Sparse Hypergraphs: from Theory to Applications

Gennian Ge

Capital Normal University

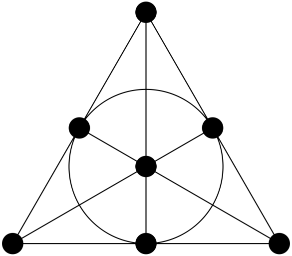
August, 2019

Outline

- Part I: Sparse hypergraphs
- Part II: Perfect hash families
- Part III: Centralized coded caching schemes

Hypergraphs

- Hypergraph $\mathcal{H} = (V, E)$: vertex set V, edge set $E \subseteq 2^V$ (*power set of* V)
- r-uniform hypergraph (henceforth r-graph): $E \subseteq$ {*r-subsets of V*}
- e.g. multi-relation
- e.g. Fano plane, Fano (1892), $r = 3, |V| = 7, |E| = 7$



(6,3)-free 3-graphs

- (6,3)-configuration: $\exists$ 3 (distinct) edges A, B, C, s.t. $|A \cup B \cup C| \leq 6$
- e.g. a typical (6,3)-configuration:

$$\begin{pmatrix} A & B & C \\ a & a & * \\ * & b & b \\ c & * & c \end{pmatrix}$$

- A 3-graph is called (6,3)-free: $\forall\ A, B, C \in E$, $|A \cup B \cup C| \geq 7$
- Question (Brown, Erdős and Sós, 1973): what is the maximum number of (3-)edges can be contained in a (6,3)-free 3-graph on n vertices?

(6,3)-free 3-graphs

- (partial) Answer (Ruzsa and Szemerédi, 1978):

$$n^{2-o(1)} < f_3(n,6,3) < o(n^2)$$

as $n \to \infty$. (also known as the (6,3)-theorem)

- State-of-the-art bound, $\exists$ constants $a, b > 0$:

$$e^{-a\sqrt{n}} n^2 < f_3(n,6,3) < \epsilon n^2,$$

where ϵ satisfies $\log\Big(\log\big(\cdots\log(n)\big)\Big) := \log^* n < 1$, for $b\log(\epsilon^{-1})$ iterations of $\log(\cdot)$, (Fox, Ann. of Math., 2011)

(6,3)-free 3-graphs

- Tools: regularity lemma (Szemerédi, 1976),
  graph removal lemma (Ruzsa and Szemerédi, 1978; Fox, 2011),
  sets of integers with no 3-term arithmetic progression (Behrend, 1946),
- Influence: Extremal Graph Theory, Ramsey Theory, Additive Combinatorics, Theoretical Computer Science, etc
- Notable mathematicians:
  - Noga Alon (ACM Fellow, AMS Fellow, Israel Prize)
  - Paul Erdős (Wolf Prize)
  - Timothy Gowers (Fields Medal)
  - Terence Tao (Fields Medal)
  - Endre Szemerédi (Abel Prize)

Sparse hypergraphs

- $\mathcal{H}$: an r-graph on n vertices
  $\mathcal{G}_r(v,e)$: all r-graphs with e edges and at most v vertices
- $\mathcal{H}$ is $\mathcal{G}_r(v,e)$-free if it does not contain any member of $\mathcal{G}_r(v,e)$.
  i.e., for arbitrary distinct $A_1,\dots,A_e \in \mathcal{H}$, $|A_1 \cup \cdots \cup A_e| \geq v+1$
- $f_r(n,v,e)$: the maximal number of edges of a $\mathcal{G}_r(v,e)$-free r-graph on n vertices. i.e., if $\mathcal{H}$ contains $f_r(n,v,e)+1$ edges, $\mathcal{H}$ contains at least one member of $\mathcal{G}_r(v,e)$
- Objective: the behavior of $f_r(n,v,e)$ with r,v,e fixed as $n \to \infty$

Known results

- Brown, Erdős and Sós (1973) proved

$$f_r(n, e(r-k)+k, e) = \Theta(n^k)$$

- Conjecture: for $r \geq k+1 \geq 3$, $e \geq 3$,

$$n^{k-o(1)} < f_r(n, e(r-k)+k+1, e) = o(n^k).$$

  (upper bound due to BES, lower bound due to Alon and Shapira, 2006)

- e.g. $r = 3, e = 3, k = 2$: $n^{2-o(1)} < f_3(n, 6, 3) = o(n^2)$

- Known matching parameters are rare. The first unsolved case: $f_3(n, 7, 4)$?

