SZAKDOLGOZAT

Kombinatorikus problémák az adatbázisok elméletében

1998

Írta: Tichler Krisztián
Témavezető: Katona Gyula

Tartalomjegyzék

1. Bevezetés 3

2. Alaptulajdonságok 5

3. Direkt szorzatok reprezentációja 11

4. $s(\mathcal{L}_k^n)$ nagyságrendje 16

5. $s(\mathcal{L}_k^n)$ meghatározása k néhány értékére 20

6. Egy kulcsrendszer reprezentációja 23

Hivatkozási jegyzék 33

1.Bevezetés

Egy adatbázis legegyszerűbb modellje egy mátrix. Egy-egy oszlopban azonos fajta adatok szerepelnek (például név, születés ideje, születés helye, stb.), míg a különböző sorokban a különböző egyének adatai. Egy bizonyos fajta adatokat (az irodalomban használatos elnevezéssel) attributumnak nevezünk, mely azonosítható a fenti mátrix egy oszlopával. Az attributumok halmazát $\Omega = \{a_1, a_2, \ldots, a_n\}$-nel fogjuk jelölni. A lehetséges értékek halmazát az i-edik oszlopban a_i értékkészletének nevezzük, és $D(a_i)$-vel jelöljük. Így tehát egy egyén adatait (a mátrix egy sorát) tekinthetjük a $D(a_1) \times D(a_2) \times \ldots \times D(a_n)$ direkt szorzat egy elemének (n-esének). Ezért az egész adatbázis (vagy mátrix) leírható az $r \subseteq D(a_1) \times D(a_2) \times \ldots \times D(a_n)$ relációval, azaz az n-esek halmazával. Ennek a modellnek a neve relációs adatbázis modell.

Az adatok között fennállhatnak bizonyos logikai összefüggések. Például a születés dátuma meghatározza a kort (egy adott évben), vagy a vezetéknév és a keresztnév meghatározza a monogramot. Azt mondjuk, hogy egy attributum (funkcionálisan) függ attributumok egy halmazától, ha az azokban felvett értékek együttesen egyértelműen meghatározzák az ottani értéket. A funkcionális függőség fogalmát W.W. Armstrong és E.F. Codd vezették be. Formálisan:

1.1 Definició [1],[5] *Legyen $A, B \subseteq \Omega$. Azt mondjuk, hogy B funkcionálisan függ A-tól, ha tetszőleges $b \in B$ esetén nem létezik r-nek 2 sora (n-ese), melyek A oszlopaiban (attributumaiban) megegyeznek, de b-ben különböznek. Ennek jelölése: $A \to B$.*

A funkcionális függőségek nagyon hasznosnak bizonyultak. Valamennyi adatbázist működtető rendszer ezen a koncepción alapul. Tekintsük a következő példát. Tegyük fel, hogy $\Omega = \{a_1, a_2, a_3, a_4\}$ és $a_1 \to a_2$ valamint $a_3 \to a_4$. Ha a teljes mátrixot tároljuk a számítógép memóriájában, akkor a legrosszabb esetben $4N_1N_3$ regiszterre van szükség, ahol $N_i = |D(a_i)|$ $(i = 1, 3)$. Valóban, a_1 és a_3 egymástól függetlenül vehetnek fel értékeket, de ezek meghatározzák a_2-t és a_4-et. Így a különböző sorok száma legfeljebb N_1N_3. Azonban az adott funkcionális függőségek használatával számottevő memóriát takaríthatunk meg. Valóban, elegendő az a_1 és a_3 oszlopaiból álló mátrixot ($2N_1N_3$ regiszter) valamint két kis mátrixot tárolni. Az egyik a_1 és a_2 értékeiből áll az első illetve második oszlopban. Az első oszlop tartalmazza a_1 lehetséges értékeit, míg a második az $a_1 \to a_2$ függőség által meghatározott értékeket. A másik kis mátrix a_3-ból és a_4-ből ugyanígy épül fel. A tárolt értékek száma legfeljebb $2N_1N_3 + 2(N_1 + N_3)$, ami általában szignifikánsan kisebb $4N_1N_3$-nál.

A szakdolgozat célja megvizsgálni a fukcionális függőség tulajdonságait. Ennek során számos érdekes kombinatorikai probléma merül fel. A 2. fejezet megismerteti az olvasót a témakör fogalmaival és problémáival. A 3. fejezet direkt szorzatok, a 4. és 5. fejezet a gyakran vizsgált $\mathcal{L}_k^n$ lezáráscsalád minimális reprezentációjával foglalkozik. A 6. fejezet új eredménye egy új típusú kulcsrendszercsalád minimális reprezentációját vizsgálja.

2. Alaptulajdonságok

Ebben a fejezetben a legfontosabb definiciók és néhány egyszerűbb, de alapvető tétel kerül tárgyalásra. Legyen $\Omega = \{a_1, a_2, \ldots, a_n\}$.

2.1 Definició [17] *Az Ω halmaz részhalmazaibol álló $A \hookrightarrow B$ párok egy $\mathcal{D}$ rendszerét determinációnak nevezzük, ha Ω bármely A, B, C, D részhalmazára*

$$A \hookrightarrow A, \tag{2.1}$$

$$A \hookrightarrow B,\ B \hookrightarrow C \Rightarrow A \hookrightarrow C, \tag{2.2}$$

$$A \subseteq C, D \subseteq B, A \hookrightarrow B \Rightarrow C \hookrightarrow D, \tag{2.3}$$

$$A \hookrightarrow B, C \hookrightarrow D \Rightarrow A \cup B \hookrightarrow C \cup D. \tag{2.4}$$

Könnyen látható, hogy egy r reláció (adatbázis) összes $A \to B$ funkcionális függőségei determinációt alkotnak Ω részhalmazain.

2.2 Definició [1] *Egy r reláció összes funkcionális függőségét teljes családnak nevezzük.*

2.3 Tétel [1] *Az $A \hookrightarrow B$ párok egy rendszere teljes család akkor és csak akkor ha determináció.* ■

2.4 Definició *Egy $\mathcal{L}$ halmazfüggvényt lezárási operációnak vagy röviden lezárásnak nevezünk Ω részhalmazain, ha minden $A, B \subseteq \Omega$-ra*

$$A \subseteq \mathcal{L}(A), \tag{2.5}$$

$$A \subseteq B \Rightarrow \mathcal{L}(A) \subseteq \mathcal{L}(B), \tag{2.6}$$

$$\mathcal{L}(\mathcal{L}(A)) = \mathcal{L}(A). \tag{2.7}$$

2.5 Állítás [17] *Egy adott Ω halmazon létezik bijekció a determinációk halmaza és a lezárások halmaza között.*

Bizonyítás: Ha adott egy $\mathcal{D}$ determináció Ω-n, akkor definiáljuk az $\mathcal{L}(\mathcal{D})$ halmazfüggvényt: $\mathcal{L}(A) = \{a \mid A \hookrightarrow a\}$ minden $A \subseteq \Omega$-ra. Könnyen látható hogy $\mathcal{L}(\mathcal{D})$ lezárás, valamint hogy $\mathcal{D} \longrightarrow \mathcal{L}(\mathcal{D})$ injekció.

Fordítva, ha adott egy $\mathcal{L}$ lezárás, akkor $\mathcal{D}(\mathcal{L})$ álljon azon $A \hookrightarrow B$ párokból, melyre $B \subseteq \mathcal{L}(A)$. Könnyen látható, hogy $\mathcal{D}(\mathcal{L})$ determináció, és most is $\mathcal{L} \longrightarrow \mathcal{D}(\mathcal{L})$ injekció. Mivel Ω véges halmaz, ezért mindkét injekció valójában bijekció. ■

További bijekciók is megadhatók a determinációk halmazával (például a metszet-félhálókkal vagy a metszetmentes halmazcsaládokkal) [4],[9],[12],[17].

2.6 Definició *$K \subseteq \Omega$ kulcs, ha $K \hookrightarrow \Omega$; minimális kulcs, ha minimális a tartalmazásra, mint részbenrendezésre nézve a kulcsok között. $A \subseteq \Omega$ antikulcs, ha nem kulcs; maximális antikulcs, ha maximális a tartalmazásra, mint részbenrendezésre nézve az antikulcsok között.*

Azaz egy r reláció esetén K kulcs, ha a K-beli értékek meghatároznak minden más értéket is.

Egy $\mathcal{L}$ lezárás (vagy $\mathcal{D}$ determináció, stb.) meghatározza a minimális kulcsok $\mathcal{K} = \mathcal{K}(\mathcal{L})$ (vagy $\mathcal{K} = \mathcal{K}(\mathcal{D})$) családját. Ez nyilván egy nem-üres Sperner-rendszer (egy $\mathcal{S}$ halmazrendszer Sperner ha tartalmazásmentes, vagyis ha $A, B \in \mathcal{S} \Rightarrow A \not\subseteq B$) Fordítva, ha adott egy $\mathcal{K}$ nem-üres Sperner-rendszer, akkor

$$\mathcal{L}(A) = \begin{cases} A & \text{ha nincs olyan } K \in \mathcal{K}, \text{ melyre } K \subseteq A \\ \Omega & \text{ha van olyan } K \in \mathcal{K}, \text{ melyre } K \subseteq A \end{cases} \tag{2.8}$$

lezárás, és a minimális kulcsok halmaza éppen $\mathcal{K}$.

2.7 Állítás [17] *Bármely nem-üres $\mathcal{K}$ Sperner-rendszerhez létezik olyan r reláció, melyben a minimális kulcsok halmaza $\mathcal{K}$.*

Bizonyítás: Az állítás következik a 2.3 Tételből és a 2.5 Állításból valamint (2.8)-ból. ■

Természetesen $\mathcal{K}$ nem határozza meg $\mathcal{L}$-et egyértelműen.

2.8 Definició *A $\mathcal{D}_k^n$ determináció álljon azon $A \hookrightarrow B$ párokból melyekre $B \subseteq A \subseteq \Omega$, valamint azon $A \hookrightarrow B$ párokból melyekre $A, B \subseteq \Omega$ és $|A| \geq k$. Ennek megfelel az*

$$\mathcal{L}_k^n = \mathcal{L}(\mathcal{D}_k^n) = \begin{cases} A & \text{ha } |A| < k, \\ \Omega & \text{ha } |A| \geq k, \end{cases} \quad A \subseteq \Omega \tag{2.9}$$

lezárás. $\mathcal{K}_k^n = \mathcal{K}(\mathcal{L}_k^n) = \binom{\Omega}{k}$ pedig az Ω alaphalmaz k elemű részhalmazainak családja.

