\mathcal{HG}(m, n, \pi)$  model, for a given distribution  $\pi$ , the *e.g.f* is

$$S_\ell(u) = e^{(1-z_\ell) \cdot u} \prod_{i=1}^{\ell} (1 - e^{-o_{\ell,i} \cdot u})$$

it can be rewritten as:

$$S_\ell(u) = e^{((1-z_\ell)u - \sum_{i=1}^{\ell} o_{\ell,i} u)} \prod_{i=1}^{\ell} (e^{o_{\ell,i} u} - 1) = e^{(1-z_\ell - \sum_{i=1}^{\ell} o_{\ell,i})u} \prod_{i=1}^{\ell} (e^{o_{\ell,i} u} - 1)$$

Our goal is to find  $m![u^m]S_\ell(u)$ , that is the probability of a set with  $\ell$  vertices to be a minimal transversal in a hypergraph with  $m$  hyperedges. The idea of the proof is the following:  $S_\ell(u)$  is analytic, we can therefore use the saddle-point method to extract its coefficient, using a circle of radius  $r$  as a closed path. Indeed, using Cauchy's integral formula we have

$$[u^m]S_\ell(u) = \frac{1}{2\pi i} \int_{|u|=r} \frac{S_\ell(u)}{u^{m+1}} du$$

where  $r$  is between 0 and the radius of convergence. Let  $h(\theta) = \ln(S_\ell(re^{i\theta}))$ , we have

$$[u^m]S_\ell(u) = \frac{1}{2\pi r^m} \int_{-\pi}^{\pi} e^{h(\theta) - im\theta} d\theta$$

Using the Taylor expansion we obtain

$$[u^m]S_\ell(u) = \frac{1}{2\pi r^m} \int_{-\pi}^{\pi} e^{h(0) + \theta h'(0) + \frac{\theta^2}{2} h''(0) + \mathcal{O}(\theta^3 \sup_{\theta \in [-\pi, \pi]} |h'''(\theta)|) - im\theta} d\theta$$

Therefore, we need to find the first, second, and third derivative of  $h(\theta)$ .

$h'(\theta)$  can be computed as follows:

$$h'(\theta) = ire^{i\theta} \frac{S'_\ell(re^{i\theta})}{S_\ell(re^{i\theta})} = ia(re^{i\theta}), \quad \text{where } a(x) = x \frac{S'_\ell(x)}{S_\ell(x)}$$

$$a(x) = x \left[ (1 - z_\ell) + \sum_{i=1}^{\ell} \frac{o_{\ell,i} e^{-o_{\ell,i} x}}{1 - e^{-o_{\ell,i} x}} \right] = x(1 - z_\ell) + \sum_{i=1}^{\ell} \omega(o_{\ell,i} x)$$

$$\text{with } \omega(x) = \frac{x e^{-x}}{1 - e^{-x}} = \frac{x}{e^x - 1}$$

Now,  $h''(\theta)$  can be computed as:

$$h''(\theta) = -re^{i\theta} a'(re^{i\theta}) := -b(re^{i\theta}), \quad \text{with } b(x) = xa'(x)$$

$$\Rightarrow b(x) = x \left[ (1 - z_\ell) + \sum_{i=1}^{\ell} o_{\ell,i} \omega'(o_{\ell,i} x) \right]$$

$$= (1 - z_\ell)x + \sum_{i=1}^{\ell} o_{\ell,i} x \omega'(o_{\ell,i} x)$$

**XX:12 On the average complexity of Berge Algorithm**

312 Where,

$$313 \quad \omega'(x) = \frac{1}{e^x - 1} - x \frac{e^x}{(e^x - 1)^2} = \frac{e^x - 1 - xe^x}{(e^x - 1)^2}$$

314 And  $h'''(\theta)$  is obtained as:

$$315 \quad h'''(\theta) = -ire^{i\theta}b'(re^{i\theta}) = -ic(re^{i\theta})$$

316 with  $c(x) = xb'(x)$

$$317 \quad = x \left[ (1 - z_\ell) + \sum_{i=1}^{\ell} o_{\ell,i} \omega'(o_{\ell,i}x) + \sum_{i=1}^{\ell} o_{\ell,i}^2 x \omega''(o_{\ell,i}x) \right]$$

$$318$$

$$319 \quad = x(1 - z_\ell) + \sum_{i=1}^{\ell} o_{\ell,i} x \omega'(o_{\ell,i}x) + \sum_{i=1}^{\ell} (o_{\ell,i}x)^2 \omega''(o_{\ell,i}x)$$

$$320$$

$$321 \quad = b(x) + \sum_{i=1}^{\ell} (o_{\ell,i}x)^2 \omega''(o_{\ell,i}x)$$

322 We have

$$323 \quad [u^m]S_\ell(u) = \frac{S_\ell(r)}{2\pi r^m} \int_{-\pi}^{\pi} e^{i\theta(a(r)-m) - \frac{\theta^2}{2}b(re^{i\theta}) + \mathcal{O}(\theta^3 \sup_{\theta \in ]-\pi, \pi[} |h'''(\theta)|)} d\theta$$

324 We define the notion of  $H$ -admissibility as it was in [19].

325 **► Definition 6** ( $H$ -admissible). Let  $G(z)$  be a generating function with a radius of convergence  
326  $R$  with  $0 < R \leq \infty$ . Given  $R_0 < R$ , let  $G(z)$  be positive for  $z \in ]R_0, R[$ .  $G(z)$  is said to be  
327  $H$ -admissible if it satisfies the following conditions:

- 328 1. (Capture condition)  $\lim_{r \rightarrow R} a(r) = +\infty$  and  $\lim_{r \rightarrow R} b(r) = +\infty$   
329 2. (Locality condition) for some function  $\theta_0(r)$  defined over  $]R_0, R[$ , and satisfying  $0 < \theta_0 < \pi$ ,  
one has

$$G(re^{i\theta}) \sim G(r)e^{i\theta a(r) - \theta^2 b(r)/2} a_s \quad r \mapsto R,$$

329 uniformly in  $|\theta| \leq \theta_0(r)$

3. (Decay condition) Uniformly in  $\theta_0(r) \leq |\theta| < \pi$

$$G(re^{i\theta}) = o\left(\frac{G(r)}{\sqrt{b(r)}}\right)$$

330 **► Lemma 7.** In the  $\mathcal{HG}(m, n, \pi)$  model, when  $\ell = o(m)$ , the probability that a set with  $\ell$   
331 vertices is a minimal transversal is:

$$332 \quad m![u^m]S_\ell(u) = m! \frac{S_\ell(r)}{\sqrt{2\pi b(r)} r^m} \left(1 + o(b(r)^{-\frac{2}{3}})\right)$$

333 **Proof.** The idea of the proof is to evaluate the integral

$$334 \quad \int_{-\pi}^{\pi} e^{i\theta(a(r)-m) - \frac{\theta^2}{2}b(re^{i\theta}) + \mathcal{O}(\theta^3 \sup_{\theta \in ]-\pi, \pi[} |h'''(\theta)|)} d\theta$$

335 We decompose the integral into two, that is  $[u^m]S_\ell(u) = I_1 + I_2$  with

$$336 \quad I_1 = \int_{|\theta| \leq \varepsilon} e^{i\theta(a(r)-m) - \frac{\theta^2}{2}b(re^{i\theta}) + \mathcal{O}(\theta^3 \sup_{\theta \in ]-\pi, \pi[} |h'''(\theta)|)} d\theta$$

$$337$$

$$338 \quad I_2 = \int_{\varepsilon \leq |\theta| \leq \pi} e^{i\theta(a(r)-m) - \frac{\theta^2}{2}b(re^{i\theta}) + \mathcal{O}(\theta^3 \sup_{\theta \in ]-\pi, \pi[} |h'''(\theta)|)} d\theta$$

339 In order to use the saddle-point method and prove Lemma 7, we prove the following:

340 ■ we show there exists a unique saddle point  $r$  such that  $a(r) = m$  (Lemma 8).

