ssdnm
środowiskowe
studia doktoranckie
z nauk matematycznych

Katarzyna Mieczkowska

Uniwersytet A. Mickiewicza w Poznaniu

On Erdős' problem on fractional matchings in hypergraphs

Praca semestralna nr 3

(semestr letni 2011/12)

Opiekun pracy: Piotr Sankowski

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

ON ERDŐS PROBLEM ON FRACTIONAL MATCHINGS IN HYPERGRAPHS

KATARZYNA MIECZKOWSKA

Abstract. In 1965 Erdős asked what is the maximum number of edges in a k-uniform hypergraph whose matching number is exactly s. He conjectured that the extremal graphs are either cliques, or consist of all edges intersecting a set of s vertices. In the paper we present actual state of knowledge on this long-standing open problem. Most of all we focus on the fractional relaxation of Erdős problem and discuss its connections and applications to other mathematical problems. In particular we study how it can be reduced to an old probabilistic conjecture of Samuels on sums of independent random variables. We also discuss an application to the problem of finding an optimal data allocation in the distributed storage system.

1. Erdős extremal problem

A *k-uniform hypergraph* $G = (V, E)$ is a set of vertices $V \subseteq \mathbb{N}$ together with a family E of k-element subsets of V, which are called edges. In this note by $v(G) = |V|$ and $e(G) = |E|$ we denote the number of vertices and edges of $G = (V, E)$, respectively. By a *matching* we mean any family of disjoint edges of G, and we denote by $\mu(G)$ the size of the largest matching contained in E.

In 1965 Erdős asked what is the maximum number of edges in a k-uniform hypergraph whose matching number is exactly s. He conjectured that it is maximized either for cliques, or for graphs which consist of all edges intersecting a set of s vertices. Neither construction is uniformly better than the other in the whole range of parameter s ($1 \leq s \leq n/k$), so the conjectured bound is the maximum of these two possibilities. More formally, let $\mathcal{H}_k(n, s)$ denote the set of all k-graphs $G = (V, E)$ such that $|V| = n$ and $\mu(G) = s$; moreover let

$$\mu_k(n, s) = \max\{e(G) : G \in \mathcal{H}_k(n, s)\}, \tag{1}$$

and

$$\mathcal{M}_k(n, s) = \{G \in \mathcal{H}_k(n, s) : e(G) = \mu_k(n, s)\}. \tag{2}$$

Let us describe two kinds of k-graphs from $\mathcal{H}_k(n, s)$ which are natural candidates for members of $\mathcal{M}_k(n, s)$. By $\mathrm{Cov}_k(n, s)$ we denote the family of k-graphs $G_1 = (V_1, E_1)$ such that $|V_1| = n$ and for some subset $S \subseteq V_1$, $|S| = s$, we have

$$E_1 = \{e \subseteq V_1 : \ e \cap S \neq \emptyset \text{ and } |e| = k\}.$$

Clearly, if $s \leq n/k$, then $\mathrm{Cov}_k(n, s) \subseteq \mathcal{H}_k(n, s)$. Furthermore, we define $\mathrm{Cl}_k(n, s)$ as the family of all k-graphs $G_2 = (V_2, E_2)$ which consists of a complete subgraph on $ks + k - 1$ and some isolated vertices, i.e. if for some subset $T \subseteq V_2$, $|T| = ks + k - 1$, we have

$$E_2 = \{e \subseteq T : \ \ |e| = k\}.$$

Again, we have $\mathrm{Cl}_k(n, s) \subseteq \mathcal{H}_k(n, s)$.

In 1965 Erdős [5] conjectured that, indeed, unless $n = 2k$ and $s = 1$, the function $\mu_k(n, s)$ is fully determined by k-graphs of these two types.