- Ruzsa and Szemerédi's (1976): $n^{2-o(1)} < f_3(n,6,3) = o(n^2)$
- Erdős, Frankl and Rödl (1986): $r \geq 3$, $k = 2$, $e = 3$,

$$n^{2-o(1)} < f_r(n, 3(r-2)+2+1, 3) = o(n^2)$$

- Alon and Shapira (2006): $r > k \geq 2$, $e = 3$,

$$n^{k-o(1)} < f_r(n, 3(r-k)+k+1, 3) = o(n^k)$$

- Sárközy and Selkow (2005): $r > k \geq 3$, $e = 4$,

$$f_r(n, 4(r-k)+k+1, 4) = o(n^k)$$

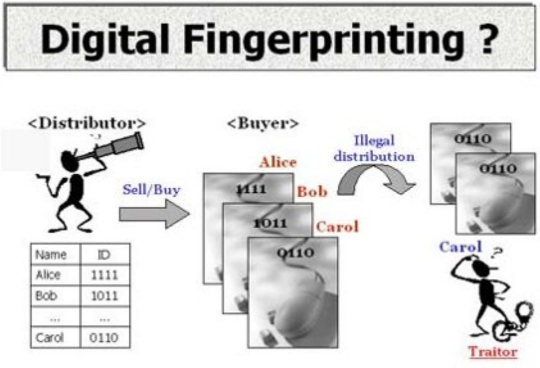
- Nagle, Rödl and Schacht (2006): $r > k \geq 2$, $e = k+1$

$$f_r(n, (k+1)(r-k)+k+1, k+1) = o(n^k)$$

- Common point in the upper bound: $r \geq k+1 \geq e$. A unified proof?

Our results

- Conjecture: $n^{k-o(1)} < f_r(n, e(r-k)+k+1, e) = o(n^k)$.
- The upper bound part holds for all $r \geq k+1 \geq e$, implying all previously known tight upper bounds.
- The lower bound part holds for $r \geq 3,\ k=2,\ e=4,5,7,8$. (first general constructions matching the lower bound for $e \geq 4$ since 1973)
- An improved lower bound for $r=3,\ k=2,\ e=6$
- Main tools: hypergraph removal lemma, additive combinatorics



- Columns of the matrix: digital fingerprints
- Insert digital fingerprints into digital data
- Distribute digital data to legal customers
- Bob sells his copy of data to other people, he will be caught by testing the digital fingerprint

Applications: perfect hash families

- Let M be an $N \times m$ matrix over a q-ary alphabet.
- M is t-perfect hashing if for arbitrary t columns $c_1, \ldots, c_t$ of M there is a row f, so that $f(c_1), \ldots, f(c_t)$ are all distinct.
- $\begin{pmatrix} & c_1 & c_2 & \cdots & \cdots & c_t \\ row\ f & f(c_1) & f(c_2) & \cdots & \cdots & f(c_t) \end{pmatrix}$
- Remark: columns of $M \Leftrightarrow$ members of a t-perfect hash family.
- Let $p_t(N, q)$ denote the maximum number of columns in such a matrix.

Known results and methods

- Here we are interested in the behavior of $p_t(N, q)$ with fixed t, N as $q \to \infty$. (other case: fixed t, q as $N \to \infty$)
- $\Omega(q^{\frac{N}{t-1}}) < p_t(N, q) \leq (t-1)q^{\lceil \frac{N}{t-1} \rceil}$
- For large q, the exponent is tight when $(t-1)|N$
- Major open problem: is the exponent tight when $\nmid$ happens?
- Conjecture (Walker II and Colbourn, 2007):

$$p_3(3, q) = o(q^2).$$

- Remark: very similar to the behavior of $f_r(n, e(r-k)+k, e)$ and $f_r(n, e(r-k)+k+1, e)$ mentioned earlier

- Known results
  - Combinatorial counting: $p_3(3, q) = \mathcal{O}(q^2)$.
  - Probabilistic method: $p_3(3, q) \geq \Omega(q^{3/2})$.
  - Construction from finite geometry (Fuji-Hara, 2015): $p_3(3, q) \geq \Omega(q^{5/3})$.
- Previous methods: probabilistic method, combinatorial design theory, algebraic combinatorics, finite geometry, etc.
- These traditional methods seem hopeless to obtain a close lower/upper bound concerning the $o(1)$ term.
- Our point of view: sparse hypergraphs and additive combinatorics.

Crucial observation

A linear family defined by a $3 \times m$ q-ary matrix is 3-perfect hashing if and only if it does not contain the following configuration.

	c_1	c_2	c_3
R_1	a	a	
R_2		b	b
R_3	c		c

- rows $\Rightarrow$ vertex parts of a 3-partite 3-graph; columns $\Rightarrow$ 3-edges
- The above configuration is indeed a (6,3)-configuration, $p_3(3, q) = \Theta(f_3(q, 6, 3))$

Theorem (Shangguan and Ge, SIAM DM 2016)

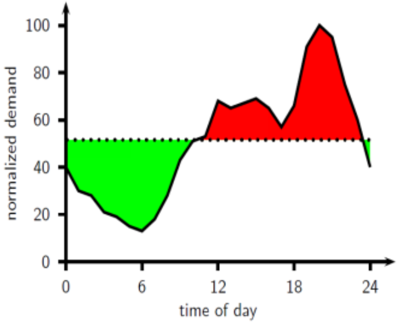
For sufficiently large q, $q^{2-o(1)} < C(3, q, \{1, 1, 1\}) = p_3(3, q) = o(q^2)$.