341 ■ we show that  $r \sim \frac{m-\ell}{1-z_\ell}$  (Lemma 9).

342 ■ we show that  $b(r)$  tends to infinity with respect to  $r$  (Lemma 10)

343 ■ in  $I_1$ , we show that if  $\varepsilon = b(r)^{-\frac{4}{9}}$ , we have  $|\theta^3 h'''(\theta)| = m^{-\frac{1}{3}}$  (Lemma 11).

344 ■ we give an upper bound of  $I_2$  and show it is negligible compared to  $I_1$  (Lemma 12).

Note that this is sufficient to prove that  $S_\ell(u)$  is H-admissible. Once all those elements are proven, we have that

$$I_1 = (1 + O(m^{-\frac{1}{3}})) \int_{-\varepsilon}^{\varepsilon} e^{-\frac{\theta^2}{2} b(r e^{i\theta})} d\theta$$

Using  $\varepsilon = b(r)^{-\frac{4}{9}}$ , all that is left to study in  $I_1$  is the Gaussian integral

$$\int_{-b(r)^{-\frac{4}{9}}}^{b(r)^{-\frac{4}{9}}} e^{-\frac{\theta^2}{2} b(r e^{i\theta})} d\theta$$

Using a change of variable  $s = \sqrt{b(r)}\theta$ , and assuming  $b(r)$  tends to infinity we obtain

$$\int_{-\infty}^{\infty} e^{-\frac{s^2}{2}} \frac{ds}{\sqrt{b(r)}}$$

345 which is equal to  $\sqrt{\frac{2\pi}{b(r)}}$ , therefore concluding the proof.

346

347 In the context of Lemma 7, since the functions  $\omega(x)$  and  $x\omega'(x)$  are bounded on  $[0, +\infty[$ ,  
348 we have  $r = \frac{m}{1-z_\ell} + \mathcal{O}(\ell)$  and  $b(r) = m + (\mathcal{O}(\ell))$ . By replacing  $m!$  using the Stirling formula,  
349 we obtain the result announced in Theorem 2 when  $m^{-\varepsilon} \leq m z_\ell \leq m^\varepsilon$ .

350 ► **Lemma 8.** *There exists a unique  $r$  such that  $a(r) = m$ .*

351 **Proof.** Recall that  $a(x) = (1 - z_\ell)x + \sum_{i=1}^{\ell} \omega(o_{\ell,i}x)$ . So,

$$352 \quad a'(x) = (1 - z_\ell) + \sum_{i=1}^{\ell} o_{\ell,i} \omega'(o_{\ell,i}x)$$

353 and

$$354 \quad \omega'(x) = \frac{e^x}{(e^x - 1)^2} (1 - x - e^{-x}) \leq 0$$

355 so  $\omega(x)$  is decreasing.

$$356 \quad \omega''(x) = \frac{e^{2x}}{(e^x - 1)^3} (x + x e^{-x} - 2 + 2e^{-x}) \geq 0$$

357 Since  $a''(x) \geq 0$ , we have  $a'(x)$  is increasing and  $a'(0) = 1 - z_\ell + \sum_{i=1}^{\ell} o_{\ell,i} \omega'(0) = 1 - z_\ell -$   
358  $\frac{1}{2} \sum_{i=1}^{\ell} o_{\ell,i} > 0$ . So  $a'(x) > 0$  and  $a(x)$  is increasing.

359 Moreover,  $a(0) = \ell$  and  $\lim_{x \rightarrow \infty} a(x) = \infty$  and we can conclude that  $r$  exists uniquely. ◀

360 ► **Lemma 9.** *If  $\ell = o(m)$ , then  $r \sim \frac{m}{1-z_\ell}$ . Otherwise we have  $r \sim \frac{m-\ell}{1-z_\ell}$ .*

361 **Proof.** Assuming  $a(r) = m$ , we have  $r(1 - z_\ell) + \sum_{i=1}^{\ell} \omega(o_{\ell,i}r) = m$ . Using the fact that  
362  $\omega(x) \in [0, 1]$ , one has  $\frac{m-\ell}{1-z_\ell} \leq r \leq \frac{m}{1-z_\ell}$ .

363 For  $k > 1$ , if  $\ell \geq k \log m$ , then  $\ell z_\ell m \rightarrow 0$

$$364 \quad r(1 - z_\ell) + \sum_{i=1}^{\ell} \omega(o_{\ell,i}r) = r(1 - z_\ell) + \sum_{i=1}^{\ell} (1 + \mathcal{O}(o_{\ell,i}r)) = r(1 - z_\ell) + \ell + \mathcal{O}(\ell z_\ell m)$$

**XX:14** On the average complexity of Berge Algorithm

365 thus,  $m = r(1 - z_\ell) + \ell + O(\ell z_\ell m)$ , so  $r \sim \frac{m-\ell}{1-z_\ell}$ .

366 If  $\ell \leq \log m$ , then  $r \sim \frac{m}{1-z_\ell} \sim \frac{m-\ell}{1-z_\ell}$ .

367

368 ► **Lemma 10.** *If  $\ell = o(m)$ , then the function  $b(r)$  tends to infinity.*

369 **Proof.** We have

$$370 \quad b(r) = r(1 - z_\ell) + \sum_{i=1}^{\ell} o_{\ell,i} r \omega'(o_{\ell,i} r)$$

371 Using the fact that  $\omega'(o_{\ell,i} r) \leq 0$ , there exists a constant like  $A \leq 0$  such that  $A \leq o_{\ell,i} r \omega'(o_{\ell,i} r)$ ,  
372  $\forall i \leq \ell$ . Therefore we obtain

$$373 \quad (1 - z_\ell)r + \ell A \leq b(r) \leq (1 - z_\ell)r$$

374 According to Lemma 9, if  $\ell = o(m)$ ,  $r \sim \frac{m-\ell}{1-z_\ell}$  and  $b(r) \sim m$ .

375 Now consider the case that  $\ell \geq k \log m$  for large enough  $k$  such that  $o_{\ell,i} r = \mathcal{O}(\frac{1}{m^3})$ . In this  
376 case,  $o_{\ell,i} r \omega'(o_{\ell,i} r) \sim -\frac{1}{2} o_{\ell,i} r$  and so  $b(r) \sim (1 - z_\ell)r + \mathcal{O}(\frac{1}{m^2})$ . Thus  $b(r) \rightarrow \infty$ . ◀