Erdős Conjecture. *For every natural number k, n and s, where $ks \leq n - k + 1$, the following holds*

$$\mu_k(n, s) = \max\left\{\binom{n}{k} - \binom{n-s}{k}, \binom{ks+k-1}{k}\right\}.$$

Moreover,

$$\mathcal{M}_k(n, s) = Cov_k(n, s) \cup Cl_k(n, s).$$

Although this problem has been intensively studied for the last fifty years, it still remains widely open and only some partial results have been obtained so far. In 1959, few years before the conjecture was stated in the whole generality, Erdős and Gallai proved it in a graph case ($k = 2$). For 3-uniform hypergraphs the conjecture has been verified just this year in [9] and [15]. Łuczak and Mieczkowska settled it in the afirmative for n large enough [15], having also shown that the only extremal 3-uniform hypergraphs are of the conjectured form. Recently, Frankl [9] proved Erdős Conjecture for 3-uniform hypergraphs for every n. He also announced to confirm the conjecture for 4-uniform hypergraphs, for $s > s_0$. The conjecture is also known to be true for the special case $s = 1$, which is actually equivalent to the celebrated Erdős-Ko-Rado theorem, which was first proved in 1961 [7].

For general k there have been series of results dealing mostly with the case when n is large compared to s, proving that

$$\mathcal{M}_k(n, s) = \mathrm{Cov}_k(n, s) \quad \text{for} \quad n \geq g(k)s, \tag{3}$$

where $g(k)$ is some function of k. The existence of such $g(k)$ was shown by Erdős [5], then Bollobás, Daykin and Erdős [3] proved that (3) holds whenever $g(k) \geq 2k^3$; Frankl and Füredi [10] showed that (3) is true for $g(k) \geq 100k^2$ and recently, Huang, Loh, and Sudakov [13] verified (3) for $g(k) \geq 3k^2$. The best published bound for $g(k)$ for general k is due to Frankl, Łuczak and Mieczkowska [11] who showed that (3) holds whenever $g(k) \geq 2k^2/\log k$.

Some close connections of Erdős Conjecture to several important problems in extremal graph theory have been recently revealed. In most of the cases it suffices to consider the weaker version of Erdős problem, its fractional relaxation, which we discuss in the next section.

2. Fractional matchings and Erdős Conjecture

A fractional matching in a k-uniform hypergraph $G = (V, E)$ is a function

$$w : E \to [0, 1] \text{ such that}$$

$$\sum_{e \ni v} w(e) \leq 1 \text{ for every vertex } v \in V.$$

Then $\sum_{e \in E} w(e)$ is the size of a matching w and the size of the largest fractional matching in G is denoted by $\nu^*(G)$. Note that $\nu^*(G)$ is always greater or equal than $\nu(G)$.

Erdős Problem in its weaker statement asks to bound the number of edges in a k-uniform hypergraph G as a function of the fractional matching number $\nu^*(G)$ instead of $\nu(G)$. As in the integral case, Erdős stated a conjecture suggesting the solution of this problem.

Erdős Conjecture (fractional version). *Let $k \geq 2$, n be any integers and let s be a real such that $0 \leq s \leq n/k$. Then, for every k-uniform hypergraph G on n vertices with fractional matching number $\nu^*(H) < s$ the following holds*

$$e(H) \leq \max\left\{ \binom{n}{k} - \binom{n - \lceil s \rceil + 1}{k}, \binom{\lceil ks \rceil - 1}{k} \right\}.$$

The lower bound is yielded either by the covering graph $\mathrm{Cov}_k(n, \lceil s \rceil - 1)$ or the clique graph $\mathrm{Cl}_k(n, \lceil ks \rceil - 1)$. In its asymptotic form the fractional conjecture says the following.

