On the number of partial Steiner systems

Armen S. Asratian and N. N. Kuzjurin

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-143286

N.B.: When citing this work, cite the original publication.

Original publication available at:

Copyright: Wiley (12 months)
http://eu.wiley.com/WileyCDA/
On The Number of Partial Steiner Systems

A. S. Asratian\textsuperscript{1}, N. N. Kuzjurin\textsuperscript{2}
\textsuperscript{1}Department of Mathematics, Lulea University, S-971 87 Lulea, Sweden,
\textsuperscript{2}Institute for System Programming, Russian Academy of Sciences, B. Kommunisticheskaya, 25, 109004 Moscow, Russia

Received May 25, 1999; accepted November 19, 1999

Abstract: We give a simple proof of the result of Grable on the asymptotics of the number of partial Steiner systems \( S(t,k,m) \).

Keywords: partial Steiner system; matching; hypergraph

1. INTRODUCTION

A partial Steiner system \( S(t,k,m) \) is a collection of \( k \)-subsets of an \( m \)-element set \( M \) such that each \( t \)-subset is contained in at most one \( k \)-subset from \( S(t,k,m) \). When every \( t \)-subset of \( M \) is contained in exactly one \( k \)-subset from \( S(t,k,m) \), we have a classical Steiner system on the set \( M \) with parameters \( t \) and \( k \). Some bounds of the number of such systems for \( t \approx 2 \), \( k \approx 3 \) and \( t \approx 3 \), \( k \approx 4 \) were obtained in [1], [9], [7] and [6]. Very little is known about the number of classical Steiner systems for large \( t \) and \( k \).

The number of distinct partial Steiner systems \( S(t,k,m) \) we denote by \( s(t,k,m) \). For two sequences \( f_m \) and \( g_m \) we write \( f_m \sim g_m \) if \( f_m / g_m \rightarrow 1 \) as \( m \rightarrow \infty \).

In [5] Grable announced that using the Rödl nibble algorithm [8] and generalizing the result in [3] he proved the following:

\textbf{Theorem 1.} Let \( t \) and \( k \) be two fixed positive integers, \( t < k \). Then

\[ \ln s(t,k,m) \sim \frac{k-t}{(k)_t} m'^t \ln m \text{ as } m \rightarrow \infty, \]

where \( (k)_t = k(k-1) \ldots (k-t+1) \).
The Rödl nibble algorithm is a very powerful but not an easy technique. In this paper we give a simple proof of Theorem 1. We use the result of Frankl and Rödl [4] concerning the existence of nearly perfect matchings in hypergraphs. A hypergraph \( H \) is a pair \((V, E)\), where \( V \) is a finite set of vertices and \( E \) is a finite family of subsets of \( V \), called edges. A hypergraph is \( r \)-uniform if every edge contains precisely \( r \) vertices. The number of edges of a hypergraph \( H \) containing a vertex \( v \) is called the degree of \( v \) and denoted by \( d_H(v) \) or simply \( d(v) \). For two distinct vertices \( u \) and \( v \) of a hypergraph \( H \), the number of edges containing both \( u \) and \( v \) is denoted by \( d_H(u,v) \) or simply \( d(u,v) \). A matching in a hypergraph is a collection of pairwise disjoint edges. We will use the result in [4] in the following slightly weaker form.

**Theorem 2.** Let integer \( r \geq 3 \) and real \( \delta > 0 \) be fixed, and \( H \) be an \( r \)-uniform hypergraph on \( n \) vertices. There exists \( \gamma > 0 \) and \( n_0 > 0 \) such that if for some \( D \) and for every pair of distinct vertices \( u \) and \( v \) of \( H \) the following two conditions hold:

1. \((1 - \gamma)D \leq d(v) \leq (1 + \gamma)D,
2. \(d(u,v) \leq D/(\ln n)^3,\)

then for all \( n \geq n_0 \) \( H \) has a matching containing at least \( \lceil (1 - \delta) \frac{n}{r} \rceil \) edges.

**2. PROOF OF THEOREM 1**

Let \( r = \binom{k}{t} \), \( n = \binom{m}{t} \), \( d = \binom{m-t}{k-t} \), and \( q = \binom{m}{k} \). Consider the family \( F = \{A_1, \ldots, A_q\} \) of all \( k \)-subsets of an \( m \)-set \( M \). We define the hypergraph \( H(t,k,m) \) corresponding to the family \( F \) in the following way: The vertex set \( V \) of \( H(t,k,m) \) is the set of all \( t \)-subsets of \( M \) and the edge set is \( E = \{e_1, \ldots, e_q\} \), where each \( e_i \) consists of all \( t \)-subsets which are contained in a \( k \)-subset \( A_i \). For each matching \( \{e_{i_1}, \ldots, e_{i_r}\} \) of \( H(t,k,m) \) the corresponding set \( \{A_{i_1}, \ldots, A_{i_r}\} \) in \( F \) is a partial Steiner system \( S(t,k,m) \). Since this correspondence is one-to-one, the number \( s(t,k,m) \) of all partial Steiner systems is equal to the number of matchings in the hypergraph \( H(t,k,m) \).