2.9 Tétel [6] *A minimális kulcsok száma legfeljebb* $\binom{n}{\lfloor \frac{n}{2} \rfloor}$, *és ez éles.*

Bizonyítás: Az egyenlőtlenség könnyen következik E. SPERNER tételéből [21], mely szerint minden $\mathcal{S}$ tartalmazásmentes halmazrendszerre $|\mathcal{S}| \leq \binom{n}{\lfloor \frac{n}{2} \rfloor}$.
Most tekintsük $\mathcal{D}^n_{\lfloor n/2 \rfloor}$-et. A 2.3 Tétel szerint van olyan r reláció, melynek teljes családja $\mathcal{D}^n_{\lfloor n/2 \rfloor}$. Nyilvánvaló, hogy a minimális kulcsok halmaza ebben a relációban $\mathcal{K}^n_{\lfloor n/2 \rfloor}$, azaz Ω $\lfloor \frac{n}{2} \rfloor$ elemű részhalmazai. Tehát az egyenlőtlenség nem javítható. ∎

2.10 Lemma (YBLM-egyenlőtlenség, [22],[3],[19],[20]) *Ha $\mathcal{S}$ Sperner-rendszer egy n elemű alaphalmazon és f_i az i elemű tagok száma, akkor*

$$\sum_{i=0}^{n} \frac{f_i}{\binom{n}{i}} \leq 1. \blacksquare \tag{2.10}$$

2.11 Tétel [17] *Tegyük fel, hogy az $A \to B$ funkcionális függőségekre $|B - A| \leq k$, ahol $k \leq n/2$. Ekkor a minimális kulcsok száma legfeljebb $\binom{n}{k}$.*

Bizonyítás: A feltétel miatt a kulcsok mérete legalább $n - k$. Alkalmazzuk a 2.10 Lemmát, ahol a Sperner-rendszer a minimális kulcsok halmaza. Ekkor $f_0 = f_1 = \ldots = f_{n-k-1} = 0$. Felhasználva, hogy

$$\binom{n}{i} \leq \binom{n}{n-k} \text{ ha } k \leq n/2, \ n - k \leq i \leq n,$$

kapjuk (2.10)-ből, hogy

$$1 \geq \sum_{i=n-k}^{n} \frac{f_i}{\binom{n}{i}} \geq \sum_{i=n-k}^{n} \frac{f_i}{\binom{n}{n-k}} = \frac{1}{\binom{n}{n-k}} \sum_{i=n-k}^{n} f_i = \frac{|\mathcal{S}|}{\binom{n}{n-k}},$$

amiből a tétel következik. ∎

A fentiekből tudjuk, hogy bármely $\mathcal{D}$ determinációhoz, $\mathcal{L}$ lezáráshoz, vagy $\mathcal{K}$ minimális kulcsok rendszeréhez létezik olyan r reláció, hogy r éppen az adott $\mathcal{D}$-t, $\mathcal{L}$-t, illetve $\mathcal{K}$-t generálja. Ekkor azt mondjuk, hogy r (vagy a neki megfelelő $\mathbf{M}$ mátrix) reprezentálja őket. Nem világos azonban, hogy mekkora a fentieket reprezentáló relációk közül $|r|$ minimuma, más szóval hány sora van a fentieket generáló legkevesebb sorral rendelkező mátrixnak. Jelölje ezeket a minimumokat $s(\mathcal{D})$, $s(\mathcal{L})$, illetve $s(\mathcal{K})$.

2.12 Tétel [6],[8] *Egy adott Ω n-elemű alaphalmazon*

$$\frac{1}{n^2}\binom{n}{\lfloor\frac{n}{2}\rfloor} < \max_{\mathcal{K}}\{s(\mathcal{K})|\ \mathcal{K}\neq\emptyset \text{ Sperner } \Omega\text{-}n\} \leq 1+\binom{n}{\lfloor\frac{n}{2}\rfloor}. \blacksquare \tag{2.11}$$

Az első egyenlőtlenség bizonyítása nem konstruktív, mivel nem ismerjük a (közel) legrosszabb Sperner-rendszereket. Egy lehetséges jelölt $\mathcal{K}^{\,n}_{\lfloor n/2\rfloor}$. Ez az egyik motivációja $s(\mathcal{L}^n_k)$ vizsgálatának. Ugyanis a következő lemma miatt ez ekvivalens $s(\mathcal{K}^n_k)$ vizsgálatával, mivel a két szám megegyezik.

2.13 Lemma [7] *Legyen $\mathcal{L}$ az n-elemű Ω alaphalmaz egy tetszőleges lezárása. Ekkor*

$$\mathcal{K}(\mathcal{L}) = \mathcal{K}^n_k \iff \mathcal{L} = \mathcal{L}^n_k. \tag{2.12}$$

Bizonyítás: Nyilván $\mathcal{K}(\mathcal{L}^n_k) = \mathcal{K}^n_k$. A másik irányhoz elég belátni, hogy $\mathcal{L}(A) = A$, ha $A \subseteq \Omega, |A| < k$, ugyanis ekkor $\mathcal{L}(A) = \Omega,\ |A| \geq k$ könnyen következik $\mathcal{K}(\mathcal{L}) = \mathcal{K}^n_k$ -ből és (2.6)-ból. Legyen tehát $a \in \mathcal{L}(A)\backslash A$. Ekkor létezik egy olyan B halmaz melyre $|B| = k,\ B \supseteq A \cup \{a\}$. (2.5) miatt $\mathcal{L}(B\backslash\{a\}) \supseteq B\backslash\{a\}$, míg (2.6) miatt $\mathcal{L}(B\backslash\{a\}) \supseteq \mathcal{L}(A) \ni a$. Ezért $\mathcal{L}(B\backslash\{a\}) \supseteq B$ Most (2.6)-ból és (2.7)-ből $\mathcal{L}(B\backslash\{a\}) = \mathcal{L}(\mathcal{L}(B\backslash\{a\})) \supseteq \mathcal{L}(B) = \Omega$. Így találtunk egy k-nál kisebb számosságú halmazt, aminek a lezártja Ω. Ez az ellentmondás mutatja (2.5) felhasználásával, hogy $\mathcal{L}(A) = A$. Tehát $\mathcal{L} = \mathcal{L}^n_k$. $\blacksquare$

A következő lemma egyszerűsége ellenére meglepően erősnek bizonyult.

2.14 Lemma [10],[7]

$$\binom{s(\mathcal{L}^n_k)}{2} \geq \binom{n}{k-1} \quad (0 \leq k \leq n). \tag{2.13}$$

Bizonyítás: Tegyük fel, hogy $\mathbf{M}$ reprezentálja $\mathcal{L}^n_k$-et és legyen $|A| = k-1$ Ω egy részhalmaza, továbbá $b \notin A$. Ekkor $\mathcal{L}^n_k$ definíciója miatt $A \not\to b$, azaz létezik két olyan sor (i és j), hogy ezek megegyeznek A-n, de különbözik a b-beli értékük. Ha lenne Ω-nak egy másik $(k-1)$-elemű B részhalmaza, hogy i és j megegyeznek B-n, akkor $A \cup B \not\to b$ állna, de mivel $|A \cup B| \geq k$, ezért $\mathcal{L}^n_k$ definiciója miatt ez lehetetlen. Így a különböző sorpárokhoz különböző oszlop-$(k-1)$-eseket rendelhetünk. $\blacksquare$

Azonban lesz amikor az előző lemma nem elegendő. A következő lemma mátrixok egy egyszerű tulajdonságát mutatja be. Legyen $\mathbf{M}$ egy $m \times n$-es mátrix és $G(\mathbf{M})$ jelölje a következő gráfot: a gráf csúcsai feleljenek meg $\mathbf{M}$ sorainak, és két csúcs pontosan akkor legyen összekötve, ha azon oszlopok A halmaza, ahol a két sor megegyezik nem-üres. Az élet cimkézzük meg A-val.

2.15 Lemma [7] *Legyen* $\mathbf{M}$ *egy mátrix és legyenek* $A_1, \ldots, A_r$ *a* $G(\mathbf{M})$ *gráf egy körének cimkéi. Ekkor*

$$\left(\bigcap_{\substack{i=1\\ i\neq j}}^{r} A_i\right)\backslash A_j = \emptyset \qquad (1 \leq j \leq r). \tag{2.14}$$

Bizonyítás: Átindexelés után feltehető, hogy $j = r$. Indirekt tegyük fel, hogy az u-adik oszlop eleme minden A_i-nek $(1 \leq i < r)$, de nem eleme A_r-nek. Legyenek a kör pontjai $k_1, \ldots, k_r$ úgy, hogy a (k_i, k_{i+1}) él cimkéje legyen A_i $(1 \leq i < r)$ és (k_r, k_1) cimkéje A_r. $u \in A_1$-ből következik, hogy az u-adik oszlop k_1-edik és k_2-edik eleme megegyezik. Hasonlóan kapjuk, hogy az u-adik oszlop k_2-edik és a k_3-adik, $\ldots$, a k_{r-1}-edik és a k_r-edik eleme megegyezik. Ezért így a k_1-edik és a k_r-edik elem megegyezik, tehát $u \in A_r$, ami ellentmond az indirekt feltevésünknek. ∎

Ha adott a minimális kulcsok $\mathcal{K}$ halmaza, akkor jelölje $\mathcal{K}^{-1}$ a maximális antikulcsok halmazát.

2.16 Lemma [7] $\mathbf{M}$ *reprezentálja a* $\mathcal{K}$ *Sperner-rendszert akkor és csak akkor, ha egyrészt bármely* $A \in \mathcal{K}^{-1}$*-hez létezik* $\mathbf{M}$*-nek két különböző sora, hogy ezekben* A *oszlopaiban ugyanaz az érték szerepel, másrészt ha bármely két olyan sor, ami megegyezik* $K \in \mathcal{K}$*-n, megegyezik mindenütt.*

Bizonyítás: Ha $\mathbf{M}$ reprezentálja $\mathcal{K}$-t, akkor $\mathcal{K} = \mathcal{K}(\mathcal{L}_{\mathbf{M}})$. Ha $K \in \mathcal{K}$, akkor $\mathcal{L}_{\mathbf{M}}(K) = \Omega$, és a második feltétel nyilván teljesül. Hasonlóan, ha $A \in \mathcal{K}^{-1}$, akkor $\mathcal{L}_{\mathbf{M}}(K) \neq \Omega$, és így a második feltétel is következik.

Fordítva, ha mindkét feltétel teljesül $\mathbf{M}$-re és $\mathcal{K}$-ra, akkor

(i) $\mathcal{L}_{\mathbf{M}}(A) \neq \Omega$ minden $A \in \mathcal{K}^{-1}$-re és

(ii) $\mathcal{L}_{\mathbf{M}}(K) = \Omega$ minden $K \in \mathcal{K}$-ra.

(ii) és (2.6) szerint $\mathcal{L}_{\mathbf{M}}(C) = \Omega$ ha $C \supseteq K$ valamely $K \in \mathcal{K}$-ra. Tegyük fel most, hogy C nem tartalmazza $\mathcal{K}$ semelyik tagját sem. Ekkor definició szerint létezik olyan $A \in \mathcal{K}^{-1}$, hogy $C \subseteq A$. (i) és (2.6) miatt $\mathcal{L}_{\mathbf{M}}(C) \neq \Omega$. Tehát a minimális kulcsok halmaza éppen $\mathcal{K}$. ∎

2.17 Lemma [7]

$$\binom{s(\mathcal{K})}{2} \geq |\mathcal{K}^{-1}| \geq s(\mathcal{K}) - 1. \tag{2.15}$$

Bizonyítás: A második egyenlőtlenséget a következő egyszerű konstrukció mutatja: A nulladik sor álljon csupa 0-ból, míg az i-edik sor ($1 \leq i \leq |\mathcal{K}^{-1}|$) 0-kból és i-kből. A nullák álljanak $\mathcal{K}^{-1}$ i-edik tagjának oszlopaiban.