Conjecture 1. *Every k-uniform hypergraph H on n vertices with fractional matching number $\nu^*(H) = xn$, where $0 \leq x < 1/k$ satisfies*

$$e(H) \leq (1 + o(1)) \max\left\{1 - (1 - x)^k, (kx)^k\right\} \binom{n}{k}.$$

As a consequence of the Erdős -Gallai theorem from [6] and the latest results of Frankl, Łuczak and Mieczkowska [9], [15], fractional version of Erdős Conjecture is asymptotically true for $k = 2$ and $k = 3$ for n going to infinity. In [1] it has been also confirmed for $k = 4$ for $x \leq 1/5$. The maximum is then achieved by the first term.

The determination of $\nu^*(G)$ is a linear programming problem. Its dual problem is to find the minimum fractional vertex cover $\tau^*(G)$. The *fractional vertex cover* in a k-uniform hypergraph $G = (V, E)$ is a function

$$w : V \to [0,1] \text{ such that}$$
$$\text{for each } e \in E \text{ we have } \sum_{v \in e} w(v) \geq 1.$$

Then $\sum_{v\in V} w(v)$ is the size of a cover w and the size of the smallest fractional cover in G is denoted by $\tau^*(G)$. Let us remind that in integral setting *a vertex cover* of G is a set of vertices $S \subset V$ such that each edge of G has got at least one vertex in S. Let $\tau(G)$ be the minimum number of vertices in a vertex cover of G. Note that $\tau^*(G)$ is always smaller or equal than $\tau(G)$. Then, by Duality Theorem, for every k-uniform hypergraph G we have

$$\nu(G) \leq \nu^*(G) = \tau^*(G) \leq \tau(G).$$

Therefore, to solve Erdős problem for fractional matchings we can consider its dual problem. Then, by the duality, getting an upper bound for the number of edges in a k-uniform hypergraph G is equivalent to finding a function $v : V(G) \to [0,1]$ satisfying $\sum_{v\in V} w(v) < s$ that maximizes the number of k-tuples e where $\sum_{v\in e} w(v) \geq 1$.

It was observed (see [1]) that the fractional version of Erdős Conjecture is closely related to an old probabilistic conjecture of Samuels. This conjecture, if true, would imply the fractional version of Erdős problem for the range $x \leq 1/(k+1)$. In the next section we will shortly discuss this dependency.

3. Fractional matchings and Samuels Conjecture

Let $\mu_1, \ldots, \mu_k$ be real numbers satisfying $0 \leq \mu_1 \leq \ldots \leq \mu_k$ and $\sum_{i=1}^{k} \mu_i < 1$. Define

$$Q_t(\mu_1, \ldots, \mu_k) = \prod_{i=t+1}^{k} \left(1 - \frac{\mu_i}{1 - \sum_{j=1}^{t} \mu_j}\right)$$

for each $1 \leq t < k$. In 1966 Samuels [20] raised the following conjecture.

Samuels Conjecture. *For all admissible values of $\mu_1, \ldots, \mu_k$ and for all possible collections of k independent nonnegative random variables $X_1, \ldots, X_k$ with expectations $\mu_1, \ldots, \mu_k$, respectively,*

$$\mathbf{P}(X_1 + \ldots + X_k < 1) \geq \min_{t=0,\ldots,k-1} Q_t(\mu_1, \ldots, \mu_k).$$

Note that for $k = 1$ Samuels Conjecture is true because of Markov's inequality. The conjecture was proved also for $k \leq 4$ by Samuels in [20], [21]. For all $k \geq 5$ this problem is still widely open.

Observe also that $Q_t(\mu_1, \ldots, \mu_k)$ is equal exactly the value of $\mathbf{P}(X_1 + \ldots + X_k < 1)$ when X_j is identically μ_j for all $j \leq t$, and X_i attains values 0 and $1 - \sum_{j=1}^{t} \mu_j$ for all $i \geq t+1$.

To solve Erdős fractional matching problem it is enough to deal with a very special case of Samuels Conjecture. Indeed, it suffices to confirm the conjecture for X_i with $\mu_i = \mu$ for some $0 < \mu \leq 1/(k+1)$ and all $i = 1, \ldots, k$. In such case, the minimum in Samuels Conjecture is attained by $Q_0(\mu_1, \ldots, \mu_k)$ (see [1]).