Note, that each edge of the hypergraph \( H(t,k,m) \) contains exactly \( r = \binom{k}{t} \) vertices, that is, the hypergraph is \( r \)-uniform.

Let \( \delta \) be real, \( 0 < \delta < 1/2 \). We define a random subfamily \( F(p) \) of the family \( F \) by choosing independently each \( k \)-subset of \( F \) with probability \( p = d^{-1+\delta} \). Taking into account the one-to-one correspondence between \( k \)-subsets of the \( m \)-set \( M \) and the edges of the hypergraph \( H(t,k,m) \) we obtain a random subhypergraph \( H_p = H_p(t,k,m) \) corresponding to the random family \( F(p) \).

**Lemma 1.** Let \( X \) be the random variable equal to the number of partial Steiner systems \( S(t,k,m) \) in the random family \( F(p) \) each containing at least \( T = \lceil (1 - \delta) \frac{m}{r} \rceil \) \( k \)-subsets. Then, \( \Pr \{X \geq 1\} \geq 1 - \delta \) for sufficiently large \( m \).

The proof of Lemma 1 will be given in Section 3. It uses the observation that since the vertex degrees and pair degrees of random subhypergraph \( H_p \) are sums of independent indicator variables, the Chernoff bounds prove that \( H_p \) almost always satisfy the conditions of Theorem 2 and, therefore, contains a matching with at least \( T = \lceil (1 - \delta) \frac{m}{r} \rceil \) edges.
Let $N$ be the number of partial Steiner systems $S(t, k, m)$ each containing at least $T$ $k$-subsets. Let us compare upper and lower bounds of the probability $P\{X \geq 1\}$.

**A lower bound.** By Lemma 1

$$P\{X \geq 1\} \geq 1 - \delta.$$ 

Markov’s inequality shows that

$$P\{X \geq 1\} \leq \mathbb{E}X \leq Np^T.$$ 

Thus,

$$N \geq (1 - \delta)p^{-T},$$

which implies the inequality

$$s(t, k, m) \geq N \geq (1 - \delta)d^{1 - \delta}n/r.$$  \hspace{1cm} (1)

**An upper bound.** The number of matchings $s(t, k, m)$ satisfies the following trivial upper bound:

$$s(t, k, m) \leq \sum_{j=1}^{\lceil \frac{n}{r} \rceil} \binom{q}{j} \leq \frac{n}{r} \left( \frac{q}{\lceil \frac{n}{r} \rceil} \right).$$

Taking into account that $\binom{q}{j} \leq \left( \frac{eq}{j} \right)^{j}$, and $q = \frac{nd}{r}$ we obtain

$$s(t, k, m) \leq \frac{n}{r} \left( \frac{nd}{\lceil r \rceil} \right) \leq \frac{n}{r} (ed)^{n/r} \leq d^{1 + \delta}n/r.$$ \hspace{1cm} (2)

Combining the inequalities (1) and (2) which hold for any fixed $0 < \delta < \frac{1}{2}$ and sufficiently large $m$, we obtain that

$$\ln s(t, k, m) \sim \frac{n}{r} \ln d \text{ as } m \to \infty.$$ 

Taking into account that

$$\frac{n}{r} = \frac{(m)_t}{(k)_t}, \quad (m)_t \sim m^t, \quad \ln d = \ln \left( \frac{m - t}{k - t} \right) \sim (k - t) \ln m,$$

we obtain that

$$\ln s(t, k, m) \sim \frac{k - t}{(k)_t} m^t \ln m \text{ as } m \to \infty.$$
3. PROOF OF LEMMA 1

Let $D(u)$ be the random variable equal to the degree of vertex $u$ in $H_p(t, k, m)$, and for each pair of distinct vertices $u$ and $v$ let $D(u, v)$ be the random variable equal to the number of edges in $H_p(t, k, m)$ containing both $u$ and $v$. For all vertices $u, v \in V$ let $A(u)$ denote the event: $|D(u) - ED(u)| > (\ln d)^{-1} ED(u)$ and $B(u, v)$ denote the event: $D(u, v) - ED(u, v) > (\ln n)^{-5} d^6$. We need the following technical lemma.

