

MINIMÁLIS RELÁCIÓS ADATBÁZISOK

FÜREDI ZOLTÁN

Budapest

Jelölje $\mathcal{F}_k^n$ az n -elemű halmaz összes k -elemű részhalmazának rendszerét. Legyen $\mathbf{M}$ egy olyan mátrix, amely $\mathcal{F}_k^n$ -t indukálja minimális kulcsrendszerként. DEMETROVICS és KATONA [7] belátták, hogy ekkor az $\mathbf{M}$ sorainak minimális száma $s(\mathcal{F}_k^n) > \sqrt{2 \binom{n}{k-1}}$ és, hogy ez a becslés pontos a $k=1$ és 2 esetben. Ismeretes, hogy $s(\mathcal{F}_3^n) = n + O(1)$ (l. [9]).

E dolgozat célja, hogy alkalmas $\mathbf{M}$ mátrix megadásával megmutassa, hogy általában is rögzített k mellett

$$s(\mathcal{F}_k^n) = O(n^{(k-1)/2}).$$

1. Alapfogalmak

Az itt következő relációs adatbázis modellt CODD [4], ARMSTRONG [1] és DEMETROVICS [5] dolgozták ki. Egy összefoglaló dolgozatot l. pl. DEMETROVICS—KATONA [7] vagy (magyar nyelven) BÉKÉSSY—DEMETROVICS—GYEPESI [3]. Itt csak a legszükségesebb definíciókra szorítkozunk.

Jelöljön $\mathbf{M}$ egy $n \times m$ -es mátrixot (adatbázist). Az oszlopok halmazát X -szel, a sorokét V -vel fogjuk jelölni. ($|X|=n$, $|V|=m$). Egy $\kappa \subset X$ részhalmazt *kulcs*-nak nevezünk, ha a κ oszlopai által alkotott részmátrix minden sora különböző. A κ kulcs *minimális*, ha nem tartalmaz további kulcsot ($\nexists \kappa' \subset \kappa, \kappa' \neq \kappa, \kappa' \text{ kulcs}$). Az $\mathbf{M}$ által indukált minimális kulcsok rendszerét $\mathcal{K}(\mathbf{M})$ -mel, vagy röviden $\mathcal{K}$ -val jelöljük. Nyilván $\mathcal{K}$ -ban nem létezik két egymást tartalmazó halmaz, más néven $\mathcal{K}$ egy *Sperner-rendszer*.

1.1. ÁLLÍTÁS (l. pl. [2]). Tetszőleges $\mathcal{F}$ *Sperner-rendszer*hez létezik olyan mátrix, amely indukálja. (Azaz $\exists \mathbf{M}$, melyre $\mathcal{F} = \mathcal{K}(\mathbf{M})$.)

1.2. KÖVETKEZMÉNY. n oszlop esetén a minimális kulcsok maximális száma éppen $\left\lfloor \begin{matrix} n \\ 2 \end{matrix} \right\rfloor$.

2. Minimális indukáló adatbázisok

Legyen $\mathcal{F}$ tetszőleges *Sperter-rendszer* n ponton. Az 1.1. állítás szerint létezik olyan $\mathbf{M}$ mátrix, amely indukálja. Természetesen felvetődő kérdés, hogy mennyi ennek minimális sorszáma. Legyen

$$s(\mathcal{F}) = \min \{m: \mathcal{F} = \mathcal{K}(\mathbf{M}), \mathbf{M} n \times m\text{-es mátrix}\}.$$

Vagyis s azt a minimális sorszámot jelöli, ahány sorú mátrixszal $\mathcal{F}$ indukálható.

Jelölje $\Delta\mathcal{F} X$ azon maximális részhalmozait, melyek nem kulcsok. Azaz $\Delta\mathcal{F} = \{G \subset X : G \text{ nem kulcs, de } \forall G \subseteq G' \subset X \text{ esetén } G' \text{ kulcs}\}$. Az $\mathcal{F}$ és $\Delta\mathcal{F}$ halmazrendszerek egyértelműen meghatározzák egymást.

$$2.1. \text{ ÁLLÍTÁS (l. pl. [2], } \binom{s(\mathcal{F})}{2} \equiv |\Delta\mathcal{F}| \equiv s(\mathcal{F}) - 1.$$

A *Sperner-tétel* és a 2.1. állítás következménye a következő állítás.