A másik irányhoz legyen $A \in \mathcal{K}^{-1}$, ekkor a 2.16 Lemma szerint van két különböző i, j sor, hogy ezek A-ban megegyeznek. Most legyen B $\mathcal{K}^{-1}$ egy másik eleme. Ha ehhez ugyanaz az (i, j) pár tartozna, akkor ezek a sorok megegyeznének $A \cup B$-n. Következésképp $\mathcal{L}(A \cup B) \neq \Omega$ és létezik $C \in \mathcal{K}^{-1}$, hogy $C \supseteq A \cup B$. $\mathcal{K}^{-1}$ definiciója miatt ez csak úgy lehetséges, ha $C = A$ és $C = B$, ami ellentmond az $A \neq B$ feltételnek. Tehát $\mathcal{K}^{-1}$ különböző elemeihez különböző sorpárok rendelhetők, ami bizonyítja az első egyenlőtlenséget. ∎

3. Direkt szorzatok reprezentációja

Keveset tudunk $s(\mathcal{L})$-ről az $\mathcal{L}_k^n$-től különböző lezárásokra. Egy eredmény azonban ismert direkt szorzatok s-függvényére.

3.1 Definició *Legyen* $\Omega = \Omega_1 \cup \Omega_2$ *az* Ω *alaphalmaz egy particiója, és legyenek* $\mathcal{L}_1$ *és* $\mathcal{L}_2$ *két lezárás* Ω_1*-n illetve* Ω_2*-n. Ekkor az* $\mathcal{L}_1 \times \mathcal{L}_2$ *direkt szorzat definiciója*

$$(\mathcal{L}_1 \times \mathcal{L}_2)(A) = \mathcal{L}_1(A \cap \Omega_1) \cup \mathcal{L}_2(A \cap \Omega_2). \tag{3.1}$$

3.2 Tétel [7]

$$s(\mathcal{L}_1 \times \mathcal{L}_2) = s(\mathcal{L}_1) + s(\mathcal{L}_2) - 1. \tag{3.2}$$

Bizonyítás: Először egy konstrukcióval bizonyítjuk az

$$s(\mathcal{L}_1 \times \mathcal{L}_2) \leq s(\mathcal{L}_1) + s(\mathcal{L}_2) - 1. \tag{3.3}$$

egyenlőtlenséget. Reprezentálják az $s(\mathcal{L}_1) \times n_1$ -es $\mathbf{M}_1$ és az $s(\mathcal{L}_2) \times n_2$ -es $\mathbf{M}_2$ mátrixok az $\mathcal{L}_1$ illetve $\mathcal{L}_2$ lezárásokat. Jelölje α az $\mathbf{M}_1$ mátrix utolsó sorát, míg β az $\mathbf{M}_2$ mátrix első sorát. Definiáljuk az $(s(\mathcal{L}_1) + s(\mathcal{L}_2) - 1) \times (n_1 + n_2)$-es $\mathbf{M}$ mátrixot a következőképpen:

$$\begin{array}{c|c|c|}
\multicolumn{1}{c}{} & \multicolumn{1}{c}{\Omega_1} & \multicolumn{1}{c}{\Omega_2} \\ \cline{2-3}
 & & \beta \\
 & \mathbf{M}_1 & \vdots \\
 & & \beta \\ \cline{3-3}
 & \alpha & \beta \\ \cline{2-2}
 & \alpha & \\
 & \vdots & \mathbf{M}_2 \\
 & \alpha & \\ \cline{2-3}
\end{array}$$

Megmutatjuk, hogy ez a mátrix reprezentálja $\mathcal{L}_1 \times \mathcal{L}_2$-t, azaz

$$a \in \mathcal{L}_{\mathbf{M}}(A) \Leftrightarrow a \in \mathcal{L}_1(A \cap \Omega_1) \cup \mathcal{L}_2(A \cap \Omega_2). \tag{3.4}$$

A szimmetria miatt feltehetjük, hogy $a \in \Omega_1$. Ekkor a (3.4) feltétel két implikációra bontható:

$$a \in \mathcal{L}_1(A \cap \Omega_1) \Rightarrow \text{ha } \mathbf{M} \text{ két sora megegyezik } A\text{-ban, akkor } a\text{-ban is,} \tag{3.5}$$

$$a \notin \mathcal{L}_1(A \cap \Omega_1) \Rightarrow \mathbf{M} \text{ két sora megegyezik } A\text{-ban, de különbözik } a\text{-ban.} \tag{3.6}$$

(3.5) bizonyításához tegyük fel, hogy $a \in \mathcal{L}_1(A \cap \Omega_1)$, és válasszuk ki $\mathbf{M}$ két olyan sorát, melyek megegyeznek A-ban. Ha mindkettő α-val kezdődik, akkor természetesen megegyeznek a-ban. Ha legalább az egyik nem α-val kezdődik, akkor ezeknek a soroknak az első fele $\mathbf{M}_1$-nek két különböző sora, így tehát ha megegyeznek $A \cap \Omega_1$-ben, akkor megegyeznek a-ban is.

(3.6) bizonyításához tegyük fel, hogy $a \notin \mathcal{L}_1(A \cap \Omega_1)$. $\mathbf{M}_1$ tartalmaz két sort, melyek megegyeznek $A \cap \Omega_1$-ben, de különböznek a-ban. Ezek a sorok ugyanúgy folytatódnak, tehát megegyeznek A-ban.

Tehát $\mathbf{M}$ reprezentálja $\mathcal{L}_1 \times \mathcal{L}_2$-t, így (3.3)-t bebizonyítottuk.

Most két lemma segítségével bebizonyítjuk a

$$s(\mathcal{L}_1 \times \mathcal{L}_2) \geq s(\mathcal{L}_1) + s(\mathcal{L}_2) - 1. \tag{3.7}$$

egyenlőtlenséget. Reprezentálja az $\mathbf{M}$ mátrix $\mathcal{L}_1 \times \mathcal{L}_2$-t és tegyük fel, hogy az első n_1 oszlop felel meg $\mathcal{L}_1$ Ω_1 alaphalmazának, míg a második n_2 oszlop $\mathcal{L}_2$ Ω_2 alaphalmazának. Azt szeretnénk bizonyítani, hogy $\mathbf{M}$-nek legalább $s(\mathcal{L}_1)+s(\mathcal{L}_1)-1$ sora van. Jelölje az első n_1 oszlop által meghatározott részmátrixot $\mathbf{M}_1$, míg a maradékot $\mathbf{M}_2$.

3.3 Lemma $\mathcal{L}_{\mathbf{M}_2} = \mathcal{L}_2$.

Bizonyítás: Tegyük fel, hogy $A \subseteq \Omega_2$, $a \in \Omega_2$ és $a \in \mathcal{L}_2(A)$. Ekkor $a \in (\mathcal{L}_1 \times \mathcal{L}_2)(A)$. Ha $\mathbf{M}$ két sora megegyezik A-ban, akkor $\mathbf{M}$ definiciója miatt a-ban is, és ez igaz marad akkor is, ha csak az $\mathbf{M}_2$ részmátrixot tekintjük. Azaz bebizonyítottuk, hogy $a \in \mathcal{L}_{\mathbf{M}_2}(A)$. Fordítva, most tegyük fel, hogy $A \subseteq \Omega_2$, $a \in \Omega_2$ és $a \notin \mathcal{L}_2(A)$. Mivel így $a \notin (\mathcal{L}_1 \times \mathcal{L}_2)(A)$, ezért $\mathbf{M}$-nek van két olyan sora, hogy ezek megegyeznek A-ban, de különböznek a-ban. Ezeknek a soroknak $\mathbf{M}_2$-be eső része mutatja, hogy $a \notin \mathcal{L}_{\mathbf{M}_2}(A)$. $\blacksquare$

Analóg módon $\mathcal{L}_{\mathbf{M}_1} = \mathcal{L}_1$. Azonban $\mathbf{M}_1$-re egy valamivel erősebb állításra van szükség:

3.4 Lemma *Tegyük fel, hogy az* $\mathbf{N}$ *mátrix sorai* k *osztályba sorolhatók úgy, hogy ha* $a \notin \mathcal{L}_{\mathbf{N}}(A)$, *akkor létezik egy osztályban két olyan sor, melyek megegyeznek* A*-ban, de különböznek* a*-ban. Ekkor*

$$(\mathbf{N} \text{ sorainak száma}) \geq s(\mathcal{L}_{\mathbf{N}}) + k - 1. \tag{3.8}$$

Bizonyítás: k-ra vonatkozó indukcióval bizonyítunk. $k = 1$-re (3.8) $s(\mathcal{L}_{\mathbf{N}})$ definiciója miatt igaz. Tegyük fel, hogy $\mathbf{N}$ már particionálva van $k \geq 2$, a lemma feltételeit teljesítő osztályra, és hogy az állítás igaz kisebb értékekre.

$\mathcal{L}_{\mathbf{N}}$ csak attól függ, hogy $\mathbf{N}$ mely elemei egyenlők, illetve melyek nem. Ezért feltehető, hogy a mátrix csak pozitív egész értékeket tartalmaz.

Töröljük $\mathbf{N}$ azon oszlopait, melyekben csak egyfajta érték áll, és jelöljük az új mátrixot $\mathbf{N}_1$-nel. $\mathbf{N}$ sorainak particiója $\mathbf{N}_1$-nek is egy "jó" particióját adja. Másfelől nyilván

$$s(\mathcal{L}_{\mathbf{N}_1}) = s(\mathcal{L}_{\mathbf{N}}). \tag{3.9}$$

Továbbá $\mathcal{L}_{\mathbf{N}_1}(\emptyset) = \emptyset$.

Legyenek $p_1, p_2, \ldots, p_k$ $\mathbf{N}_1$ minden eleménél nagyobb prímszámok. Szorozzuk be az i-edik sorosztály minden elemét p_i-vel $(1 \leq i \leq k)$. Jelölje az új mátrixot $\mathbf{N}_2$. Könnyen látható, hogy

$$\mathcal{L}_{\mathbf{N}_2} = \mathcal{L}_{\mathbf{N}_1}, \tag{3.10}$$

mivel $\mathcal{L}_{\mathbf{N}_1}(\emptyset) = \emptyset$ és különböző sorosztályokban $\mathbf{N}_2$ nem tartalmaz egyforma elemeket.