377 ► **Lemma 11.** *if  $\varepsilon = b(r)^{-\frac{4}{9}}$ , we have  $|\theta^3 h'''(\theta)| = m^{-\frac{1}{3}}$*

378 **Proof.** Recall that  $h'''(\theta) = -ic(re^{i\theta})$ , and

$$379 \quad c(x) = x(1 - z_\ell) + \sum_{i=1}^{\ell} o_{\ell,i} x \omega'(o_{\ell,i} x) + \sum_{i=1}^{\ell} (o_{\ell,i} x)^2 \omega''(o_{\ell,i} x)$$

380 Define  $\bar{\omega}(u) := u\omega'(u) + u^2\omega''(u)$ , so

$$381 \quad c(x) = x(1 - z_\ell) + \sum_{i=1}^{\ell} \bar{\omega}(o_{\ell,i} x)$$

382 We have

$$383 \quad |x^k e^{\alpha x}| = |x|^k e^{\alpha \operatorname{Re}(x)} = |x|^k e^{\alpha |x| \cos(\theta)}$$

384 If  $|x| \rightarrow \infty$ , we have

$$385 \quad \omega''(x) \sim \frac{x e^{2x}}{(e^x - 1)^3} \quad \text{and} \quad \bar{\omega}(x) \sim \frac{|x|^3 e^{2|x| \cos(\theta)}}{e^{3|x| \cos(\theta)}}$$

386 which tends to zero if  $|\theta| < \frac{\pi}{2} - \varepsilon$ . So  $\bar{\omega}$  is bounded on the cone  $\{z; |\arg(z)| < \frac{\pi}{2} - \varepsilon\}$ . Thus  
387  $|c(x)| \leq (1 - z_\ell)|x| + O(\ell) \sim O(m)$ .

388 Now, define  $\varepsilon = b(r)^{-\gamma} \sim m^{-\gamma}$  such that  $\frac{1}{3} < \gamma < \frac{1}{2}$ . In particular let  $\gamma = \frac{4}{9}$  or  $\varepsilon = m^{-\frac{4}{9}}$ .  
389 Therefore,

$$390 \quad |\theta^3 h'''(\theta)| = m^{-3\gamma} m = m^{-\frac{1}{3}}$$

391

392 ► **Lemma 12.** *if  $\ell = o(m)$ , then  $I_2 = o(I_1)$*

393 **Proof.** Our objective here is to bound  $|S_\ell(re^{i\theta})|$  relatively to  $S_\ell(r)$ .

394 We have

$$395 \quad S_\ell(u) = e^{(1-z_\ell - \sum_{i=1}^{\ell} o_{\ell,i})u} \prod_{i=1}^{\ell} (e^{o_{\ell,i} u} - 1)$$

$e^{o_{\ell,i}u} - 1$  is a function with positive coefficients, therefore

$$|e^{o_{\ell,i}re^{i\theta}} - 1| \leq e^{o_{\ell,i}r} - 1$$

Also  $|e^{\alpha re^{i\theta}}| = |e^{\alpha r \cos(\theta) + i\alpha r \sin(\theta)}| = e^{\alpha r \cos(\theta)} = e^{\alpha r} e^{-\alpha r(1-\cos(\theta))}$ . For  $|\theta| \geq \varepsilon$  we obtain:

$$|e^{\alpha re^{i\theta}}| \leq e^{\alpha r} e^{-\alpha r(1-\cos(\varepsilon))}$$

And finally we have:

$$|S_{\ell}(re^{i\theta})| \leq e^{(1-z_{\ell}-\sum_{i=1}^{\ell} o_{\ell,i})r} \prod_{i=1}^{\ell} (e^{o_{\ell,i}r} - 1) e^{-(1-z_{\ell}-\sum_{i=1}^{\ell} o_{\ell,i})r(1-\cos(\varepsilon))}$$

using the definition of  $S_{\ell}(r)$  the inequality simplifies into

$$|S_{\ell}(re^{i\theta})| \leq S_{\ell}(r) e^{-(1-z_{\ell}-\sum_{i=1}^{\ell} o_{\ell,i})r(1-\cos(\varepsilon))} \leq S_{\ell}(r)$$

396 So we have

$$397 \quad I_2 = S_{\ell}(r) (1 + O(b(r)^{-\frac{1}{3}})) \int_{\varepsilon}^{2\pi-\varepsilon} e^{-\frac{\theta^2}{2}b(r)}$$

398

$$399 \quad \leq S_{\ell}(r) (1 + O(b(r)^{-\frac{1}{3}})) 2(\pi - \varepsilon) e^{-\frac{\varepsilon^2}{2}b(r)} = S_{\ell}(r) (1 + O(b(r)^{-\frac{1}{3}})) 2(\pi - \varepsilon) e^{-m^{\frac{4}{9}}}$$

400

## 401 7.2 Details for the OGF

402 In this section, we analyze the asymptotic behavior of the coefficients  $c_{m,\ell}$  and  $\bar{c}_{m,\ell}$  for large  
403 and small values of  $\ell$ , where  $\hat{S}_{\ell}(u) = \sum_{m \geq 0} c_{m,\ell} u^m$  and  $\bar{S}_{\ell}(u) = \sum_{m \geq 0} \bar{c}_{m,\ell} u^m$ . For the sake  
404 of simplicity, the results are given in the  $\mathcal{HG}(m, n, p)$  model but similar results can be stated  
405 in the  $\mathcal{HG}(m, n, \pi)$  model with some additional notations.

► **Lemma 13.** *In the  $\mathcal{HG}(m, n, p)$  model with  $\ell \geq \log_{\frac{1}{q}} m^{1+\varepsilon}$ , the coefficients satisfy*

$$c_{m,\ell} = \binom{m}{\ell} \ell! o_{\ell}^{\ell} \left( 1 + O\left(\frac{\log m}{m^{\varepsilon}}\right) \right),$$

$$\bar{c}_{m,\ell} = \binom{m+1}{\ell+1} \ell! o_{\ell}^{\ell} \left( 1 + O\left(\frac{\log m}{m^{\varepsilon}}\right) \right),$$

406 where the constants in  $O$  do not depend on  $\ell$ .