Proposition 1. *For every integer $k \geq 2$ and every real number x saisfying $0 < x \leq 1/(k+1)$, if $\mu_1 = \ldots, \mu_k = x$, then*

$$\min_{t=0,\ldots,k-1} Q_t(\mu_1, \ldots, \mu_k) = Q_0(\mu_1, \ldots, \mu_k) = (1-x)^k.$$

Thus, we may focus on this very special case of Samuels Conjecture, stated below.

Conjecture 2. *For all $\mu \leq \frac{1}{k+1}$ and all choices of k independent random variables $X_1, \ldots, X_k$ with a common expectation μ*

$$\mathbb{P}(X_1 + \ldots X_k < 1) \geq (1-\mu)^k.$$

Let now see how this conjecture, if true, implies the fractional version of Erdős problem. We follow the proof of Theorem 2.1. from [1].

Theorem 1. *For any $k \geq 3$ and $0 < x \leq \frac{1}{k+1}$, if Conjecture 2 is true for k and $\mu = x$ then Conjecture 1 holds for $x \leq \frac{1}{k+1}$.*

Proof. Let $H = (V, E)$ be a k-uniform hypergraph on n vertices such that $\nu^*(H) < xn$ for some $0 < x \leq \frac{1}{k+1}$. By duality we have $\tau^*(H) = \nu^*(H) < xn$. Thus, there exists a fractional vertex cover $w : V \to [0,1]$ such that $\sum_{v \in V} w(v) < xn$ and for every $e \in E$ we have $\sum_{v \in e} w(v) \geq 1$. Without loss of generality we may assume that $\sum_{v \in V} w(v) = xn$.

Let now choose at random k vertices $v_1, \ldots, v_k$ of H, independently and uniformly from the set $V(H)$. Let X_i be a random variable such that $X_i = w(v_i)$ for every $i = 1, \ldots, k$. Observe that X_i are independent and each of them attains value $w(v)$ with probability $1/n$ for every vertex $v \in V(H)$, so they are identically distributed. Let now compute the expected value of X_i.

$$\mu_i = \mathbb{E}(X_i) = \sum_{v \in V} w(v) \frac{1}{n} = \frac{\sum_{v \in V} w(v)}{n} = \frac{xn}{n} = x.$$

Let denote by E' the set of all k-tuples contained in V that are not an edge in H. Since $|E'| = \binom{n}{k} - |E|$ we can bound the number of edges in H, by estimating the size of the set E'. As for every $e \in E$ we have $\sum_{v\in e} w(v) \geq 1$, then number N of all $k-$element subsets $S \subset V$ for which $\sum_{v\in S} w(v) < 1$ is a lower bound on the number $|E'|$ of non-edges in H. Let N_1 and N_2 be the numbers of all k-element sequences of vertices of V and all k-element sequences of distinct vertices of V, respectively, with the sums of weights smaller than 1. As the number of all k-element subsets of V is $\binom{n}{k} = (1 + o(1))n^k/k!$ and $N = N_2/k!$, we have

$$\mathbb{P}\left(\sum_{i=1}^{k} w(v_i) < 1\right) = \frac{N_1}{n^k} \leq \frac{N_2 + O(n^{k-1})}{\binom{n}{k}k!} = (1 + o(1))\frac{N}{\binom{n}{k}}.$$

Assuming, Conjecture 2 holds for a given k, then by Proposition 1,

$$\mathbb{P}\left(\sum_{i=1}^{k} w(v_i) < 1\right) = \mathbb{P}\left(\sum_{i=1}^{k} X_i < 1\right) \geq (1 - x)^k.$$

Thus,

$$|E'| \geq N \geq (1 + o(1))(1 - x)^k \binom{n}{k},$$

and, consequently

$$e(H) = \binom{n}{k} - |E'| \leq (1 + o(1)) \left(1 - (1 - x)^k\right) \binom{n}{k}.$$

□

Applying this probabilistic approach, by combining Samuels results for $k = 3, 4$ with Theorem 1, Alon et al. [1] proved Conjecture 1 for $k = 3$, $x < 1/4$ and for $k = 4$, $x < 1/5$. Using Samuels Conjecture in the higher range of x, one gets a bound on the number of edges which is larger than that in Conjecture 1.