Legyen $\gamma = (\gamma_1, \ldots, \gamma_u)$ illetve $\delta = (\delta_1, \ldots, \delta_u)$ az $\mathbf{N}_2$ első illetve második osztályának egy-egy sora. Most töröljük a γ sort és cseréljünk ki minden γ_i-t δ_i-re az i-edik oszlopban (minden $1 \leq i \leq u$ -ra). Jelölje az új mátrixot $\mathbf{N}_3$. Ennek eggyel kevesebb sora van, mint $\mathbf{N}_2$-nek. Most bizonyítsuk be, hogy

$$\mathcal{L}_{\mathbf{N}_3} = \mathcal{L}_{\mathbf{N}_2}. \tag{3.11}$$

Tegyük fel először, hogy $a \in \mathcal{L}_{\mathbf{N}_2}(A)$ és válasszuk ki $\mathbf{N}_3$ két sorát, μ_3-t és ν_3-t, melyek megegyeznek A-ban. A megfelelő sorokat $\mathbf{N}_2$-ben jelölje μ_2 illetve ν_2. Ha $\mathbf{N}_2$- ben μ_2 és ν_2 azonos osztályban van, de nem az elsőben, akkor $\mu_2 = \mu_3$, $\nu_2 = \nu_3$. Ezért μ_2 és ν_2 megegyeznek A-ban és így $a \in \mathcal{L}_{\mathbf{N}_2}(A)$ miatt a-ban is. Ugyanez áll μ_3-ra és ν_3-ra, tehát $a \in \mathcal{L}_{\mathbf{N}_3}(A)$. Ha μ_2 és ν_2 egyaránt az első osztályban van, akkor μ_3 csak annyiban különbözik μ_2-től, hogy γ_i helyett δ_i áll mindenütt. Ugyanez igaz ν_2-re és ν_3-ra. Következésképp μ_2 és ν_2 megegyezik A-ban, és így a-ban is. Kapjuk tehát, hogy μ_3 és ν_3 is megegyezik a-ban és így $a \in \mathcal{L}_{\mathbf{N}_3}(A)$. Az utolsó esetben μ_2

és ν_2 különböző osztályban vannak. Mivel a feltevés szerint μ_3 és ν_3 megegyezik A-ban, ezért vagy $A = \emptyset$, vagy μ_2 és ν_2 az első és a második sorosztályból való. $\mathcal{L}_{\mathbf{N}_2}(\emptyset) = \emptyset$ miatt az első esetet kizárhatjuk. Tehát $\mu_2 \neq \gamma$ az első osztályban van, míg ν_2 a másodikban. Nyilván $\nu_2 = \nu_3$. Mivel μ_3 és ν_3 megegyezik A-ban, ezért mindkettőben az i-edik helyen $i \in A$-ra δ_i-nek kell állnia. Ekkor $\nu_2 = \nu_3$ és δ megegyeznek A-ban, következésképp a-ban is. Jelölje a közös értéket itt δ_a. Ha μ_3-ban az i-edik helyen δ_i áll $i \in A$-ra, akkor μ_2 ott γ_i-t tartalmaz. Tehát μ_2 és γ megegyezik A-ban és így a-ban is. A közös értékük itt γ_a. Ebből azt kapjuk μ_3-nak is δ_a az értéke ebben az oszlopban. Ezért μ_3 és ν_3 megegyezik a-ban, és így $a \in \mathcal{L}_{\mathbf{N}_3}(A)$.

Tegyük most fel, hogy $a \notin \mathcal{L}_{\mathbf{N}_2}(A)$. Ekkor van $\mathbf{N}_2$-nek két sora, hogy ezek megegyeznek A-ban, de különböznek a-ban. Ha $A \neq \emptyset$, akkor ez a két sor ugyanabban az osztályban van, következésképp a megfelelő sorok $\mathbf{N}_3$-ban szintén megegyeznek A-ban és különböznek a-ban (γ-nak δ felel meg). Tehát ebben az esetben $a \notin \mathcal{L}_{\mathbf{N}_3}(A)$. $A = \emptyset$, $a \in \mathcal{L}_{\mathbf{N}_3}(\emptyset)$ azt jelen tené, hogy van egy olyan oszlop, amiben ugyanazok az értékek szerepelnek. Ez $k \geq 3$-ra lehetetlen. $k = 2$-re ez csak úgy lehetséges, ha $\mathbf{N}_2$ csak γ_a-t és δ_a-t tartalmaz az a-nak megfelelő oszlopban. Azonban ebben az esetben nem tudnánk találni két sort ugyanabban az osztályban, melyek teljesítik a lemma feltételét $a \notin \mathcal{L}_{\mathbf{N}_2}(\emptyset)$-ra. Ez az ellentmondás mutatja, hogy $a \notin \mathcal{L}_{\mathbf{N}_3}(A)$, és így (3.11) -t bebizonyítottuk.

Sőt $\mathbf{N}_3$-ra teljesül a lemma feltétele (legalább) $k-1$ osztállyal. Így az indukciós feltevés alapján:

$$(\mathbf{N}_3 \text{ sorainak száma}) \geq s(\mathcal{L}_{\mathbf{N}_3}) + k - 2.$$

Ebből és (3.9)-(3.11)-ből kapjuk:

$$(\mathbf{N} \text{ sorainak száma}) \geq s(\mathcal{L}_{\mathbf{N}}) + k - 1,$$

amivel a lemmát bebizonyítottuk. ■

Térjünk most vissza a tétel bizonyításához, pontosabban (3.7)-hez. Particionáljuk $\mathbf{M}_1$ sorait aszerint, hogy a folytatásuk $\mathbf{M}_2$-ben megegyezik-e. Célunk a 3.4 Lemmát alkalmazni $\mathbf{M}_1$-re. A 3.3 Lemma alapján $\mathcal{L}_{\mathbf{M}_1} = \mathcal{L}_1$. Tegyük fel, hogy valamely $A \subseteq \Omega_1$-re és $a \in \Omega_1$-re $a \notin \mathcal{L}_{\mathbf{M}_1}(A) = \mathcal{L}_1(A)$. Ekkor $a \notin \mathcal{L}_1(A) \cup \Omega_2 = (\mathcal{L}_1 \times \mathcal{L}_2)(A \cup \Omega_2)$ teljesül és ezért $\mathbf{M}$-nek van olyan két sora, melyek megegyeznek $A \cup \Omega_2$-n, de különböznek a-ban. Másszóval van $\mathbf{M}_1$-nek két sora, melyek megegyeznek A-ban, különböznek a-ban, és ugyanabban a particióosztályban vannak. Alkalmazzuk a 3.3 Lemmát $\mathbf{M}_1$-re:

$$(\mathbf{M}_1 \text{ sorainak száma}) \geq s(\mathcal{L}_1) + (\mathbf{M}_2 \text{ különböző sorainak száma}) - 1. \qquad (3.12)$$

Ismét felhasználva a 3.3 Lemmát kapjuk

$$(\mathbf{M}_2 \text{ különböző sorainak száma}) \geq s(\mathcal{L}_2). \tag{3.13}$$

(3.12)-ből és (3.13)-ból:

$$(\mathbf{M} \text{ sorainak száma}) \geq s(\mathcal{L}_1) + s(\mathcal{L}_2) - 1,$$

ami bizonyítja (3.7)-t és így a tételt is. ■

Analóg módon definiálhatjuk Sperner-rendszerek direkt szorzatát.

3.5 Definició *Legyen $\Omega = \Omega_1 \cup \Omega_2$ az Ω alaphalmaz egy particiója, és legyenek $\mathcal{K}_1$ és $\mathcal{K}_2$ Sperner-rendszerek Ω_1-n illetve Ω_2-n. Ekkor a direkt szorzat definiciója $\mathcal{K}_1 \times \mathcal{K}_2 = \{A \cup B | \; A \in \mathcal{K}_1, B \in \mathcal{K}_2\}$*

Az előző tétel bizonyításában szereplő konstrukció mutatja, hogy $s(\mathcal{K}_1 \times \mathcal{K}_2) \leq s(\mathcal{K}_1) + s(\mathcal{K}_2) - 1$. Bár számos esetben fennáll az egyenlőség, a 3.2 Tétel állítása nem igaz kulcsrendszerekre.

3.6 Példa [7] Legyen $\Omega_1 = \{1,2,3,4,5\}$, $\Omega_2 = \{6,7,8,9,10\}$, $\mathcal{K}_1 = \{\{1,2\}, \{3,4\}, \{1,5\}, \{2,5\}, \{3,5\}, \{4,5\}\}$, $\mathcal{K}_2 = \{\{6,7\}, \{8,9\}, \{6,10\}, \{7,10\}, \{8,10\}, \{9,10\}\}$. Könnyen látható, hogy

$$\mathcal{K}_1^{-1} = \{\{5\}, \; \{1,3\}, \; \{1,4\}, \; \{2,3\}, \; \{2,4\}\}.$$

A 2.14 Lemma szerint $s(\mathcal{K}_1) \geq 4$. Tegyük fel, hogy $s(\mathcal{K}_1) = 4$ és ezt az $\mathbf{M}$ mátrix realizálja. Ekkor a 2.15 Lemmában szereplő $G(\mathbf{M})$ gráfnak 4 csúcsa van és 5 éle $\mathcal{K}^{-1}$ elemeivel cimkézve. Néhány eset megkülönböztetésével látható, hogy a 2.15 Lemma miatt a hatodik él cimkéje $\{1,2\}$, $\{1,2,5\}$, $\{3,4\}$ vagy $\{3,4,5\}$ lehet. Ezek azonban mind kulcsok, ami ellentmondás. Tehát $s(\mathcal{K}_1) = s(\mathcal{K}_2) \geq 5$. Másfelől a következő mátrix mutatja, hogy $s(\mathcal{K}_1 \times \mathcal{K}_2) \leq 8$:

$$\left|\begin{array}{ccccc|ccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 2 \\
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 2 & 1 & 1 & 1 & 1 & 0
\end{array}\right|$$

(nem triviális, $(\mathcal{K}_1 \times \mathcal{K}_2)^{-1}$ bonyolult struktúrája miatt).

4. $s(\mathcal{L}_k^n)$ nagyságrendje

Ebben a fejezetben rögzített k mellett $n \to \infty$-re vizsgáljuk $s(\mathcal{L}_k^n)$ és $s(\mathcal{L}_{n-k}^n)$ értékét.

4.1 Tétel [7]

$$\sqrt{2}\left(\tfrac{1}{k-1}\right)^{(k-1)/2} n^{(k-1)/2} < s(\mathcal{L}_k^n) < 2^{3k/2} n^{(k-1)/2} \quad (2 \le k < n). \tag{4.1}$$

Bizonyítás: (4.1) első egyenlőtlensége a 2.14 Lemma egyszerű következménye. Most mutatunk egy konstrukciót, mely bizonyítja a második egyenlőtlenséget.