**Proof.** Recall that

$$\hat{S}_{\ell}(u) = u^{\ell} \ell! o_{\ell}^{\ell} \prod_{i=0}^{\ell} \frac{1}{1 - u(1 - z_{\ell} - io_{\ell})}$$

So,

$$u^{\ell} \ell! \frac{o_{\ell}^{\ell}}{(1 - u(1 - z_{\ell} - \ell o_{\ell}))^{\ell+1}} \preceq \hat{S}_{\ell}(u) \preceq u^{\ell} \ell! \frac{o_{\ell}^{\ell}}{(1 - u(1 - z_{\ell}))^{\ell+1}}$$

where  $f(u) \preceq g(u)$  means the coefficients of  $f$  are smaller than the coefficients of  $g$ . Using the following equality

$$\frac{1}{(1 - au)^{\ell+1}} = \sum_{k \geq 0} \binom{\ell + k}{\ell} a^k u^k$$

**XX:16 On the average complexity of Berge Algorithm**

407 the coefficients satisfy

$$408 \quad \ell! o_\ell^\ell \binom{m}{\ell} (1 - z_\ell - \ell o_\ell)^{m-\ell} \leq c_{m,\ell} \leq \ell! o_\ell^\ell \binom{m}{\ell} (1 - z_\ell)^{m-\ell}. \quad (3)$$

Since  $\ell \geq \log_{\frac{1}{q}} m^{1+\epsilon}$ ,  $(1 - z_\ell - \ell o_\ell)^{m-\ell} = 1 + O(\frac{\log m}{m^\epsilon})$  and  $(1 - z_\ell)^{m-\ell} = 1 + O(\frac{1}{m^\epsilon})$ . This concludes the proof for  $c_{m,\ell}$ . For the cumulative coefficients  $\bar{c}_{m,\ell}$ , the bounds on  $\bar{c}_{m,\ell}$  can be reused and we obtain for all  $\ell \leq \ell' \leq m$ ,

$$\ell! o_\ell^\ell \binom{\ell'}{\ell} (1 - z_\ell - \ell o_\ell)^{m-\ell} \leq c_{\ell',\ell} \leq \ell! o_\ell^\ell \binom{\ell'}{\ell}.$$

409 Summing over  $\ell'$  gives the announced result.

410 ◀

411 The following lemma is also written for  $\mathcal{HG}(m, n, p)$  model. The argument used in it works  
412 directly for  $\mathcal{HG}(m, n, \pi)$  model.

► **Lemma 14.** *In the  $\mathcal{HG}(m, n, p)$  model with  $\ell \leq \log_{\frac{1}{q}} m^{1-\epsilon}$ , the coefficients satisfy*

$$c_{m,\ell} = (1 - z_\ell)^m \left( 1 + O\left(\frac{\log m}{m^\epsilon}\right) \right),$$

$$\bar{c}_{m,\ell} = \Gamma\left(\frac{q}{p}\right) \frac{\ell!}{\Gamma\left(\frac{q}{p} + \ell + 1\right) o_\ell} \frac{1}{o_\ell} \left( 1 + O\left(\frac{\log m}{m^\epsilon}\right) \right)$$

413 **Proof.** The dominant singularity of  $\hat{S}_\ell(u)$  is  $\frac{1}{1-z_\ell}$ . To obtain  $c_{m,\ell}$ , we use the Cauchy  
414 coefficient formula. By choosing a circle with radius  $\frac{1}{1-z_\ell - \frac{o_\ell}{2}}$  as the contour, we have

$$415 \quad \frac{1}{2\pi i} \int_{C_r} \frac{\hat{S}_\ell(u)}{u^{m+1}} du = [u^m] \hat{S}_\ell(u) + \text{Re}\left(\frac{\hat{S}_\ell(u)}{u^{m+1}}, u = \frac{1}{1-z_\ell}\right)$$

416 and

$$417 \quad \text{Res}\left(\frac{\hat{S}_\ell(u)}{u^{m+1}}, u = \frac{1}{1-z_\ell}\right) = \text{Res}\left(\frac{u^\ell}{u^{m+1}} \ell! o_\ell^\ell \prod_{i=0}^{\ell} \frac{1}{1-u(1-z_\ell - i o_\ell)}, u = \frac{1}{1-z_\ell}\right)$$

$$418$$

$$419 \quad = \ell! (1-z_\ell)^{m+1-\ell} o_\ell^\ell \prod_{i=1}^{\ell} \frac{1-z_\ell}{i o_\ell} \text{Res}\left(\frac{1}{1-u(1-z_\ell)}, u = \frac{1}{1-z_\ell}\right) = -(1-z_\ell)^m$$

420 So,

$$421 \quad \frac{1}{2\pi i} \int_{C_r} \frac{\hat{S}_\ell(u)}{u^{m+1}} du = [u^m] \hat{S}_\ell(u) - (1-z_\ell)^m$$

422 On the other hand,  $|\frac{1}{2\pi i} \int_{C_r} \frac{\hat{S}_\ell(u)}{u^{m+1}} du| \leq \frac{\hat{S}_\ell(r)}{r^m}$ , and we have

$$423 \quad \frac{\hat{S}_\ell(r)}{r^m} = \ell! r^{\ell-m} o_\ell^\ell \prod_{i=0}^{\ell} \frac{1}{1-r(1-z_\ell - i o_\ell)} = -\frac{1}{2} (1-z_\ell - \frac{o_\ell}{2})^{m+1} \frac{1}{o_\ell} \frac{\Gamma(\frac{1}{2})}{\Gamma(\ell + \frac{1}{2})} \ell!$$

$$424$$

$$425 \quad = O\left((1-z_\ell - \frac{o_\ell}{2})^m \frac{1}{o_\ell} r^{\frac{1}{2}}\right)$$

426 To compute  $\bar{c}_{m,\ell}$ , we use the same argument as for computing  $c_{m,\ell}$ . Note that the  
 427 dominant singularity of  $\bar{S}_\ell(u)$  is at  $u = 1$ . Therefore, the contour is a circle of radius  
 428  $r = \frac{1}{1-\frac{z_\ell}{2}}$ . Consider the term  $\prod_{i=0}^{\ell} \frac{1}{1-u(1-z_\ell-io_\ell)}$ . We can simplify it as follows:  
 429 define  $\beta(u) = \frac{1-(1-z_\ell)u}{uo_\ell}$ , then

$$430 \quad \prod_{i=0}^{\ell} \frac{1}{1-u(1-z_\ell-io_\ell)} = \frac{1}{(uo_\ell)^{\ell+1}} \prod_{i=0}^{\ell} \frac{1}{\beta(u)+i} = \frac{1}{(uo_\ell)^{\ell+1}} \frac{\Gamma(\beta(u))}{\Gamma(\beta(u)+\ell+1)}$$

431 Note that  $\beta(1) = \frac{q}{p}$ , so we have:

$$432 \quad \text{Res}\left(\frac{\bar{S}_\ell(u)}{u^{m+1}}, u=1\right) = \ell! \frac{1}{o_\ell} \frac{\Gamma(\frac{q}{p})}{\Gamma(\frac{q}{p}+\ell+1)} \text{Res}\left(\frac{1}{1-u}, u=1\right) = -\ell! \frac{1}{o_\ell} \frac{\Gamma(\frac{q}{p})}{\Gamma(\frac{q}{p}+\ell+1)}$$