2.2. ÁLLÍTÁS. Ha $\mathcal{F}$ egy *Sperner-rendszer* n ponton, akkor

$$s(\mathcal{F}) \equiv \left\lfloor \frac{n}{2} \right\rfloor + 1.$$

A következő eredmény, melyet bizonyítás nélkül közlünk, alsó becslést ad a lehetséges maximális $s(\mathcal{F})$ értékre (DEMETROVICS—GYEPESI [6]):

2.3. TÉTEL. Létezik olyan $\mathcal{F}$ *Sperner-rendszer* n ponton, melyre

$$\frac{1}{n^2} \left\lfloor \frac{n}{2} \right\rfloor \equiv s(\mathcal{F}).$$

3. Uniform rendszereket indukáló adatbázisok

A 2.3. tétel bizonyítása (l. még [3] és [7]) nem konstruktív. A konkrét konstrukciók keresése vezethet az uniform rendszerek vizsgálatához. Jelölje $\mathcal{F}_k^n$ az n -elemű alaphalmaz összes k -elemű részhalmozából álló halmazrendszert. Nyilván $\Delta\mathcal{F}_k^n = \mathcal{F}_{k-1}^n$, így a 2.1. állításból kapjuk, hogy

$$3.1. \text{ KÖVETKEZMÉNY. } \sqrt{2 \binom{n}{k-1}} < s(\mathcal{F}_k^n) \equiv \binom{n}{k-1} + 1.$$

DEMETROVICS és KATONA [7] meghatározták $s(\mathcal{F}_k^n)$ értékét, ha $k=1, 2, n-1, n$.

$$3.2. \text{ TÉTEL. } s(\mathcal{F}_1^n) = 2, \quad s(\mathcal{F}_2^n) = \lfloor (1 + \sqrt{8n+1})/2 \rfloor \quad \text{és}$$

$$s(\mathcal{F}_n^n) = n, \quad s(\mathcal{F}_{n-1}^n) = n+1.$$

Továbbá bebizonyították, hogy

$$(3.1) \quad n \equiv s(\mathcal{F}_3^n) \equiv \mathcal{O}(n^{3/2}).$$

A felső becslés a nyilvánvaló

$$(3.2) \quad s(\mathcal{F}_k^n) \equiv s(\mathcal{F}_k^{n-1}) + s(\mathcal{F}_{k-1}^{n-1})$$

összefüggésből adódik.

Legutóbb feloldható blokkrendszerek segítségével sikerült bizonyítani sejtésüket, miszerint $s(\mathcal{F}_3^n) = n + \mathcal{O}(1)$.

$$3.3. \text{ TÉTEL (l. [8]). } n \equiv s(\mathcal{F}_3^n) \equiv n + 8.$$

4. Eredmények

A fent említett eredmények közül azokra koncentrálunk, melyekben a kulcsok nagysága rögzített. A 3.2. és a 3.3. tétel azt mutatja, hogy a 3.1. következmény alsó becslése adja $s(\mathcal{F}_k^n)$ pontos nagyságrendjét (legalábbis a $k=1, 2$ és 3 esetekben). E dolgozat célja, hogy alkalmas M mátrix megadásával bizonyítsa, hogy ez minden k -ra így van, azaz

4.1. TÉTEL. Ha $n > k \geq 1$ egészek, akkor $s(\mathcal{F}_k^n) < n^{(k-1)/2} \cdot 2^k$.

Felmerülhet 2.3 tétellel analóg kérdés a k -uniform kulcsrendszerek esetén is.

4.2. TÉTEL. (1. [8]). Létezik olyan $\mathcal{H} \subset \mathcal{F}_k^n$ melyre $s(\mathcal{H}) > \sqrt{2} \binom{2k-2}{k-1}^{1/(2k-2)1/2}$.

Jelölje $f_k(n) = \max \{s(\mathcal{F}) : \mathcal{F} \subset \mathcal{F}_k^n\}$. Könnyű ellenőrizni, hogy $f_1(n) = 2$. A fenti tétel adja, hogy $f_k(n) > 2^{n/4}$. Kissé talán meglepő, hogy egy egyszerű konstrukció is csak ilyen sok sorral imprikálható. Ezzel kapcsolatos a következő probléma:

4.3. PROBLÉMA. Határozzuk meg $\max \{|\Delta \mathcal{F}| : \mathcal{F} \subset \mathcal{F}_k^n\}$ -et. Más szóval: Egy n pontú k -gráfnak hány maximális független pontrendszere lehet? A $k=2$ eset megoldott: diszjunkt háromszögek rendszere $3^{\lfloor n/3 \rfloor}$ -ot ad.