Legyen p egy prímszám. Megmutatjuk, hogy létezik egy $2\lfloor\sqrt{p}\rfloor$ számosságú D halmaz, hogy minden i egész számhoz léteznek $d_1, d_2 \in D$, hogy

$$i \equiv d_1 - d_2 \pmod{p}. \tag{4.2}$$

Egy ilyen D halmaz definiciója például:

$$D = \{0, 1, 2, \ldots, a-1, 2a, 3a, \ldots, (a-1)a\},$$

ahol $a = \lceil\sqrt{p}\rceil$. Tegyük fel, hogy valamely i-re $0 \le i < p$ és $i = al + r$ $(0 \le r < a)$. Ha $1 \le l \le a-2$ és $0 < r < a$, akkor $d_1 = (l+1)a$ és $d_2 = a - r$ teljesíti (4.2)-t. Ha $i = al$ $(2 \le l \le a-1)$, akkor $d_1 = al$ és $d_2 = 0$ megfelelő. $a = 3a - 2a$, míg a többi kifejezhető 0 és az $1, 2, \ldots, a-1$ számok különbségeként ($p = 2$-re vagy $p = 3$-ra $D = \{0, 1\}$ jó). Ekkor D számossága:

$$|D| = 2a - 2 = 2(\lceil\sqrt{p}\rceil - 1) = 2\lfloor\sqrt{p}\rfloor.$$

Most definiáljuk polinomok egy $2^k\lfloor\sqrt{p}\rfloor^{k-1}$ számosságú $\mathcal{P}$ halmazát:

$$\mathcal{P} = \{c_{k-1}x^{k-1} + c_{k-2}x^{k-2} + \ldots + c_1x + c_0 \mid c_0, c_1, \ldots, c_{k-1} \in D, c_{k-1} = 0 \text{ vagy } 1\}.$$

Legyen $\mathbf{M}$ egy $|\mathcal{P}| \times p$-es mátrix. A sorok feleljenek meg $\mathcal{P}$ tagjainak. A $z(x) \in \mathcal{P}$-nek megfelelő sor j. eleme legyen $z(j) \pmod{p}$ $(0 \le j \le p-1,$

$0 \le z(j) \le p-1$). Belátjuk, hogy $\mathbf{M}$ reprezentálja $\mathcal{L}_k^p$ -t. A 2.13 Lemma alapján elegendő belátni, hogy $\mathbf{M}$ reprezentálja $\mathcal{K}_k^p$-t. Elég azonban csak a 2.16 Lemma feltételeit ellenőrizni $\mathcal{K} = \binom{\Omega}{k}$, $\mathcal{K}^{-1} = \binom{\Omega}{k-1}$-re (ahol $|\Omega| = p$).

Tegyük fel, hogy a $z_1(x)$-nek és $z_2(x)$-nek megfelelő sorok k helyen megegyeznek:

$$z_1(t_i) \equiv z_2(t_i) \pmod p \; (0 \le t_1 < t_2 < \ldots < t_k < p).$$

A $z_1(x) - z_2(x)$ legfeljebb $(k-1)$-edfokú polinomnak k különböző gyöke van. Ez az ellentmondás mutatja, hogy z_1 és z_2 ugyanaz, és a "két" sor valójában egy.

Válasszuk most a $0 \le t_1 < t_2 < \ldots < t_{k-1} < p$ egész számokat tetszőlegesen. Találnunk kell két sort, melyek megegyeznek a t_1-edik, t_2-edik,...,t_{k-1}-edik helyen. Tekintsük a

$$w(x) = (x-t_1)(x-t_2)\ldots(x-t_{k-1}) = x^{k-1} + a_{k-2}x^{k-2} + \ldots + a_1x + a_0$$

polinomot. a_i-hez ($0 \le i \le k-1$) található D-nek két c_i és c_i' eleme, hogy $a_i \equiv c_i - c_i' \pmod p$. Ekkor $w(x) = z(x) - z'(x)$, ahol

$$z(x) = x^{k-1} + c_{k-2}x^{k-2} + \ldots + c_1x + c_0 \text{ és}$$

$$z'(x) = c'_{k-2}x^{k-2} + \ldots + c'_1x + c'_0.$$

$z(x)$ és $z'(x)$ nyilván két különböző eleme $\mathcal{P}$-nek. Másfelől valóban $z(t_i) \equiv z'(t_i) \pmod p$. A 2.16 Lemma mindkét feltételét ellenőriztük, így $\mathbf{M}$ tényleg reprezentálja $\mathcal{L}_k^p$-t. Tehát:

$$s(\mathcal{L}_k^p) \le 2^k p^{(k-1)/2}.$$

Tetszőleges n-re válasszunk egy p prímet, melyre $n \le p \le 2n$. Ilyen P.L. Csebisev tétele (például [13]-ban) szerint létezik. Most készítsünk egy mátrixot, mely reprezentálja $\mathcal{L}_k^p$-t, és hagyjuk el $p-n$ oszlopát. Ez a mátrix reprezentálja $\mathcal{L}_k^n$-t, ezért

$$s(\mathcal{L}_k^n) \le 2^k p^{(k-1)/2} \le 2^k (2n)^{(k-1)/2} \le 2^{3k/2} n^{(k-1)/2},$$

tehát a tételt bebizonyítottuk. ■

Azt mondjuk, hogy egy n oszlopból álló $\mathbf{M}$ mátrix egy tökéletes k-hibajavító adatbázis, ha tetszőleges k oszlopát törölve a megmaradt mátrix bármely két sora különbözik, de bármelyik $k+1$ sorát törölve lesz két sor, melyek megegyeznek. Egy n attributumú tökéletes k-hibajavító adatbázis lehetséges sorai számának meghatározása nyilván ekvivalens $s(\mathcal{L}^n_{n-k})$ meghatározásával. A következő tétel első egyenlőtlensége megjavítja erre az esetre a 2.14 Lemmát.

4.2 Tétel [14] $n > n_0(k)$, $k \geq 1$-re

$$\tfrac{1}{(k+1)!} n^{(2k+1)/3} < s(\mathcal{L}^n_{n-k}) < \tfrac{1}{k!} n^k. \tag{4.3}$$

Bizonyítás: A második egyenlőtlenséget egy egyszerű konstrukció mutatja. Legyenek Ω k-elemű részhalmazai $E_1, E_2, \ldots, E_{\binom{n}{k}}$. Az i-edik sor j-edik helyén álljon 0 ha $j \notin E_i$, ha pedig $j \in E_i$, akkor i. Az így kapott mátrix $k \geq 1$-re reprezentálja $\mathcal{L}^n_{n-k}$-t.

A másik irány bizonyításához tegyük fel, hogy $\mathbf{M}$ reprezentálja $\mathcal{L}^n_{n-k}$-t. Felhasználjuk Kővári-Sós-Turán egy extremális gráfelméleti eredményét.

4.3 Lemma [18] *Tegyük fel, hogy a $\mathcal{G}(V,E)$ gráf nem tartalmazza részgráfként a $\mathcal{K}(2,p)$ teljes páros gráfot. Ekkor*

$$|E| < \sqrt{\frac{p-1}{2}} |V|^{3/2} + \frac{|V|}{2}. \ \blacksquare \tag{4.4}$$

Szükségünk van még néhány jelölésre. Jelölje V az adatbázis sorainak a halmazát. Legyenek $\alpha, \beta \in V$,ekkor $D(\alpha,\beta) = \{i | \ \alpha(i) \neq \beta(i)\}$. Jelölje $H(\alpha,\beta) = |D(\alpha,\beta)|$ az α és β sorok Hamming-távolságát. Definiáljuk az $\mathbf{M}$-hez tartozó $\mathcal{G}$ minimális távolsággráfot a V csúcshalmazon úgy, α-t és β-t pontosan akkor kössük össze éllel, ha $H(\alpha,\beta) = k+1$. $(\alpha,\beta) \in \mathcal{G}$ színe legyen D (ahol $D \subseteq \Omega$, $|D| = k+1$), ha $D = D(\alpha,\beta)$. Jelölje szokásos módon $\deg(\alpha)$ α fokát, $\Gamma(\alpha)$ pedig α szomszédainak halmazát. Minden $k+1$ elemű $D \subseteq \Omega$-hoz válasszuk ki $\mathcal{G}$ egy D színű élét. Ez az $\binom{n}{k+1}$ pár alkotja a $\mathcal{G}_0$ redukált minimális távolsággráfot. Ebben $\deg_0$-lal illetve Γ_0-lal jelöljük egy pont fokát illetve szomszédai halmazát.

4.4 Lemma
(i) amennyiben $(\alpha,\beta) \notin \mathcal{G}$, akkor $|\Gamma(\alpha) \cap \Gamma(\beta)| < 3^{2k+2}$,
(ii) $n > n_0(k)$ esetén tetszőleges $\alpha, \beta \in V$-re $|\Gamma_0(\alpha) \cap \Gamma_0(\beta)| < n3^k$.

Bizonyítás: Legyen $C = \{x \in \Omega | \ \alpha(x) \neq \beta(x)\}$, ekkor $|C| = H(\alpha,\beta)$. Feltehető, hogy $|C| \leq 2k+2$, ugyanis különben a lemma triviálisan igaz, mert ekkor nem létezik $\gamma \in V$, hogy $\{\alpha,\gamma\} \in \mathcal{G}, \{\beta,\gamma\} \in \mathcal{G}$.

Vegyük észre, hogy minden $x \in C$-re $\gamma(x)$ különbözik $\alpha(x)$-től vagy $\beta(x)$-től, ezért

$$C \subseteq D(\alpha,\gamma) \cup D(\beta,\gamma). \tag{4.5}$$

Továbbá $D(\alpha,\gamma)\backslash C = D(\beta,\gamma)\backslash C$.

Szükségünk van egy állításra, mely szerint γ -t (majdnem) meghatározza $D(\alpha,\gamma)$ és $D(\beta,\gamma)$ C-beli nyoma.