433 Thus

$$434 \quad \frac{1}{2\pi i} \int_{C_r} \frac{\bar{S}_\ell(u)}{u^{m+1}} du = [u^m] \hat{S}_\ell(u) - \frac{\ell!}{o_\ell} \frac{\Gamma(\frac{q}{p})}{\Gamma(\frac{q}{p}+\ell+1)}$$

435 But  $|\frac{1}{2\pi i} \int_{C_r} \frac{\bar{S}_\ell(u)}{u^{m+1}} du| \leq \frac{\bar{S}_\ell(r)}{r^m}$  and

$$436 \quad \frac{\bar{S}_\ell(r)}{r^m} = \ell! o_\ell^\ell (1-\frac{z_\ell}{2})^{m+2} \frac{2}{z_\ell} \frac{\Gamma(\frac{q}{2p})}{o_\ell^{\ell+1} \Gamma(\frac{q}{2p}+\ell+1)}$$

437

► **Lemma 15.** In the  $\mathcal{HG}(m, n, p)$  model with  $\log_{\frac{1}{q}} m^{1-\epsilon} \leq \ell \leq \log_{\frac{1}{q}} m^{1+\epsilon}$ , the cumulative sums satisfy

$$\ell! o_\ell^\ell \binom{j_0}{\ell} (1-z_\ell - \ell o_\ell)^{j_0-\ell} \leq \bar{c}_{m,\ell} \leq (m-\ell)! o_\ell^\ell \binom{j_1}{\ell} (1-z_\ell)^{j_1-\ell}$$

where  $j_0$  and  $j_1$  are the indexes such that the bounds in Inequality (3) are maximal. In particular, we have

$$j_0 = \min\left(m, \left\lfloor \frac{\ell}{z_\ell + \ell o_\ell} \right\rfloor\right), \quad j_1 = \min\left(m, \left\lfloor \frac{\ell}{z_\ell} \right\rfloor\right).$$

**Proof.** The proofs for the lower and upper bounds are similar. The inequalities in(3) give

$$\ell! o_\ell^\ell \max_{i=\ell..m} \binom{i}{\ell} (1-z_\ell - \ell o_\ell)^{i-\ell} \leq \bar{c}_{m,\ell} \leq (m-\ell)! o_\ell^\ell \max_{i=\ell..m} \binom{i}{\ell} (1-z_\ell)^{i-\ell}.$$

438 Studying the ratio between two consecutive elements in the max, it can be proved that the  
 439 sequences are increasing for  $j \leq j_0$  or  $j \leq j_1$  and then decreasing (if  $j_0 \neq m$  or  $j_1 \neq m$ ). This  
 440 concludes the proof. ◀

### 441 7.3 Details for the average number of minimal transversals

Let  $\ell' = \log_{\frac{1}{1-p}}(m^{2(\alpha+1)})$  and  $av(n, m, \ell) = \binom{n}{\ell} m! [u^m] S_\ell(u)$ . The average number of minimal transversals is

$$\sum_{\ell=1}^{\min\{m,n\}} av(n, m, \ell)$$

**XX:18 On the average complexity of Berge Algorithm**

Assuming that that  $n = m^\alpha$  with  $\alpha < 1$ , we split the formula into two parts:

$$\sum_{\ell=1}^{\ell'-1} av(n, m, \ell) + \sum_{\ell=\ell'}^n av(n, m, \ell)$$

442 The two following lemmas gives the asymptotic.

443 ► **Lemma 16.** *The asymptotic of the first sum is equal to  $av(n, m, \lfloor j \rfloor) + av(n, m, \lfloor j + 1 \rfloor) +$*   
 444  *$av(n, m, \lfloor j + 2 \rfloor)$*

**Proof.** there is a unique constant  $\kappa$  such that

$$\alpha + \log_{\frac{1}{1-p}} \left( \frac{1 - e^{-(1-p)\kappa}}{1 - e^{-\kappa}} \right) = 0$$

we define the real value  $j$  such that  $p(1-p)^{j-1} = \kappa$ . Let  $r_\ell$  be the solution of  $a(r_\ell) = m$ . We study the ratio:

$$\tilde{R}(n, m, \ell) = \frac{av(n, m, \ell + 1)}{av(n, m, \ell)} = \frac{n - \ell}{\ell + 1} \frac{S_{\ell+1}(r_{\ell+1})}{S_\ell(r_\ell)} \left( \frac{r_\ell}{r_{\ell+1}} \right)^m$$

445 It can be proven that:

- 446 ■  $\tilde{R}(n, m, \lfloor j \rfloor) \geq 1$
- 447 ■  $\tilde{R}(n, m, \lfloor j + 1 \rfloor) \leq 1$
- 448 ■ for all  $\ell < \lfloor j \rfloor$ ,  $\tilde{R}(n, m, \ell)$  tends to infinity uniformly in  $m$
- 449 ■ for all  $\ell > \lfloor j + 1 \rfloor$ ,  $\tilde{R}(n, m, \ell)$  tends to 0 uniformly in  $m$ .

450

451 ► **Lemma 17.** *The second sum is equal to  $\mathcal{O}(\frac{1}{m})$*

452 **Proof.** ■ first we prove that  $\binom{n}{\ell'} [u^m] S_{\ell'}(u) = \mathcal{O}(\frac{1}{m^{\alpha+1}})$ . We have

$$\begin{aligned} \binom{n}{\ell'} [u^m] S_{\ell'}(u) &\sim \binom{n}{\ell'} \frac{m!}{(m-\ell)! p^{\ell'} (1-p)^{\ell'(\ell'-1)}} \\ &\leq n^{\ell'} m^{\ell'} (1-p)^{\ell'(\ell'-1)} \\ &\leq \frac{m^{(\alpha+1)\ell'}}{m^{2(\alpha+1)(\ell'-1)}} \\ &\leq \frac{1}{m^{\alpha+1}} \end{aligned}$$

456 ■ Then we show that  $\frac{av(n, m, \ell+1)}{av(n, m, \ell)} < 1$ , for all  $\ell \geq \ell'$ . For all  $\ell \in \{1, \dots, n-1\}$ , we have

$$\frac{av(n, m, \ell + 1)}{av(n, m, \ell)} = \frac{p(n-\ell)(m-\ell)(1-p)^{2\ell}}{(\ell+1)} \leq pnm(1-p)^{2\ell}$$

For all value  $\ell \geq \ell'$  we obtain

$$\frac{av(n, m, \ell + 1)}{av(n, m, \ell)} \leq p$$

457 Therefore we have the  $\sum_{\ell=\ell'}^n av(n, m, \ell) = \mathcal{O}(\frac{1}{m})$ .

458

## 459 7.4 Details for the average case complexity of the Berge algorithm

460 This section is devoted to the proof of Proposition 5 on the average case complexity of the  
 461 Berge algorithm. By hypothesis, we have  $n = m^\alpha$  with  $\alpha < 1$ . The average complexity  
 462 satisfies the upper bound

$$463 \quad \mathbb{E}_p[B(\mathcal{H})] \leq n \cdot m \cdot \min\{n, m\} \sum_{i=1}^m \mathbb{E}_\pi[Tr_{i,n}] \leq m^{2\alpha+1} \sum_{\ell=1}^n \binom{n}{\ell} \bar{c}_{m,\ell}$$

464 with  $\bar{c}_{m,\ell} = \sum_{i=\ell}^m c_{m,i}$ . According to Proposition 3, the sum is split into three parts:  $\ell \leq \ell^-$ ,  
 465  $\ell \geq \ell^+$  and  $\ell^- \leq \ell \leq \ell^+$  with  $\ell^- = \log_{\frac{1}{q}} m^{1-\epsilon}$  and  $\ell^+ = \log_{\frac{1}{q}} m^{1+\epsilon}$  for an arbitrary fixed  
 466 positive  $\epsilon$ .