**Lemma 2.** For sufficiently large $m$

$$ P\{A(u)\} \leq \exp \{-d^{b/2}\}, \quad P\{B(u, v)\} \leq \exp \{-d^{b/2}\}. $$

**Proof of Lemma 2.** We use the following bounds for probabilities of large deviations of sums $Z = \sum_{i=1}^{L} z_i$ of independent random variables $z_1, \ldots, z_L$ such that $z_i$ takes two values 0 and 1, and $P\{z_i = 1\} = p$, $P\{z_i = 0\} = 1 - p$ (see [2]):

$$ P\{|Z - EZ| > \varepsilon EZ\} \leq 2 \exp \{-((\varepsilon^2/3)EZ\}, \quad \text{if } 0 \leq \varepsilon \leq 1, \quad (3) $$

$$ P\{Z - EZ > \varepsilon EZ\} \leq \exp \{((\varepsilon - (1 + \varepsilon)\ln(1 + \varepsilon))EZ\}, \quad \text{if } \varepsilon \geq 0. \quad (4) $$

For each vertex $u$ we define independent random variables $y_1, \ldots, y_{d(u)}$ such that $y_j = 1$ iff the $j$th edge containing $u$ is in $H_p(t, k, m)$, and $y_j = 0$, otherwise. It is clear, that

$$ D(u) = \sum_{j=1}^{d(u)} y_j, \quad ED(u) = \sum_{j=1}^{d(u)} Ey_j = pd = d^b. $$

Using (3) we obtain

$$ P\{|D(u) - ED(u)| > (\ln d)^{-1} ED(u)\} \leq 2 \exp \{-2 \ln d^{-2} d^6\} \leq \exp \{-d^{b/2}\}. $$

Similarly, for every two distinct vertices $u$ and $v$ we define independent random variables $x_1, \ldots, x_{d(u,v)}$ such that $x_j = 1$ iff the $j$th edge containing both $u$ and $v$ is in $H_p(t, k, m)$, and $x_j = 0$ otherwise. Clearly,

$$ D(u, v) = \sum_{j=1}^{d(u,v)} x_j. $$

Since $d = \binom{m-t}{k-t} \leq (m-t)^{k-t}$, we obtain, for sufficiently large $m$,

$$ ED(u, v) \leq p \max_{u \neq v} d(u, v) \leq p \binom{m-t-1}{k-t-1} \leq (k-t)d^{b-1/(k-t)} \leq d^b (\ln n)^{-6}. \quad (5) $$
Let $d^\delta = gE D(u, v)$, $\theta = (\ln n)^{-5}$. Then (5) implies $\theta g > \ln n$. This and (4) imply

$$P\{D(u, v) - E D(u, v) > \theta d^\delta\} = P\{D(u, v) - E D(u, v) > \theta g E D(u, v)\}$$

$$\leq \exp \{((\theta g - (1 + \theta g) \ln (1 + \theta g)) E D(u, v)\}$$

$$\leq \exp \{-\theta g (\ln (\theta g) - 1) E D(u, v)\} \leq \exp \{-d^\delta (\ln n)^{-5}\} \leq \exp \{-d^\delta/2\}.$$ 

The proof of Lemma 2 is complete.

Lemma 2 and inequality (4) imply, for sufficiently large $m$,

$$P\{\bigcup_{u \in V} A(u)\} \leq n \max_{u} P\{A(u)\} \leq m^\delta \exp \{-m^{\delta/4}\} \leq \delta/2,$$

$$P\{\bigcup_{u, v \in V, u \neq v} B(u, v)\} \leq n^2 \max_{u \neq v} P\{B(u, v)\} \leq m^2 \exp \{-m^{\delta/4}\} \leq \delta/2.$$ 

Note, that the event $\bigcup_{u \in V} A(u)$ implies, for every vertex $u$, the inequality

$$|D(u) - d^\delta| \leq (\ln d)^{-1} d^\delta,$$  \hspace{1cm} (6)

and the event $\bigcup_{u, v \in V, u \neq v} B(u, v)$ implies for every two distinct vertices $u, v$ the inequality

$$D(u, v) \leq E D(u, v) + (\ln n)^{-5} d^\delta \leq 2 d^\delta (\ln n)^{-5}.$$  \hspace{1cm} (7)

Let $\gamma$ and $n_0$ be the constants in Theorem 2. It is clear that for sufficiently large $m$ the inequality (6) implies the condition 1 of Theorem 2 with $D = d^\delta$, and the inequality (7) implies the condition 2 of Theorem 2 with $D = d^\delta$. Thus, by Theorem 2

$$P\{X \geq 1\} \geq 1 - (P\{\bigcup_{u \in V} A(u)\} + P\{\bigcup_{u, v \in V, u \neq v} B(u, v)\}) \geq 1 - \delta.$$ 

ACKNOWLEDGMENTS

We thank Svenska Institute for the financial support. We also thank the referees for their valuable remarks.

REFERENCES