4.5 Állítás *Tegyük fel, hogy $\gamma,\gamma' \in \Gamma(\alpha)\cap\Gamma(\beta), \gamma \neq \gamma'$, valamint azt, hogy $D(\alpha,\gamma)\cap C = D(\alpha,\gamma')\cap C$ és $D(\beta,\gamma)\cap C = D(\beta,\gamma')\cap C$. Ekkor $|C| = k+1$ és $(D(\alpha,\gamma)\backslash C) = (D(\alpha,\gamma')\backslash C) = \emptyset$.*

Bizonyítás: (4.5) szerint $D(\gamma,\gamma') \subseteq D(\alpha,\gamma)\cup D(\alpha,\gamma')$. γ és γ' megegyezik $C\backslash(D(\alpha,\gamma)\cap D(\beta,\gamma))$-n, ezért

$$H(\gamma,\gamma') \leq |D(\alpha,\gamma)\cap D(\beta,\gamma)\cap C| + |D(\alpha,\gamma)\backslash C| + |D(\alpha,\gamma')\backslash C|. \tag{4.6}$$

Mivel $|D(\alpha,\gamma')\backslash C| = |D(\beta,\gamma)\backslash C|$, azért (4.6) jobb oldala egyenlő $2(k+1) - |C|$-kel. Ha $|C| > k+1$, akkor ez a $H(\gamma,\gamma') < k+1$ ellentmondáshoz vezet. Így $|C| = k+1$, és (4.6)-ban egyenlőség áll fenn. Ezért a $D(\alpha,\gamma)\backslash C$ és $D(\alpha,\gamma')\backslash C$ halmazok diszjunktak. ■

Most térjünk vissza 4.4 Lemma bizonyításához. Az (i) esetben a 4.5 Állítás szerint $|\Gamma(\alpha)\cap\Gamma(\beta)|$ legfeljebb annyi, mint azon $A, B \subseteq C = D(\alpha,\beta)$ halmazpárok száma, melyekre $|A| = |B|$, $A\cap B = \emptyset$ és $|A| \geq |C|-k-1$. (Itt $A = D(\alpha,\gamma)\backslash D(\beta,\gamma)$ és $B = D(\beta,\gamma)\backslash D(\alpha,\gamma)$.) Ezért

$$|\Gamma(\alpha)\cap\Gamma(\beta)| \leq \sum_{i\geq|C|-k-1} \binom{|C|}{2i}\binom{2i}{i} < 3^{|C|} \leq 3^{2k+2}.$$

(ii) bizonyításához most már feltehető, hogy $|C| = k+1$. Válasszunk $A, B \subseteq C$ halmazokat, melyekre $|A| = |B| = i$, $A\cap B = \emptyset$, és tekintsük az összes $\gamma \in \Gamma_0(\alpha)\cap\Gamma_0(\beta)$-t, hogy $A = D(\alpha,\gamma)\backslash D(\beta,\gamma)$ és $B = D(\beta,\gamma)\backslash D(\alpha,\gamma)$. Ekkor $\mathcal{G}_0$ definiciója miatt tetszőleges $\gamma \in \Gamma_0(\alpha)$-ra $|D(\alpha,\gamma)\backslash C| = i > 0$. Így a 4.5 Állítás szerint az ilyen γ-k száma legfeljebb $(n-|C|)/i$. Ezért

$$|\Gamma_0(\alpha)\cap\Gamma_0(\beta)| \leq \sum_{i\geq 1}\binom{|C|}{2i}\binom{2i}{i}\frac{n-k-1}{i} < n\sum_{i\geq 1}\binom{k+1}{2i}\binom{2i}{i}\frac{1}{i} < n3^k. \quad ■$$

A 4.2 tétel bizonyításához tekintsük a $\mathcal{G}_0$ **M**-hez tartozó redukált minimális távolsággráfot. A 4.4 Lemma szerint $\mathcal{G}_0$ nem tartalmazza részgráfként a $\mathcal{K}(2, n3^k)$ teljes páros gráfot. Ekkor a 4.3 Lemma szerint

$$\binom{n}{k+1} = |\mathcal{G}_0| \leq \sqrt{\frac{n3^k}{2}}|V|^{3/2} + \frac{|V|}{2},$$

amiből a tétel következik. ■

5. $s(\mathcal{L}_k^n)$ meghatározása k néhány értékére

Ebben a fejezetben $s(\mathcal{L}_k^n)$ néhány konkrét értékét határozzuk meg.

5.1 Tétel [10],[7]

$$s(\mathcal{L}_1^n) = 2, \qquad s(\mathcal{L}_2^n) = \lceil (1+\sqrt{1+8n})/2 \rceil,$$
$$s(\mathcal{L}_{n-1}^n) = n, \qquad s(\mathcal{L}_n^n) = n+1.$$

Bizonyítás: Az első három esetben a tétel azt állítja, hogy a 2.14 Lemmából származó becslés (természetesen ha nem egész, akkor a felső egészrésze) éles. Tehát elég 3 konstrukciót mutatni. Hogy ezek tényleg jók, a 2.16 Lemma alapján könnyen ellenőrizhető.

$k=1$ esetén:

$$\begin{array}{|cccc|}
\hline
0 & 0 & \ldots & 0 \\
1 & 1 & \ldots & 1 \\
\hline
\end{array}$$

$k=2$ esetén: (minden oszlopban pontosan 2 darab nulla áll úgy, hogy különböző oszlopokhoz tartozó nullák különböző sorpárokhoz tartoznak)

$$\begin{array}{|ccccccccc|}
\hline
0 & 0 & 0 & \ldots & 0 & 1 & 1 & \ldots & 1 \\
0 & 2 & 2 & \ldots & 2 & 0 & 0 & \ldots & 2 \\
3 & 0 & 3 & \ldots & 3 & 0 & 3 & \ldots & 3 \\
4 & 4 & 0 & \ldots & 4 & 4 & 0 & \ldots & 4 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
s-1 & s-1 & s-1 & \ldots & s-1 & s-1 & s-1 & \ldots & s-1 \\
s & s & s & \ldots & 0 & s & s & \ldots & s \\
\hline
\end{array}$$

$k=n-1$ esetén:

$$\begin{array}{|cccc|}
\hline
1 & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1 \\
\hline
\end{array}$$

A $k = n$ esetben a 2.15 Lemma van segítségünkre. Tegyük fel, hogy egy $\mathbf{M}$ mátrix reprezentálja $\mathcal{K}_n^n = \{\Omega\}$-t. Ekkor Ω bármely $(n-1)$-elemű A részhalmazához létezik egy él $G(\mathbf{M})$-ben, melynek cimkéje A. Ezen élekből alkotott részgráf a lemma szerint körmentes. Így $G(\mathbf{M})$-nek legalább $n+1$ csúcsa van, tehát $s(\mathcal{L}_n^n) \geq n+1$. Az egyenlőséget a következő példa mutatja:

$$\begin{array}{|cccc|} \hline 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ \hline \end{array}$$

Ezzel a tétel bizonyítását befejeztük. ■

Még két k értékre ismertek tételek. Ezeket bizonyítás nélkül megemlítjük.

$s(\mathcal{L}_{n-2}^n)$-re ad jobb becslést a 4.2 Tétel első egyenlőtlenségénél a következő tétel, mely egyben megmondja a pontos nagyságrendet is.

5.2 Tétel [14]

$$\tfrac{1}{12}n^2 < s(\mathcal{L}_{n-2}^n) < \tfrac{1}{2}n^2. \blacksquare \tag{5.1}$$

A 2.14 Lemmából éppen $s(\mathcal{L}_3^n) \geq n$ adódik, egyenlőség akkor áll, ha tudunk konstruálni egy $n \times n$-es $\mathbf{M}$ mátrixot, melyre:

1) bármely különböző $a, b, c \in \Omega$-hoz létezik két sor, melyek megegyeznek a-ban és b-ben, de különböznek c-ben,
2) bármely különböző $a, b, c \in \Omega$-hoz nem létezik két sor, melyek mindháromban megegyeznek.

Tekintsük a duális problémát. Egy oszlop természetes módon meghatározza a sorok V halmazának egy particióját az elemek egyenlősége alapján. Azt mondjuk, hogy egy partició lefedi az (α, β) párt $(\alpha, \beta \in V, \alpha \neq \beta)$ akkor és csak akkor, ha α és β a particiónak ugyanabban az osztályában van. Ekkor a fenti két feltétel a következőképpen fogalmazható:

Találjunk n particióját V-nek ($|V| = n$), melyekre:

1') bármely két particióhoz létezik egy (α, β) pár, melyet mindkettő lefed,
2') semelyik (α, β) párt nem fedi le három partició.

Azonban a particiópárok száma szintén $\binom{n}{2}$ és 2') miatt különböző particiópárok nem fedhetik le ugyanazt a párt. Tehát 1') és 2') ekvivalens a következőkkel:

(i) bármely két particióhoz pontosan egy pár létezik, melyet mindkettő lefed,
(ii) bármelyik párt pontosan két partició fedi le.

5.3 Definició [15] *Az (i)-t és (ii)-t teljesítő partíciók egy osztályát ortogonális dupla fedésnek nevezzük.*

A mi szempontunkból a következő eredmény a leglényegesebb.

5.4 Tétel [2] [16] *$n \geq 7$, $n \neq 8$-ra létezik az n-elemű halmaznak n partícióval történő ortogonális dupla fedése.* ■

Az ortogonális dupla fedések elmélete szerteágazó, általában design-típusú problémákra vezet. Rövid összefoglalók találhatók például [17]-ben és [11]-ben a legújabb eredményekről, ugyanitt további hivatkozások is találhatók.

6. Egy kulcsrendszer reprezentációja

Ebben a fejezetben egy, az eddigiektől eltérő $\mathcal{K}_n$ kulcsrendszer minimális reprezentációja kerül tárgyalásra. Először egy, a bizonyítás lényegét adó kombinatorikai segédtételt (6.5 Tétel) bizonyítok. Majd ez alapján a 6.6 Tételben belátom, hogy $s(\mathcal{K}_n)$ $\sqrt{|\mathcal{K}_n^{-1}|}$ és $|\mathcal{K}_n^{-1}|$ közül (2.17 Lemma) az utóbbihoz van közelebb. Végül a 6.7 tételben $|\mathcal{K}_n^{-1}|$ nagyságrendjének meghatározása teszi teljessé a bizonyítást.

6.1 Definició *Egy F fát nevezzünk irányított fának, ha fa, és ha van az éleknek egy olyan irányítása, hogy pontosan egy csúcsból (gyökér) érhető el a többi az irányítás szerint. Jelöljük egy v csúcs kimenő szomszédai halmazát $N(v)$-vel, míg (a gyökér kivételével) az egyértelműen létező bemenő szomszédot $n(v)$-vel. A fa levelei halmazát jelölje $l(F)$.*

6.2 Definició *Legyen adva egy U (véges) halmaz. Egy $F = F(U)$ fát nevezzünk cimkézett fának, ha F minden v csúcsához hozzá van rendelve U-nak egy $A(v)$ részhalmaza.*

Legyen U egy tetszőleges m elemű halmaz, mondjuk $U = \{1, 2, ..., m\}$ $(m \geq 2)$. Tekintsük irányított cimkézett fáknak azt az $\mathcal{F} = \mathcal{F}^{(m)}$ családját, amely család F fáinak csúcsai a következőképpen vannak cimkézve. A g gyökér cimkéje $A(g) = U$. Tetszőleges gyökértől különböző v csúcsra pedig valamely $N(v) = N_0(v) \dot\cup N_1(v)$ diszjunkt felbontás mellett:

$$A(v) \subseteq A(n(v)), \tag{6.1}$$

$$|A(v)| \geq 2, \tag{6.2}$$

$$w_1, w_2 \in N_i(v) \Rightarrow A(w_1) \cap A(w_2) = \emptyset \ (i = 0, 1), \tag{6.3}$$

$$w_1 \in N_i(v), w_2 \in N_{1-i}(v) \Rightarrow |A(w_1) \cap A(w_2)| \leq 1 \ (i = 0, 1). \tag{6.4}$$

Jelölje $T(m) = \max\limits_{F \in \mathcal{F}^{(m)}} |l(F)|$. $T(m)$ becslése előtt szükségünk van két egyszerű technikai lemmára.