467 ► **Lemma 18.** *The first and third sums satisfy*

$$468 \quad \sum_{\ell=1}^{\ell^-} \binom{n}{\ell} \bar{c}_{m,\ell} = \binom{n}{\ell^-} \bar{c}_{m,\ell^-} (1 + o(1)), \quad \text{and} \quad \sum_{\ell=\ell^+}^n \binom{n}{\ell} \bar{c}_{m,\ell} = \binom{n}{\ell^+} \bar{c}_{m,\ell^+} (1 + o(1)).$$

469 **Proof.** The proof is similar in the two cases. First the ratio  $R_\ell = \binom{n}{\ell+1} \bar{c}_{m,\ell+1} / \binom{n}{\ell} \bar{c}_{m,\ell}$  is  
 470 calculated using the equivalences of Proposition 3 and we prove it is decreasing. After, we  
 471 show that  $R_{\ell^-}$  tends to infinity whereas  $R_{\ell^+}$  tends to 0. Then we can conclude and the  $o(1)$   
 472 can be replaced by  $\mathcal{O}(1/R_{\ell^-})$  for the first sum and  $\mathcal{O}(R_{\ell^+})$  for the second one. Indeed the  
 473 equality in Proposition 3 gives for  $\ell \leq \ell^-$  and  $\ell \geq \ell^+$

$$474 \quad R_\ell = \frac{n-\ell}{\frac{p}{q} + \ell + 1} \frac{1}{q} \left( 1 + \mathcal{O}\left(\frac{\log m}{m^\epsilon}\right) \right), \quad R_\ell = \frac{(n-\ell)(m-\ell)}{\ell+2} pq^{2\ell} \left( 1 + \mathcal{O}\left(\frac{\log m}{m^\epsilon}\right) \right).$$

475 It proves that the ratio  $R_\ell$  is decreasing with  $\ell$  in both cases. The ratio  $R_{\ell^-}$  is a  $\Theta\left(\frac{n}{\log m}\right)$   
 476 which tends to  $+\infty$ . The ratio  $R_{\ell^+}$  is a  $\Theta\left(\frac{1}{m^{1-\alpha+2\epsilon} \log m}\right)$  which tends to 0 for  $\epsilon$  small enough.  
 477 By induction on  $\ell$ ,  $\binom{n}{\ell} \bar{c}_{m,\ell} \leq \binom{n}{\ell^-} \bar{c}_{m,\ell^-} R_{\ell^-}^{\ell-\ell^-}$  and  $\binom{n}{\ell} \bar{c}_{m,\ell} \leq \binom{n}{\ell^+} \bar{c}_{m,\ell^+} R_{\ell^+}^{\ell-\ell^+}$ . Summing over  
 478  $\ell$  gives the two assertions. ◀

479 Note that Lemma 18 entails that the asymptotic in the complete sum is given by the  
 480 asymptotic over the  $\ell$  in the intermediate range  $\ell^- \leq \ell \leq \ell^+$ . Using once more the upper  
 481 bound in Proposition 3, we obtain

$$482 \quad \sum_{\ell=1}^n \binom{n}{\ell} \bar{c}_{m,\ell} = (1 + o(1)) \sum_{\ell=\ell^-}^{\ell^+} \binom{n}{\ell} \bar{c}_{m,\ell} \\ 483 \quad \leq (1 + o(1)) m \sum_{\ell=\ell^-}^{\ell^+} \binom{n}{\ell} \ell! o_\ell^\ell \binom{j_\ell}{\ell} (1 - z_\ell)^{j_\ell - \ell} \quad (4)$$

with  $j_\ell = j_\ell(m, \ell) = \min\left(m, \left\lfloor \frac{\ell}{z_\ell} \right\rfloor\right)$ . We now deal with the asymptotic of the right hand  
 side, using once more the upper bound in Proposition 3. For this, we consider once more the  
 ratio  $R_\ell$  of two successive elements,

$$R_\ell = \frac{n-\ell}{\ell+1} pq^{2\ell} \frac{j_{\ell+1}!}{(j_{\ell+1} - \ell - 1)!} \frac{(j_\ell - \ell)!}{j_\ell!} \frac{(1 - z_{\ell+1})^{j_{\ell+1} - \ell - 1}}{(1 - z_\ell)^{j_\ell - \ell}}.$$

**XX:20 On the average complexity of Berge Algorithm**

► **Lemma 19.** *The ratio  $R_\ell$  is decreasing and the maximum value in the sum (4) is obtained for  $\ell = \log_{\frac{1}{q}}(m) - \log_{\frac{1}{q}}(\ln m) - \log_{\frac{1}{q}}(\frac{1-\alpha}{p}) + o(1)$  when  $\alpha \geq 1 - \frac{p}{\ln(1-p)}$ . In particular, we have*

$$\sum_{\ell=1}^n \binom{n}{\ell} \bar{c}_{m,\ell} \leq (1 + o(1))m(\ell^+ - \ell^-) \binom{n}{\ell} \ell! o_\ell^\ell \binom{m}{\ell} (1 - z_\ell)^{m-\ell}.$$

485 **Proof.** To study the ratio, it is necessary to distinguish the case  $j_\ell < m$  and the case  $j_\ell = m$ .  
 486 Though, in both cases, the ratio is decreasing. The  $\ell$  given in the lemma only comes from  
 487 exact calculations. Note that the condition on  $\alpha$  ensures that  $j_\ell = m$ . ◀

**7.5 Details for the Poisson-Mellin-Newton Cycle**

The generating function is in the following form:

$$S_\ell(u) = e^{(1-(1-p)^\ell)u} \left(1 - e^{-p(1-p)^{\ell-1}u}\right)^\ell$$

491 The Poisson generating function is  $P_l(u) = e^{-u} S_l(u) = e^{-u(1-p)^\ell} \left(1 - e^{-p(1-p)^{\ell-1}u}\right)^\ell$

492 The Mellin transform is as follows:

$$M_l(s) = \frac{1}{(1-p)^{sl}} \int_0^\infty v^{s-1} e^{-v} (1 - e^{-\frac{p}{1-p}v})^l dv$$

