Exercise XI, Sublinear Algorithms for Big Data Analysis 2024-2025

These exercises are for your own benefit. Feel free to collaborate and share your answers with other students, and solve as many problems as you can. Problems marked (*) are more difficult, but also more rewarding. These problems have been taken from various sources on the Internet, too numerous to cite individually.

1 Let $G = (P, Q, E)$ be defined as follows. Let $P = \{0,1\}^n, Q = \{0,1\}^n, N = 2^n$, and let $E = \bigcup_{i=1}^n M_i$, where M_i is a matching between $A_i := \{x \in \{0,1\}^n : x_i = 1\}$ and $B_i := \{x \in \{0,1\}^n : x_i = 0\}$ defined as follows. For every $x \in A_i$ insert an edge $(x, x + e_i)$ into M_i, where e_i is the i-th coordinate vector and addition is modulo 2. Show that the matchings M_i are induced, and hence G is a $(1/2, \log_2 N, 2N)$-Ruzsa-Szemerédi graph.

Solution. Suppose towards a contradiction that the matchings are not induced, i.e. there exist a, b such that $(a, a+e_i) \in M_i, (b, b+e_i) \in M_i$ and $(b, a+e_i) \in M_j$. Then $b+e_j = a+e_i$. However, since $a, b \in M_i$, we get that $a_i = b_i = 1$, and since $a + e_j \in M_j$, we have $(a + e_j)_i = a_i = 0$, a contradiction. □

2 Extend the construction above to obtain $(1/2^k, \binom{\log_2 N}{k}, 2N)$-Ruzsa-Szemerédi graphs for every integer $k \geq 1$.

Solution. For every subset S of $[\log_2 N]$ of size k match $A_S = \{x \in \{0,1\}^n : x_S = 0^{|S|}\}$ to $B_S = \{x \in \{0,1\}^n : x_S = 1^{|S|}\}$ by matching $x \in A_S$ to $x + \sum_{i \in S} e_i \in B_S$. It follows similarly to exercise 1 that the matchings are induced. The size bounds and the bounds on the number of matchings follows. $\square$

3 In this problem you will construct $(1/12, 2^{\Omega((\log\log N)^2, N)})$-Ruzsa-Szemerédi graphs. We will use

Theorem 0.1 (Grolmusz'2000) *Let m have $r > 1$ distinct prime divisors. Then for $h > 0$, there exists a uniform set system $\mathcal{H}$ over $[h]$ with:*

1. $|\mathcal{H}| \geq \exp\left(c\frac{(\log h)^r}{(\log\log h)^{r-1}}\right)$
2. $|H| \equiv 0 \mod m$ *for all* $H \in \mathcal{H}$
3. $|G \cap H| \not\equiv 0 \mod m$ *for* $G \neq H$

This is very suprising since such a system does not exist if m is prime:

Theorem 0.2 (Frankl–Wilson) *Let $\mathcal{F}$ be a set system over a universe of n elements, and let $\mu_0, \mu_1, \ldots, \mu_s$ be distinct residues modulo a prime p, such that:*

- *For all $F \in \mathcal{F}$, $|F| = k \equiv \mu_0 \pmod{p}$,*
- *For any two distinct $F, G \in \mathcal{F}$, $|F \cap G| \equiv \mu_i \pmod{p}$ for some $i \in \{1, \ldots, s\}$.*

If $k + s \leq n$, then

$$|\mathcal{F}| \leq \binom{n}{s}.$$

Let $v_1, \ldots, v_m \in \{0,1\}^n$, with $|v_i| \equiv 0 \mod 6$, and $\langle v_i, v_j \rangle \not\equiv 0 \mod 6$ for $i \neq j$, as per the Grolmusz construction. We define a graph G by letting $V = \mathbb{Z}_6^n$ and defining edges as follows. For every $i = 1, \ldots, m$ define $A_i = \{u \in V : \langle u, v_i \rangle \equiv 0 \mod 6\}$. For each $u \in A_i$, include an edge $\{u, u + v_i\}$. Prove that this defines a $(1/6, 2^{\Omega((\log\log N)^2}, N)$–Ruzsa-Szemerédi graph.

Solution. For every $x \in \mathbb{Z}_6$ consider the line through x in direction v_i, i.e. the set of points $x + t \cdot v_i, t \in \{0,1,2,3,4,5\}$. Such lines are disjoint for all x such that $\langle x, v_i \rangle \pmod 6 = 0$. Include edges $\{x, x+v_i\}, \{x+2v_i, x+3v_i\}$ and $\{x+4v_i, x+5v_i\}$, where the arithmetic is mod 5, for every x such that $\langle x, v_i \rangle \pmod 6 = 0$. These are matchings of size $V/12$, since $1/6$ fraction of points satisfy $\langle x, v_i \rangle \pmod 6 = 0$, and the matching is perfect on all of them. It remains to argue the induced property.

Suppose that $\{a, a+v_j\}$ is an edge of a matching M_j, where a and $a+v_j$ are matched in matching M_i. The latter means that $\langle a, v_i \rangle \pmod 6 = \langle a+v_j, v_i \rangle \pmod 6 = 0$. Thus

$$\langle v_j, v_i \rangle \pmod 6 = 0,$$

a contradiction with the choice of the family of v_i's.

There are 6^m vertices, and the size of the set of v_i's is $2^{\Omega(\log^2 m)}$. Thus, we get $2^{\Omega((\log\log N)^2)}$ matchings of size $N/12$, as required. Here $r = 2$ since 6 has two distinct prime numbers. $\square$