6.3 Lemma

$$T(m_1) + T(m_2) + \ldots + T(m_k) \leq T(m_1 + m_2 + \ldots m_k). \tag{6.5}$$

Bizonyítás: Elég belátni, hogy $T(m_1) + T(m_2) \leq T(m_1 + m_2)$. Ha adott két irányított cimkézett fa az U_1 illetve U_2 m_1 illetve m_2 elemű diszjunkt halmazokkal cimkézve a gyökerüknél, melyek rendelkeznek a (6.1)-(6.4) tulajdonságokkal, akkor az a fa, amelyiknek a gyökerének a cimkéje $U_1 \cup U_2$ és a gyökér össze van kötve a két kisebb fa gyökerével szintén rendelkezik a (6.1)-(6.4) tulajdonságokkal. ∎

A következő lemmában a 6.5 Tételben használt függvények néhány tulajdonságát említem meg. Most, és a fejezet további részeben is az e alapú logaritmust log jelöli.

6.4 Lemma *Rögzített* $0 < \varepsilon, \delta < 1$*-re* $x > 1$ *esetén:*

$$f(x) = e^{\log^{1-\delta} x} \text{ szigorúan monoton növő,} \tag{6.6}$$

$$g(x) = x^{\varepsilon + \log^{-\delta} x} \text{ szigorúan monoton növő,} \tag{6.7}$$

$$h(x) = x^{1+\varepsilon+\log^{-\delta} x} \text{ szigorúan monoton növő,} \tag{6.8}$$

$$h(x_1) + h(x_2) + \ldots + h(x_k) \leq h(x_1 + x_2 + \ldots x_k). \tag{6.9}$$

Bizonyítás: $f'(x) = \frac{1-\delta}{x \log^{\delta} x} e^{\log^{1-\delta} x} > 0$ bizonyítja (6.6)-t. $g(x)$ illetve $h(x)$ két szigorúan monoton függvény szorzata. (6.7) alapján

$$h(x_1) + h(x_2) = x_1^{1+\varepsilon+\log^{-\delta} x_1} + x_2^{1+\varepsilon+\log^{-\delta} x_2} \leq x_1(x_1 + x_2)^{\varepsilon+\log^{-\delta}(x_1+x_2)} +$$
$$+ x_2(x_1 + x_2)^{\varepsilon+\log^{-\delta}(x_1+x_2)} = h(x_1 + x_2),$$

ami bizonyítja (6.9)-t. Tehát beláttuk a lemma összes állítását. ∎

6.5 Tétel *Minden* $(\varepsilon_0 \geq)\varepsilon > 0$*-hoz létezik csak* ε*-tól függő* δ*, hogy minden* $m \geq 2$ *egészre*

$$T(m) \leq m^{1+\varepsilon+\log^{-\delta} m}. \tag{6.10}$$

Bizonyítás: Legyen ε rögzített. m-re vonatkozó indukcióval bizonyítunk. Mivel a jobb oldalon m kitevője fix m-re $\delta \to 0$ esetén $(2+\varepsilon)$-hoz tart és nyilván $T(m) \leq \binom{m}{2}$, ezért ha M csak ε -tól függő (elég nagy) konstans (amit később választunk meg), akkor feltehető hogy $m \leq M$ esetén igaz az állítás.

Tegyük most fel, hogy minden $m(> M)$-nél kisebbre igaz az állítás. Tekintsük azt az F fát, amelyre $|l(F)|$ maximális. Legyen $N(g) = \{v_1, v_2, \ldots, v_s, v_{s+1}, \ldots, v_t\}$ valamint $N_1(g) = \{v_1, v_2, \ldots, v_s\}$ és $N_2(g) = \{v_{s+1}, v_{s+2}, \ldots, v_t\}$ az F definiciójában szereplő felbontás. Jelölje $m_i = |A(v_i)|$ valamint F_i a v_i-ből mint gyökérből induló részfát $1 \leq i \leq t$-re. Ez csak úgy lehetséges, ha $T(m_i) = |l(F)|$ minden $1 \leq i \leq t$ -re. Tehát azt elég belátni, hogy

$$\sum_{i=1}^{t} T(m_i) \leq m^{1+\varepsilon+\log^{-\delta} m}. \tag{6.11}$$

Legyen $c = c(\varepsilon)$ olyan nagy, hogy $(1 - \left(\frac{1}{6}\right)^{1/\varepsilon})^{\varepsilon} \leq 1 - \frac{1}{c}$ és $2 < c^{\varepsilon}$ valamint $M = M(\varepsilon)$ olyan nagy, hogy $M^{\varepsilon} > 4c^2 \max_{2 \leq h \leq c^2} T(h)$ Most 4 részre bontjuk az $1 \leq i \leq t$ indexhalmazt:

$$\begin{aligned} P &= \{i|\ m_i \leq c^2\}, \\ Q &= \{i|\ c^2 < m_i \leq \tfrac{m}{c}\}, \\ R &= \{i|\ \tfrac{m}{c} < m_i \leq m(1-\tfrac{1}{c})^{1/\varepsilon}\}, \\ S &= \{i|\ m(1-\tfrac{1}{c})^{1/\varepsilon} < m_i\}. \end{aligned}$$

Vegyük észre, hogy c választása miatt

$$S \subseteq \{i|\ \left(\tfrac{m-m_i}{m}\right)^{\varepsilon} < \tfrac{1}{6}\}. \tag{6.12}$$

1. Eset: $S \neq \emptyset$. Legyen $j \in S$. F definiciójának szimmetriája miatt feltehető, hogy $1 \leq j \leq s$. Ekkor

$$\sum_{i=1}^{t} T(m_i) = T(m_j) + \sum_{\substack{i=1 \\ i \neq j}}^{t} T(m_i) \leq T(m_j) + T(m - m_j) + T(2m - 2m_j), \tag{6.13}$$

felhasználva (6.5)-t, valamint hogy (6.3) miatt $\sum_{\substack{i=1 \\ i \neq j}}^{s} m_i \leq m - m_j$ és hogy (6.2)-(6.4) miatt $\sum_{i=s+1}^{t} m_i \leq m - m_j + (t - s) \leq 2(m - m_j)$. Az indukciós feltevést, valamint (6.7)-t felhasználva:

$$T(m_j) \leq m^{1+\varepsilon+\log^{-\delta} m} - (m - m_j)m^{\varepsilon+\log^{-\delta} m}, \tag{6.14}$$

$$T(m - m_j) \leq (m - m_j)^{1+\varepsilon+\log^{-\delta}(m-m_j)}, \tag{6.15}$$

$$T(2m - 2m_j) \leq (2m - 2m_j)^{1+\varepsilon+\log^{-\delta}(2m-2m_j)}. \tag{6.16}$$

Vegyük észre, hogy

$$\begin{aligned}
&(2m-2m_j)^{1+\varepsilon+\log^{-\delta}(2m-2m_j)} \leq \\
\leq &2^{1+\varepsilon+\log^{-\delta}(m-m_j)} \cdot (m-m_j)^{1+\varepsilon+\log^{-\delta}(m-m_j)} \leq \\
\leq &5 \cdot (m-m_j)^{1+\varepsilon+\log^{-\delta}(m-m_j)},
\end{aligned} \tag{6.17}$$

felhasználva, hogy $m_j \leq m-3$ esetén 2 kitevőjében a harmadik tag 1-gyel becsülhető felülről. Az $m_j = m-1$ vagy $m_j = m-2$ esetek külön ellenőrizhetők (például ekkor $\sum_{i=1}^{t} T(m_i) \leq T(m-1)+3$, ami elég nagy m esetén az indukciót felhasználva kisebb mint a tétel jobboldalán lévő függvény). A fentiek alapján

$$\begin{aligned}
&(m-m_j)^{1+\varepsilon+\log^{-\delta}(m-m_j)} + (2m-2m_j)^{1+\varepsilon+\log^{-\delta}(2m-2m_j)} \leq \\
&\leq 6 \cdot (m-m_j)^{1+\varepsilon+\log^{-\delta}(m-m_j)} = \\
&= (m-m_j) m^{\varepsilon+\log^{-\delta} m} \Big(6 \cdot \big(\tfrac{m-m_j}{m}\big)^{\varepsilon} e^{\log^{1-\delta}(m-m_j) - \log^{1-\delta} m}\Big).
\end{aligned} \tag{6.18}$$

Most (6.12)-(6.18)-t összevetve és felhasználva (6.6)-t kapjuk a (6.11) egyenlőtlenséget.

2. Eset: $S = R = \emptyset$. Ekkor az 1-től t-ig menő összegzés $P \cup Q$ -n megy. Az indukciós feltevés szerint

$$\begin{aligned}
\sum_{i=1}^{t} T(m_i) \leq &\sum_{i \in P \cup Q} {m_i}^{1+\varepsilon+\log^{-\delta} m_i} = \\
= \sum_{i \in P \cup Q} &m_i \big(\tfrac{m_i}{m}\big)^{\varepsilon} m^{\varepsilon+\log^{-\delta} m} \; e^{\log^{1-\delta} m_i - \log^{1-\delta} m}.
\end{aligned} \tag{6.19}$$

Q definiciója és a c-re tett második feltevés miatt minden $1 \leq i \leq t$-re $\big(\frac{m_i}{m}\big)^{\varepsilon} \leq \frac{1}{2}$. (6.6) felhasználásával:

$$\sum_{i \in P \cup Q} m_i \big(\tfrac{m_i}{m}\big)^{\varepsilon} m^{\varepsilon+\log^{-\delta} m} \; e^{\log^{1-\delta} m_i - \log^{1-\delta} m} \leq \frac{1}{2} \sum_{i \in P \cup Q} m_i m^{\varepsilon+\log^{-\delta} m}. \tag{6.20}$$

A nyilvánvaló $\sum_{i=1}^{t} m_i = \sum_{i=1}^{s} m_i + \sum_{i=s+1}^{t} m_i \leq m+m = 2m$ egyenlőtlenségből (alkalmaztuk (6.3)-t) (6.19) és (6.20) felhasználásával adódik a (6.11) egyenlőtlenség.