Applications: centralized coded caching schemes

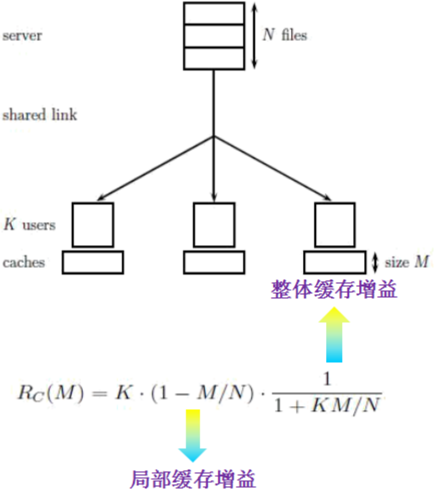
数字产品的普及带来了无线网络的拥挤，尤其以视频点播服务为例，在高峰期面对大量的需求，网络堵塞，传输效率低下，进而严重影响用户体验。

应对方案：
利用网络系统中及用户客户端中的缓存空间，以一定编码技巧在网络使用低峰期预先存储部分内容，从而缓解高峰期的传输压力。

- Network burden in peak-traffic times and off-peak times

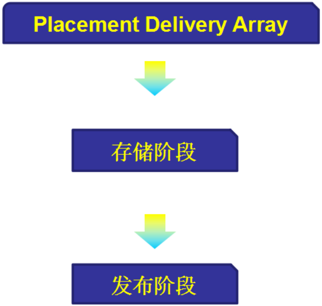


- Fundamental Limits of Caching, Maddah-Ali and Niesen, IT 2014 (SCI Citation: 502, IEEE Information Theory Society Best Paper Award (2016))



Example

文件：**A=(A1,A2,A3,A4)，B=(B1,B2,B3,B4)**



$$\mathcal{P}_{4,2} = \begin{pmatrix} & \text{用户1} & \text{用户2} \\ \text{文件分割1} & * & 1 \\ \text{文件分割2} & 1 & * \\ \text{文件分割3} & * & 2 \\ \text{文件分割4} & 2 & * \end{pmatrix}$$

用户1缓存：**(A1,A3,B1,B3)**
用户2缓存：**(A2,A4,B2,B4)**

用户需求	第一次传输	第二次传输
(A, A)	$A2 \bigoplus A1$	$A4 \bigoplus A3$
(A, B)	$A2 \bigoplus B1$	$A4 \bigoplus B3$
(B, A)	$B2 \bigoplus A1$	$B4 \bigoplus A3$
(B, B)	$B2 \bigoplus B1$	$B4 \bigoplus B3$

- In general, $F = F(K) = \exp(K)$, try to reduce it!

- Placement delivery array (Yan, Cheng, Tang and Chen, IT 2017)
  - PDA: An array of size $F \times K$, $\mathcal{P} = [p_{j,k}]_{F\times K}$, F is a given integer such that $Z := FM/N$ is an integer.
  - $\mathcal{P}$ consists of a specific symbol $*$ and a set of S integers.
  - The transmission rate is $R = S/F$.
  - Set $\mathcal{F} = \{1, ..., F\}$, $\mathcal{K} = \{1, ..., K\}$, $\mathcal{S} = \{1, ..., S\}$, $\mathcal{N} = \{1, ..., N\}$.

- The following constraints are required:

  C1. $*$ appears $Z = FM/N$ times in each column. Each column has $F - Z$ integer entries.

  C2. In each row or each column there do not exist identical integers.

  C3. For any two distinct entries $p_{j_1,k_1} = p_{j_2,k_2} = s \in \mathcal{S}$, $j_1 \neq j_2$ and $k_1 \neq k_2$, we have $p_{j_1,k_2} = p_{j_2,k_1} = *$.

- A hypergraph perspective (Shangguan, Zhang and Ge, IT 2018)

(K, F, Z, S)-PDA		超图
F×K的矩阵，字母集为S	→	顶点集为F×K×S的三部三均衡超图
每列有Z个*	→	K中每个点的度数为(F-Z)
每行每列不存在相同整数	→	线性性质
对角线性质	→	(6, 3)-free性质

Our results

- PDA $\Leftrightarrow$ a linear 3-uniform 3-partite (6,3)-free hypergraph exists.
- (6,3)-theorem $\Rightarrow$ constant rate PDA with $F(K)$ linear in K does not exist.
- Two new constructions: constant rate PDAs with $F(K) = \exp(\sqrt{K})$. (previous constructions: $F(K) = \exp(K)$)
- Open question: does $F(K) = \mathrm{poly}(K)$ exist?

References

- Theory

  - Gennian Ge and Chong Shangguan, Sparse hypergraphs: New bounds and constructions, submitted to J. Combin. Theory (B), 2017 (arXiv:1706.03306).

- Applications

  - Chong Shangguan and Gennian Ge, Separating Hash Families: A Johnson-type bound and new constructions. SIAM J. Discrete Math., 30(4):2243–2264, 2016.
  - Chong Shangguan, Yiwei Zhang and Gennian Ge, Centralized coded caching schemes: A hypergraph theoretical approach, IEEE Trans. Inform. Theory, 64(8):5755–5766, 2018.

$\mathscr{THANK}$

$\mathscr{YOU}$