A $\max s(\mathcal{H})$ meghatározása, a minimális sorszámra alsó becslést nyújtó technikák híján, nagyon nehéznek látszik.

5. A konstrukció előkészítése, véges testek és differenciahalmazok

A 4.1. tétel bizonyításához szükség lesz néhány eddig nem szerepelt fogalomra.

Jelöljön Q egy véges q -elemű testet. (Pl. ha q prím, akkor $Q = \langle 0, 1, \dots, q-1 \rangle$ a modulo q -vett szorzással és összeadással.) Az $A \subset Q$ részhalmazt *differenciahalmaznak* mondjuk, ha két A -beli elem különbségeként minden testelem megkapható. Azaz, ha az $x = a_i - a_j$ egyenlet megoldható minden $x \in Q$ esetén ($a_i, a_j \in A$).

5.1. LEMMA. A q -elemű véges test tartalmaz egy legfeljebb $2\sqrt{q}$ tagú differenciahalmazt.

Bizonyítás. Legyen e a Q egységeleme és $a = \lceil \sqrt{q} \rceil$. Végül legyen $A = \{0, e, 2e, \dots, ae\} \cup \{2ae, 3ae, \dots, a^2e\}$. Ez az A nyilván differenciahalmaz, hiszen minden $(q-a)$ -nál kisebb x egész felírható $x = \alpha a - \beta$ alakban, ahol $0 \leq \beta < a$ és $\alpha a < q$. Végül a $q-a$ és $q-1$ között egészek is előállíthatóak (modulo q) $0 - \beta$ alakban. Ennek az A -nak a számosságára:

$$|A| = a + \left\lfloor \frac{q}{a} \right\rfloor = \lceil \sqrt{q} \rceil + \left\lfloor \frac{q}{\lceil \sqrt{q} \rceil} \right\rfloor \leq \lceil \sqrt{q} \rceil + \lfloor \sqrt{q} \rfloor \leq 2\sqrt{q}.$$

(Ugyanis minden $x \geq 0$ valósra $\lfloor x \rfloor + \lfloor x \rfloor \leq \lfloor 2x \rfloor$.)

Végezetül könnyen látható, hogy ha $\sqrt{q}$ nem egész szám, akkor az $A - \{0\}$ is differenciahalmaz. Ezzel a lemmát igazoltuk.

5.2. *Megjegyzés.* Véges test helyett a differenciahalmaz véges csoportra is értelmezhető. Mi itt tulajdonképpen azt bizonyítottuk, hogy a q -elemű véges ciklikus csoport minimális differenciahalmaza legfeljebb $2\sqrt[3]{q}$ elemű. Általános tételt (amely minden csoportra igaz) nem ismerek, de a *mohó algoritmussal* (vagy LOVÁSZ LÁSZLÓ [11] egy tételének következményeként) könnyen igazolható, hogy

(5.1) tetszőleges q -elemű véges csoport minimális differenciahalmaza legfeljebb $q/\log q$ elemű.

Azt hiszem az (5.1) általában nem javítható.

Az itt ismertetendő konstrukciónk csupán véges gyűrűket használ, így a fenti megjegyzésnek egy pontosabb felső becslés kidolgozásában lehet jelentősége.

5.3. *Megjegyzés.* Az 5.1. lemma egyszerűsége ellenére azért is jól használható, mert bizonyos értelemben elég közel van az igazsághoz. (Konstans szorzó erejéig.) Ugyanis

(5.2) Tetszőleges $A \subset Q$ differenciahalmazra $|A| \geq \sqrt[3]{q}$.

Valóban az összes $a_i - a_j$ párok száma éppen $|A|^2$ ($a_i, a_j \in A$), így ha ezek kiadják a Q összes elemét, akkor $|A|^2 \geq |Q|$. (Ráadásul e párok nem is mind különbözőek.)