494 Using the Binomial theorem we have:

$$\begin{aligned} M_l(s) &= \frac{1}{(1-p)^{sl}} \int_0^\infty v^{s-1} e^{-v} \left(\sum_{k=0}^l \binom{l}{k} (-1)^{l-k} e^{-\frac{(l-k)p}{1-p}v}\right) dv \\ &= \sum_{k=0}^l \binom{l}{k} (-1)^{l-k} \frac{1}{(1-p)^{sl}} \int_0^\infty v^{s-1} e^{-v(\frac{1-p+lp-kp}{1-p})} dv = \end{aligned}$$

498 Then by changing the variable  $v(\frac{1-p+lp-kp}{1-p}) = u$  we have  $\int_0^\infty v^{s-1} e^{-v(\frac{1-p+lp-kp}{1-p})} dv =$   
 499  $\int_0^\infty u^{s-1} e^{-u(\frac{1-p}{1-p+lp-kp})} du = (\frac{1-p}{1-p+lp-kp})^s \Gamma(s)$   
 500 and therefore

$$M_l(s) = \sum_{k=0}^l (-1)^{l-k} \binom{l}{k} \frac{\Gamma(s)}{(1-p)^{(l-1)s} (1-p+lp-kp)^s}$$

502 According to Poisson-Mellin-Newton cycle, we know that the coefficients of the p.g.f. are  
 503 obtained as follows by simplifying the Nørlund-Rice Integral:

$$C_{m,l} = \frac{1}{2\pi i} \int_\Lambda \frac{\Gamma(m+1)}{\Gamma(m+s+1)} M_l(s) ds$$

505 But

$$\int_\Lambda \frac{M_l(s)}{\Gamma(m+s+1)} ds = \sum_{k=0}^l (-1)^{l-k} \binom{l}{k} \int_\Lambda \frac{ds}{(1-p)^{(l-1)s} (1-p+lp-kp)^s (s)(s+1) \cdots (s+m)}$$

507 Let

$$F(s) = \frac{1}{(1-p)^{(l-1)s} (1-p+lp-kp)^s (s)(s+1) \cdots (s+m)}$$

508

509 then according to the Residue Theorem we have:

$$510 \quad \int_{\Lambda} \frac{ds}{(1-p)^{(l-1)s}(1-p+lp-kp)^s(s)(s+1)\cdots(s+m)} = 2\pi i \sum_{j=0}^m \text{Res}[F(s), s = -j]$$

511 But

$$512 \quad \text{Res}[F(s), s = -j] = \frac{(-1)^j}{j!(m-j)!} \left( \frac{1}{(1-p)^{(l-1)}(1-p+lp-kp)} \right)^{-j}$$

513 So

$$514 \quad \sum_{j=0}^m \text{Res}[F(s), s = -j] = \frac{1}{m!} \sum_{j=0}^m \binom{m}{j} (-1)^j ((1-p)^{(l-1)}(1-p+lp-kp))^j = \frac{1}{m!} (1 - (1-p)^{l-1}(1-p+lp-kp))^m$$

515 Thus

$$516 \quad 2\pi i \sum_{j=0}^m \text{Res}[F(s), s = -j] = \frac{2\pi i}{m!} (1 - (1-p)^{l-1}(1-p+lp-kp))^m$$

517 Therefore  $C_{m,l}$  is as follows:

$$518 \quad C_{m,l} = \sum_{k=0}^l (-1)^{l-k} \binom{l}{k} (1 - [(1-p)^{l-1}(1-p+lp-kp)])^m$$

$$519 \quad = \sum_{k=0}^l (-1)^{l-k} \binom{l}{k} \left[ \sum_{i=0}^m \binom{m}{i} (1 - (1-p)^l - lp(1-p)^{l-1})^{m-i} (p(1-p)^{l-1}k)^i \right]$$

$$520 \quad = \sum_{i=0}^m \binom{m}{i} (1 - (1-p)^l - lp(1-p)^{l-1})^{m-i} (p(1-p)^{l-1})^i \left( \sum_{k=0}^l (-1)^{l-k} \binom{l}{k} k^i \right)$$

$$521$$

$$522 \quad = \sum_{i=0}^m \binom{m}{i} i! \left\{ \begin{matrix} i \\ l \end{matrix} \right\} (1 - (1-p)^l - lp(1-p)^{l-1})^{m-i} (p(1-p)^{l-1})^i \quad (5)$$

523 in which  $\left\{ \begin{matrix} i \\ l \end{matrix} \right\}$  is Stirling numbers of the second kind.

## 524 ——— References ———

- 525 **1** Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of  
526 items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference*  
527 *on Management of data*, pages 207–216, 1993.
- 528 **2** James Bailey, Thomas Manoukian, and Kotagiri Ramamohanarao. A fast algorithm for  
529 computing hypergraph transversals and its application in mining emerging patterns. In *Third*  
530 *IEEE International Conference on Data Mining*, pages 485–488. IEEE, 2003.
- 531 **3** James Bailey and Zhou Zhu. Fast discovery of interesting collections of web services. In *2006*  
532 *IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference*  
533 *Proceedings)(WI'06)*, pages 152–160. IEEE, 2006.
- 534 **4** Alexandre Bazin, Laurent Beaudou, Giacomo Kahn, and Kaveh Khoshkhan. Bounding the  
535 number of minimal transversals in tripartite 3-uniform hypergraphs. *Discrete Mathematics &*  
536 *Theoretical Computer Science*, 23(Special issues), 2023.