4. Applications to other problems

Some generalized results, obtained while working on Erdős problem, revealed close connections of Erdős Conjecture to several important problems in extremal graph theory. For example, it is known that it can be used to study a Dirac-type question of Daykin and Häggkvist [4] about perfect matchings in hypergraphs. Moreover, in [2] it turned out that its fractional version might be used to attack an old problem of Manickam-Miklos-Singhi.

Conjecture 3 (Manickam, Mikloś, Singhi, 1987). *For any integers n, k satisfying $n \geq 4k$, every set of n real numbers with nonnegative sum has at least $\binom{n-1}{k-1}$ k-element subsets whose sum is also nonnegative.*

Furthermore, in [1] it was discovered that the fractional version of Erdős Conjecture have some interesting applications in information theory. In particular, studies on the uniform model of distributed storage allocation performed in [22] led to a question which is asymptotically equivalent to the fractional version of Erdős problem. This model has been studied in information theory in [14], [19], [22] and we will describe it below.

Consider a distributed storage system comprising n storage nodes. A file is to be split into multiple chunks, replicated redundantly and stored in a distributed manner over these nodes. The size of the whole file is normalized to unit size and the amount of data to be stored in each node i is equal to x_i. In reality, because there is limited storage space or transmission bandwidth, we require that the total amount of data stored does not exceed a given budget T, i.e. $x_1 + \ldots + x_n \leq T$. At some time after the creation of this coded storage, a data collector attempts to recover the original data object by accessing only the data stored in a random subset R of r nodes which is chosen uniformly at random. It is known that there always exists a coding scheme such that we can recover the file whenever the total amount of data accessed is at least 1. Our goal is to find an optimal allocation $(x_1, \ldots, x_n)$ to maximize the probability of successful recovery. This problem can be reformulated as follows.

Problem 1. *For a nonnegative sequence $(x_1, \ldots, x_n)$, let*

$$\Phi(x_1, \ldots, x_n) = \left| \{S \subset [n], |S| = r \text{ such that } \sum_{i \in S} x_i \geq 1\} \right|.$$

Then the probability of successful recovery of the file equals

$$\frac{\Phi(x_1, \ldots, x_n)}{\binom{n}{r}}.$$

Given integers $n \geq r \geq 1$ and a real number $T > 0$, determine

$$F^T(r, n) = max_{\sum x_i = T, x_i \geq 0 \forall_i} \Phi(x_1, \ldots, x_n)$$

and find an allocation optimizin $F^T(r, n)$.

It can be shown that Problem 1 may be translated to a hypergraph settings in such a way that $F^T(r, n)$ is the maximum number of edges in an r-uniform hypergraph on n vertices with fractional matching number at most T. The only difference in Erdős problem is that we assume the strict inequality $\nu^*(H) < T$ in its definition. Hence, Problem 1 is only asymptotically equivalent to the Erdős Problem, and thus, if Conjecture 1 is true, then

$$F^T(r,n) \approx \max\left\{\binom{rT}{r}, \binom{n}{r} - \binom{n-T}{r}\right\}.$$

The bounds are achieved when H is a clique or a complement of clique. A corresponding (asymptotically) optimal storage allocation is $x_1 = \ldots = x_{rT} = 1/r$, $x_{rT+1} = \ldots = x_n = 0$ or $x_1 = \ldots = x_T = 1$, $x_{T+1} = \ldots = x_n = 0$, respectively.