5.4. *Megjegyzés.* A számelméletben és egyes algebrai struktúrákban sokat használt a *perfekt* differenciahalmaz fogalma. Ez olyan differenciahalmaz, mely az $x \neq 0$ -ra az $x \in Q$ elemet egyértelműen állítja elő $a_i - a_j$ alakban. Ilyen perfekt differenciahalmaz csak a $q = n^2 + n + 1$ esetén létezik (ahol n egész szám).

Bővebbet erről l. [10].

6. A 4.1. tétel bizonyítása, a konstrukció

Legyen $A \subset Q$ a q -elemű véges test egy minimális számosságú differenciahalmaza, mely tartalmazza a 0-t és az 1-et. Az 5.1. lemma szerint $|A| = a < 2\sqrt[3]{q}$. Definálunk egy $q \times 2a^{k-1}$ -es M mátrixot, melynek elemei testelemek lesznek és egy-egy sora egy-egy legfeljebb $(k-1)$ -edfokú Q feletti polinom, $0, 1, \dots, g-1$ helyen felvett helyettesítési értékeiből áll. Legyen a $\mathcal{P}$ polinomhalmaz a következő:

$$\mathcal{P} = \{c_{k-1}x^{k-1} + c_{k-2}x^{k-2} + \dots + c_1x + c_0 : c_{k-1} = 0 \text{ vagy } 1, \text{ és minden } c_i \in A\}.$$

Legyen M egy $Q \times \mathcal{P}$ -es mátrix, ahol a p -edik sor x -edik eleme $p(x)$ ($p \in \mathcal{P}$, $x, p(x) \in Q$). Belátjuk, hogy ez az M rendelkezik a kívánt tulajdonságokkal.

6.1. **LEMMA.** Az M mátrix p' és p'' -edik sorának közös elemeinek halmaza éppen a $(p' - p'')$ polinom zérushelyeinek halmaza.

A 6.1. lemma bizonyítása triviális, hiszen ahol $p'(x) = p''(x)$, ott $(p' - p'')(x) = 0$.

6.2. **KÖVETKEZMÉNY.** Az összes legalább k -elemű halmaz kulcs.

A 6.2. következmény bizonyítása. Itt használjuk, hogy a Q véges test és ezért a Q feletti legfeljebb $(k-1)$ -edfokú polinomoknak legfeljebb $(k-1)$ gyökük lehet.

6.3. LEMMA. Ha a Q feletti p polinom legfeljebb $(k-1)$ -edfokú és fő együtt-hatója 1, akkor felírható $p' - p''$ alakban, ahol $p', p'' \in \mathcal{P}$.

A 6.3. lemma bizonyítása a differenciahalmaz fogalmából adódik, hiszen ha $p(x) = c_{k-1}x^{k-1} + \dots + c_1x + c_0$, akkor az összes c_i -t szézbontva $a'_i - a''_i$ alakra ($a'_i, a''_i \in A$) a $p'(x) = \sum_{i \leq k-1} a'_i x^i$, $p''(x) = \sum_{i \leq k-1} a''_i x^i$ polinomok kielégítik a lemma feltételeit.

6.4. KÖVETKEZMÉNY. Az M által indukált kulcsok között nincs egyetlen k -nál kisebb elemű.

A 6.4. következmény bizonyítása. Legyen $T \subset Q$ $|T| \leq k-1$, $T = \langle t_1, \dots, t_i, \dots \rangle$. Tekintsük a $p_T = (x-t_1)(x-t_2)\dots(x-t_i)\dots$ legfeljebb $(k-1)$ -edfokú 1 főegyütt-hatójú polinomot. Ehhez található a 6.3. lemma szerint egy olyan p' és $p'' \in \mathcal{P}$, melyre $p' - p'' = p_T$. De ekkor a 6.1. lemma szerint a p' és p'' -edik sorok a T halmazban megegyeznek, így T nem lehet M által indukált kulcs.