## XX:22 On the average complexity of Berge Algorithm

- 537 5 Claude Berge. *Hypergraphs*. Elsevier, 3rd edition, 1989.
- 538 6 Endre Boros, Khaled Elbassioni, Vladimir Gurvich, and Leonid Khachiyan. Generating  
539 maximal independent sets for hypergraphs with bounded edge-intersections. In *Latin American*  
540 *Symposium on Theoretical Informatics*, pages 488–498. Springer, 2004.
- 541 7 Endre Boros, Khaled Elbassioni, Vladimir Gurvich, Leonid Khachiyan, and Kazuhisa Makino.  
542 An intersection inequality for discrete distributions and related generation problems. In  
543 *International Colloquium on Automata, Languages, and Programming*, pages 543–555. Springer,  
544 2003.
- 545 8 Endre Boros and Vladimir Gurvich. On nash-solvability in pure stationary strategies of  
546 finite games with perfect information which may have cycles. *Mathematical social sciences*,  
547 46(2):207–241, 2003.
- 548 9 Andreas Brandstädt, Van Bang Le, and Jeremy P Spinrad. *Graph classes: a survey*. SIAM,  
549 1999.
- 550 10 Peter Damaschke. Parameterized enumeration, transversals, and imperfect phylogeny recon-  
551 struction. *Theoretical Computer Science*, 351(3):337–350, 2006.
- 552 11 Julien David, Loïck Lhote, Arnaud Mary, and François Rioult. An average study of hypergraphs  
553 and their minimal transversals. *Theoretical Computer Science*, 596:124–141, 2015.
- 554 12 Pierpaolo Degano. *Automata, Languages and Programming: 24th International Colloquium,*  
555 *ICALP'97, Bologna, Italy, July 7-11, 1997, Proceedings*, volume 24. Springer Science &  
556 Business Media, 1997.
- 557 13 Guozhu Dong and Jinyan Li. Mining border descriptions of emerging patterns from dataset  
558 pairs. *Knowledge and Information Systems*, 8:178–202, 2005.
- 559 14 Jean-Guillaume Dumas, Frank Heckenbach, David Saunders, and Volkmar Welker. Computing  
560 simplicial homology based on efficient smith normal form algorithms. In *Algebra, Geometry*  
561 *and Software Systems*, pages 177–206. Springer, 2003.
- 562 15 Thomas Eiter and Georg Gottlob. Identifying the minimal transversals of a hypergraph and  
563 related problems. *SIAM Journal on Computing*, 24(6):1278–1304, 1995.
- 564 16 Thomas Eiter and Georg Gottlob. Hypergraph transversal computation and related problems  
565 in logic and ai. In *European Workshop on Logics in Artificial Intelligence*, pages 549–564.  
566 Springer, 2002.
- 567 17 Thomas Eiter, Georg Gottlob, and Kazuhisa Makino. New results on monotone dualization and  
568 generating hypergraph transversals. In *Proceedings of the thirty-fourth annual ACM symposium*  
569 *on Theory of computing*, pages 14–22, 2002.
- 570 18 P ERDdS and A R&wi. On random graphs i. *Publ. math. debrecen*, 6(290-297):18, 1959.
- 571 19 Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. cambridge University press,  
572 2009.
- 573 20 Michael L Fredman and Leonid Khachiyan. On the complexity of dualization of monotone  
574 disjunctive normal forms. *Journal of algorithms*, 21(3):618–628, 1996.
- 575 21 Hector Garcia-Molina and Daniel Barbara. How to assign votes in a distributed system.  
576 *Journal of the ACM (JACM)*, 32(4):841–860, 1985.
- 577 22 Daya Ram Gaur and Ramesh Krishnamurti. Average case self-duality of monotone boolean  
578 functions. In *Advances in Artificial Intelligence: 17th Conference of the Canadian Society for*  
579 *Computational Studies of Intelligence, Canadian AI 2004, London, Ontario, Canada, May*  
580 *17-19, 2004. Proceedings 17*, pages 322–338. Springer, 2004.
- 581 23 Dimitrios Gunopulos, Roni Khardon, Heikki Mannila, Sanjeev Saluja, Hannu Toivonen, and  
582 Ram Sewak Sharma. Discovering all most specific sentences. *ACM Transactions on Database*  
583 *Systems (TODS)*, 28(2):140–174, 2003.
- 584 24 Dimitrios Gunopulos, Heikki Mannila, Roni Khardon, and Hannu Toivonen. Data mining,  
585 hypergraph transversals, and machine learning. In *Proceedings of the sixteenth ACM SIGACT-*  
586 *SIGMOD-SIGART symposium on Principles of database systems*, pages 209–216, 1997.
- 587 25 Matthias Hagen. Lower bounds for three algorithms for transversal hypergraph generation.  
588 *Discrete Applied Mathematics*, 157(7):1460–1469, 2009.

- 589 26 Utz-Uwe Haus, Steffen Klamt, and Tamon Stephen. Computing knock-out strategies in  
590 metabolic networks. *Journal of Computational Biology*, 15(3):259–268, 2008.
- 591 27 Céline Hébert, Alain Bretto, and Bruno Crémilleux. A data mining formalization to improve  
592 hypergraph minimal transversal computation. *Fundamenta Informaticae*, 80(4):415–433, 2007.
- 593 28 David S Johnson, Mihalis Yannakakis, and Christos H Papadimitriou. On generating all  
594 maximal independent sets. *Information Processing Letters*, 27(3):119–123, 1988.
- 595 29 Mamadou Moustapha Kanté, Vincent Limouzy, Arnaud Mary, and Lhouari Nourine. On  
596 the enumeration of minimal dominating sets and related notions. *SIAM Journal on Discrete  
597 Mathematics*, 28(4):1916–1929, 2014.
- 598 30 Naouel Karam, Serge Linckels, and Christoph Meinel. Semantic composition of lecture subparts  
599 for a personalized e-learning. In *The Semantic Web: Research and Applications: 4th European  
600 Semantic Web Conference, ESWC 2007, Innsbruck, Austria, June 3-7, 2007. Proceedings 4*,  
601 pages 716–728. Springer, 2007.
- 602 31 Dimitris J Kavvadias and Elias C Stavropoulos. An efficient algorithm for the transversal  
603 hypergraph generation. *J. Graph Algorithms Appl.*, 9(2):239–264, 2005.
- 604 32 Dogan Kesdogan and Lexi Pimenidis. The hitting set attack on anonymity protocols. In  
605 *International Workshop on Information Hiding*, pages 326–339. Springer, 2004.
- 606 33 Leonid Khachiyan. Transversal hypergraphs and families of polyhedral cones. *Advances  
607 in Convex Analysis and Global Optimization: Honoring the Memory of C. Caratheodory  
608 (1873–1950)*, pages 105–118, 2001.
- 609 34 Leonid Khachiyan, Endre Boros, Khaled Elbassioni, and Vladimir Gurvich. An efficient  
610 implementation of a quasi-polynomial algorithm for generating hypergraph transversals and  
611 its application in joint generation. *Discrete Applied Mathematics*, 154(16):2350–2372, 2006.
- 612 35 Leonid Khachiyan, Endre Boros, Khaled Elbassioni, and Vladimir Gurvich. Generating all  
613 minimal integral solutions to and–or systems of monotone inequalities: Conjunctions are  
614 simpler than disjunctions. *Discrete applied mathematics*, 156(11):2020–2034, 2008.
- 615 36 Steffen Klamt, Julio Saez-Rodriguez, and Ernst D Gilles. Structural and functional analysis of  
616 cellular networks with cellnetanalyzer. *BMC systems biology*, 1(1):1–13, 2007.
- 617 37 Heikki Mannila and Kari-Jouko Rähkä. On the complexity of inferring functional dependencies.  
618 *Discrete Applied Mathematics*, 40(2):237–243, 1992.
- 619 38 Ronald C Read. Every one a winner or how to avoid isomorphism search when cataloguing  
620 combinatorial configurations. In *Annals of discrete mathematics*, volume 2, pages 107–120.  
621 Elsevier, 1978.
- 622 39 Saswati Sarkar and Kumar N Sivarajan. Hypergraph models for cellular mobile communication  
623 systems. *IEEE Transactions on Vehicular Technology*, 47(2):460–471, 1998.
- 624 40 Ken Takata. A worst-case analysis of the sequential method to list the minimal hitting sets of  
625 a hypergraph. *SIAM Journal on Discrete Mathematics*, 21(4):936–946, 2008.