References

[1] N. Alon, P. Frankl, H. Huang, V. Rödl, A. Ruciski, B. Sudakov, *Large matchings in uniform hypergraphs and the conjectures of Erdős and Samuels*, J. Combin. Th., Series A **119** (2012), 1200–1215.

[2] N. Alon, H. Huang, and P. Sudakov, *Nonnegative k-sums, fractional covers, and probability of small deviations*, J. Combin. Th. Ser. B **102** (2012), 784–796.

[3] B. Bollobás, E. Daykin, and P. Erdős, *Sets of independent edges of a hypergraph*, Quart. J. Math. Oxford Ser. (2) **27** (1976), 25–32.

[4] D. E. Daykin, R. Häggkvist, *Degrees giving independent edges in a hypergraph*, Bull. Austral. Math. Soc. **23** (1981), 103–109.

[5] P. Erdős, *a problem on independent r-tuples*, Ann. Univ. Sci. Budapest. Etvs Sect. Math. **8** (1965), 93–95.

[6] P. Erdős and T. Gallai, *On maximal paths and circuits of graphs*, Acta Math. Acad. Sci. Hungar. **10** (1959), 337–356.

[7] P. Erdős, C. Ko, and R. Rado, *Intersection theorems for systems of finite sets*, Quart. J. Math. Oxford Set. 2 **12** (1961), 313–320.

[8] P. Frankl, *The shifting technique in extremal set theory*, in: Surveys in Combinatorics, Lond. Math. Soc. Lect. Note Ser. **123** (1987), 81–110.

[9] P. Frankl, *On the maximum number of edges in a hypergraph with given matching number*, `arXiv:1205.6847`.

[10] P. Frankl and Z. Füredi, unpublished.

[11] P. Frankl, T. Łuczak, and K. Mieczkowska, *On matchings in hypergraphs*, Elec. J. Combin. **19** 2 (2012), P42.

[12] P. Frankl, V. Rödl, and A. Ruciński, *On the maximum number of edges in a triple system not containing a disjoint family of a given size*, Combinatorics, Probability & Computing **21** (2012), 141–148.

[13] H. Huang, P. Loh, and B. Sudakov, *The size of a hypergraph and its matching number*, Combinatorics, Probability & Computing **21** (2012), 442–450.

[14] D. Leong, A. G. Dimakis, T. Ho, *Symmetric allocations for distributed storage*, Proceedings of the IEEE Global Telecommunications Conference, (GLOBECOM), (2010).

[15] T. Łuczak, K. Mieczkowska, *On Erdős' extremal problem on matchings in hypergraphs*, `arXiv:1202.4196`.

[16] T. Łuczak, K. Mieczkowska, *Erdős conjecture on matchings in hypergraphs*, http://ssdnm.mimuw.edu.pl/.
[17] N. Manickam, D. Miklós, *On the number of non-negative partial sums of a non-negative sum*, Colloq. Math. Soc. Janos Bolyai, **52** (1987), 385–392.
[18] N. Manickam, D. Miklós, *First distribution invariants and EKR theorems*, J. Combinatorial Theory, Series A **48** (1988), 91–103.
[19] M. Naor, R. M. Roth, *Optimal file sharing in distributed networks*, **24** (1991), 158–183.
[20] S. M. Samuels, *On a Chebyshev-type inequality for sums of independent random variables*, Ann. Math. Statist., **37** (1966), 248–259.
[21] S. M. Samuels, *More on a Chebyshev-type inequality for sums of independent random variables*, Purdue Stat. Dept. Mimeo., Series no. **155** (1968).
[22] M. Sardari, R. Restrepo, F. Fekri, E. Soljanin, *Memory allocation in distributed storage networks*, Proceedings of IEEE International Symposium on Information Theory (ISIT), (2010).

Adam Mickiewicz University, Collegium Mathematicum, ul. Umultowska 87, 61-614 Poznań, Poland