Összefoglalva

6.5. TÉTEL. Ha $n > k \geq 1$, akkor a fent definiált M -re $\mathcal{H}(M) = \mathcal{F}_k^Q$.

A 4.1. tétel bizonyítása. Nyilvánvaló, hogy ha Q oszlopaiból elhagyunk néhányat, akkor (a megmaradt oszlopok halmazát Y -nal jelölve)

$$(6.1) \quad \mathcal{H}(M|Y) = \mathcal{F}_k^{Q|Y}.$$

Ebből a Csebisev-tételt használva (mely szerint minden n és $2n$ között létezik prím-szám) legyen $n \leq q \leq 2n$ adódik, hogy

$$s(\mathcal{F}_k^n) \leq s(\mathcal{F}_k^{Q/n}) \leq s(\mathcal{F}_k^Q) = 2a^{k-1} \leq 2(2q)^{\frac{k-1}{2}} < 2(4n)^{\frac{k-1}{2}} = 2^k n^{(k-1)/2}.$$

7. További bizonyítások

A 4.2. tétel bizonyítása szintén alkalmas konstrukció megadásával történik. Particionáljuk X -et, $X = X_1 \cup \dots \cup X_q \cup Y$, ahol $q = \lfloor n/(2k-2) \rfloor$ és $|X_i| = 2k-2$ ($1 \leq i \leq q$). Legyen $\mathcal{F} = \{F: |F| = k, F \subset X_i \text{ valamely } i\text{-re}\}$. Ekkor $\Delta\mathcal{F} = \{A: |A \cap X_i| = k-1 \text{ minden } i\text{-re, és } Y \subset A\}$. Következésképp $|\Delta\mathcal{F}| = \binom{2k-2}{k-2}^q$. Végül a 2.1 állítást használva kap-hatun alsó becslést $s(\mathcal{F})$ -re.

IRODALOM

- [1] ARMSTRONG, W. W., "Dependency structures of data base relationship", *Information Processing* 74 (North-Holland Publishing Company, Amsterdam, 1974) 580—583.
- [2] BÉKÉSSY, A., DEMETROVICS, J., HANNÁK, L., KATONA, G. O. H. and FRANKL, P., "On the number of maximal dependencies in a data relation of fixed order" *Discrete Mathematics* 30 (1980) 83—88.
- [3] BÉKÉSSY, A., DEMETROVICS, J. és GYEPESI, GY., „Relációs adatbázis logikai szintű vizsgálata funkcionális függőségek szempontjából”, *MTA SZTAKI Tanulmányok* 109 (1980) 1—56.

- [4] CODD, E. F., "A relational model of data for large shared data banks", *Communications of the ACM* **13** (1970) 377—387.
- [5] DEMETROVICS, J., "Candidate keys and antichains", *SIAM Journal on Algebra and Discrete Methods* **1** (1980) 1—92.
- [6] DEMETROVICS, J. and GYEPESI, GY., "On the functional dependency and some generalization of it", *Acta Cybernetica* **5** (1981) 296—305.
- [7] DEMETROVICS, J. and KATONA, G. O. H., "Extremal combinatorial problems in relational data base", *Lecture Notes in Computer Sciences* **117** (1981) 110—119.
- [8] DEMETROVICS, J., FÜREDI, Z. and KATONA, G. O. H., "Maximum matrix representation of closure operations", *Discrete Mathematics*, megjelenés alatt.
- [9] FÜREDI, Z., „Relációs adatbázisokkal kapcsolatos kombinatorikai problémák”, kézirat, Budapest, 1981.
- [10] HALBERSTAM, H. and ROTH, F. K., *Sequences* (Clarendon Press, Oxford, 1966).
- [11] LOVÁSZ, L., Minimax tételek a kombinatorikában, Doktori értekezés, Budapest, 1978.
- [12] SPERNER, E., „Ein Satz über Untermengen einer endlich Menge“, *Mathematische Zeitschrift* **27** (1928) 544—548.

(Beérkezett: 1982. december 15.)

FÜREDI ZOLTÁN

MTA MATEMATIKAI KUTATÓ INTÉZET
1053 BUDAPEST, RÉALTANODA U. 13—15.

MINIMAL RELATIONAL DATA BASES

Z. FÜREDI

Let F_k^n denote the system of all k -element subsets of an n -element set. Let $\mathbf{M}$ be a matrix which generates F_k^n as a minimal key system. DEMETROVICS and KATONA have proved ([7]) that in this case the minimal number of the rows of $\mathbf{M}$ is $s(F_k^n) \geq \sqrt{2 \binom{n}{k-1}}$ and this bound is exact for $k=1$ and 2. It is also well known that $s(F_3^n) = n + O(1)$ (see [8]).

The purpose of this paper is to show that generally for a fixed value of k

$$s(F_k^n) = O(n^{(k-1)/2}).$$

This is done by definition of an appropriate matrix $\mathbf{M}$.