3. Eset: $S=\emptyset, R\neq\emptyset$. Ekkor $\sum\limits_{i=1}^{t} T(m_i) = \sum\limits_{i\in P} T(m_i) + \sum\limits_{i\in Q} T(m_i) + \sum\limits_{i\in R} T(m_i)$. Továbbá az indukciós feltevés szerint:

$$\begin{aligned}\sum_{i\in R} T(m_i) &\le \sum_{i\in R} {m_i}^{1+\varepsilon+\log^{-\delta} m_i} = \\ =&\frac{1}{2}\sum_{i\in R} m_i m^{\varepsilon+\log^{-\delta} m}\Big(2\cdot\big(\tfrac{m_i}{m}\big)^{\varepsilon} e^{\log^{1-\delta} m_i-\log^{1-\delta} m}\Big).\end{aligned} \tag{6.21}$$

Felhasználva (6.6)-t és R definicióját:

$$\begin{aligned}&\frac{1}{2}\sum_{i\in R} m_i m^{\varepsilon+\log^{-\delta} m}\Big(2\cdot\big(\tfrac{m_i}{m}\big)^{\varepsilon} e^{\log^{1-\delta} m_i-\log^{1-\delta} m}\Big) \le \\ \le&\frac{1}{2}\sum_{i\in R} 2m_i\big(1-\tfrac{1}{c}\big)m^{\varepsilon+\log^{-\delta} m} = \\ =&\frac{1}{2}\Big(\sum_{i\in R}\big(m_i+m_i-\tfrac{m_i}{c}\big)\Big)m^{\varepsilon+\log^{-\delta} m}-\frac{1}{2}\Big(\sum_{i\in R}\tfrac{m_i}{c}\Big)m^{\varepsilon+\log^{-\delta} m}.\end{aligned} \tag{6.22}$$

R definiciója miatt és mivel $R\neq\emptyset$

$$\frac{1}{2}\Big(\sum_{i\in R}\tfrac{m_i}{c}\Big)m^{\varepsilon+\log^{-\delta} m} \ge \frac{1}{2c^2}m^{1+\varepsilon+\log^{-\delta} m}. \tag{6.23}$$

M választása miatt

$$2c^2\sum_{i\in P} T(m_i) < m\cdot M^{\varepsilon} \le m^{1+\varepsilon+\log^{-\delta} m}. \tag{6.24}$$

A Q-ra történő összegzést hasonlóan becsülhetjük, mint a 2. esetben:

$$\begin{aligned}\sum_{i\in Q} T(m_i) \le& \sum_{i\in Q} m_i\big(\tfrac{m_i}{m}\big)^{\varepsilon} m^{\varepsilon+\log^{-\delta} m}\ e^{\log^{1-\delta} m_i-\log^{1-\delta} m} \le \\ &\le \frac{1}{2}\sum_{i\in Q} m_i m^{\varepsilon+\log^{-\delta} m}.\end{aligned} \tag{6.25}$$

Azonban vegyük észre, hogy

$$\sum_{i\in Q} m_i + \sum_{i\in R} (m_i + (m_i - \tfrac{m_i}{c})) \leq 2m. \tag{6.26}$$

Ugyanis egy tetszőleges $i \in R$-hez (feltehető, hogy $1 \leq i \leq s$) tartozó $A(v_i)$ halmazba (6.3) miatt csak az $s+1 \leq j \leq t$-hez tartozó $A(v_j)$ diszjunkt halmazok metszhetnek bele. Mivel minden $j \in Q \cup R$-re $m_j > c^2$, ezért legfeljebb $\frac{m}{c^2}$ darab $A(v_j)$, $j \in Q \cup R$ metszhet bele $A(v_i)$-be, mivel azonban R választása miatt $|A(v_i)| > \frac{m}{c}$ ezért (6.4) miatt $A(v_i)$ elemeinek legfeljebb c-edrésze lehet valamelyik $A(v_j)$, $j \in Q \cup R$ eleme, másszóval legalább $m_i - \frac{m_i}{c}$ elem nem lehet $s+1 \leq j \leq t$, $j \in Q \cup R$-hez tartozó $A(v_j)$-vel lefedve. Ezért

$$\sum_{\substack{i=s+1\\ i\in Q\cup R}}^{t} m_i + \sum_{\substack{i=s+1\\ i\in R}}^{t} \left(m_i - \tfrac{m_i}{c}\right) \leq m \quad \text{és} \quad \sum_{\substack{i=1\\ i\in Q\cup R}}^{s} m_i \leq m.$$

Így ebben az esetben is (6.21)-(6.26) alapján fennáll a (6.11) egyenlőtlenség. Ezzel tehát a tétel bizonyítását befejeztük. ■

6.6 Tétel *Legyen* $\mathcal{K}_n = \{\{1,2\}, \{2,3\}, \ldots, \{n-2,n-1\}, \{n-1,n\}, \{n,1\}\}$. *Ekkor minden* $(\varepsilon_0 \geq)\varepsilon > 0$ *létezik* n_0, *hogy* $n > n_0$ *esetén* $|\mathcal{K}_n{}^{-1}|^{1-\varepsilon} \leq s(\mathcal{K}_n) \leq |\mathcal{K}_n{}^{-1}| + 1$.

Bizonyítás: Legyen ε tetszőleges (elég kis) pozitív szám. A második egyenlőtlenség a 2.17 Lemma miatt igaz.

Határozzuk meg most $\mathcal{K}_n^{-1}$-t. Azt állítom, hogy

$$\mathcal{K}_n{}^{-1} = \{\{a_1, \ldots, a_k\} |\ a_1 < \ldots < a_k,\ 2 \leq a_{i+1} - a_i \leq 3 \\ \text{és } n-3 \leq a_k - a_1 \leq n-2\}. \tag{6.27}$$

Valóban, ezek nyilván antikulcsok, hiszen nem tartalmazzák $\mathcal{K}_n$ egyetlen elemét sem, maximális antikulcsok, hiszen egy ilyennél bővebb már tartalmazna szomszédos elemet. Más nem tartozik hozzájuk, hiszen aminek van két (ciklikusan) szomszédos eleme az kulcs, ami pedig legalább hármat kihagy, az nem lehet maximális, hiszen valamelyik középsőt hozzávéve még mindig antikulcsot kapunk.

A 2.16 Lemma szerint egy $\mathbf{M}$ mátrix reprezentálja $\mathcal{K}_n$ -et pontosan akkor, ha minden $A \in \mathcal{K}_n{}^{-1}$-hez létezik két sor, hogy A-ban azonos elemek állnak, és ha $K \in \mathcal{K}_n$ és két sor megegyezik K-ban, akkor mindenütt. Az utóbbi viszont minimális sorszámú reprezentáció esetén nem lehetséges, hiszen a két egyforma sor közül az egyik elhagyásával egy kevesebb sorral rendelkező mátrixot kapunk, ami szintén reprezentálja $\mathcal{K}_n$-et. Így tehát feltehető, hogy semelyik $K \in \mathcal{K}_n$-re nem létezik $\mathbf{M}$-nek két sora melyek megegyeznek K-ban.

Nézzük meg, hogy milyen "sűrűn" lehetnek a ${\mathcal{K}_n}^{-1}$ -beli elemekhez tartozó sorpárok. Ezek a párok persze különböző antikulcsra különbözőek, mert ha nem, akkor lenne két sor, mely két ${\mathcal{K}_n}^{-1}$-beli unióján megegyezne, ami viszont egy kulcs. Legyen $\mathbf{M}$ egy olyan mátrix ami reprezentálja $\mathcal{K}_n$ -et és $s(\mathcal{K}_n)$ sora van. Rögzítsük $\mathbf{M}$ egy oszlopát, mondjuk az elsőt. Most felépítünk F_i cimkézett irányított fákat (annyit, ahányféle elemből legalább 2 szerepel az első oszlopban).

Legyen $U = \{1, \dots, s(\mathcal{K}_n)\}$ a sorok egy sorszámozása és $U \supseteq \bigcup U_i, \quad |U_i| \geq 2$, ahol U_i az első oszlopban szereplő azon i-edik fajta elemhez tartozó sorok sorszámainak halmaza, amely elemből legalább 2 van. Legyen $|U_i| = m_i$. Az F_i fa g gyökerének cimkéje legyen $A(g) = U_i$. Tegyük fel, hogy már definiáltuk a fa egy v csúcsát és $A(v)$ cimkéjét, és hogy $A(v)$ úgy volt definiálva, hogy valamely j-edik oszlopban bizonyos egyenlő elemekhez tartozó sorok sorszámainak halmaza. Legyen $A(v) \supseteq \bigcup_{i=1}^{s} A_i$, ahol A_i a (ciklikusan) $(j+2)$-edik oszlopban szereplő azon $A(v)$-beli i-edik fajta elemekhez tartozó sorok sorszámainak halmaza, amiből legalább 2 van, míg $A(v) \supseteq \bigcup_{i=s+1}^{t} A_i$, ahol A_i a (ciklikusan) $(j+3)$-adik oszlopban szereplő azon $A(v)$-beli i-edik fajta elemekhez tartozó sorok sorszámainak halmaza, amiből legalább 2 van. Legyen $N(v) = \{v_1, \dots, v_t\}$ és $A(v_i) = A_i \quad (1 \leq i \leq t)$. Vegyük észre még azt, hogy $|A_k \cap A_l| \leq 1 \quad (1 \leq k \leq s,\ 1 \leq l \leq t)$, ugyanis ha legalább kételemű lenne a metszet, akkor lenne két sor mely a $\{j+2, j+3\}$ kulcs oszlopaiban megegyezne, ami viszont a fentiek szerint nem lehet. Így tehát azt is látjuk, hogy az F_i fák rendelkeznek a (6.1)-(6.4) tulajdonságokkal.

A 6.5 Tételt alkalmazva:

$$|l(F_i)| \leq T(m_i) \leq {m_i}^{1+\varepsilon+\log^{-\delta} m_i}. \tag{6.28}$$

Ebből (6.9) alapján

$$\sum_i |l(F_i)| \leq s(\mathcal{K}_n)^{1+\varepsilon+\log^{-\delta} s(\mathcal{K}_n)}. \tag{6.29}$$

$s(\mathcal{K}_n) \to +\infty$ miatt (ami például a korábbi állításokat nem használó (6.36) és (6.32) állítások, valamint a 2.17 Lemma miatt igaz) $s(\mathcal{K}_n) > m_0$-ra (ami $n > n_0$ esetén igaz):

$$s(\mathcal{K}_n)^{1+\varepsilon+\log^{-\delta} s(\mathcal{K}_n)} \leq s(\mathcal{K}_n)^{1+2\varepsilon}. \tag{6.30}$$

Vegyük észre, hogy

$$\sum_i |l(F_i)| \geq |\{K \in \mathcal{K}_n{}^{-1} \text{ melyre } 1 \in K\}|. \tag{6.31}$$

Ugyanis az F_i fák definiciója miatt a jobboldalon szereplő halmaz egy tetszőleges eleméhez (azaz egy 1-et tartalmazó antikulcshoz) tartozó azon sorok (legalább 2) halmaza, melyek megegyeznek ebben az antikulcsban részhalmaza valamely $A(v)$-nek ($v \in \bigcup l(F_i)$), másrészt egy $\bigcup l(F_i)$-beli elemhez nem tartozhat két $\{K \in \mathcal{K}_n{}^{-1}$ melyre $1 \in K\}$-beli elem, ugyanis akkor lenne két sor, mely egy $\mathcal{K}_n{}^{-1}$-belinél bővebb halmazon, azaz egy kulcson megegyezne.

Könnyen láthatóan legalább minden harmadik $\mathcal{K}_n{}^{-1}$-beli antikulcs áthalad 1-n, és így:

$$|\mathcal{K}_n{}^{-1}| \geq |\{K \in \mathcal{K}_n{}^{-1} \text{ melyre } 1 \in K\}| \geq \frac{1}{3}|\mathcal{K}_n{}^{-1}|. \tag{6.32}$$

Mivel $s(\mathcal{K}_n) \to +\infty$, ezért (6.29)-(6.32) alapján $n > n_0$ esetén tényleg

$$|\mathcal{K}_n{}^{-1}| \leq s(\mathcal{K}_n)^{1+3\varepsilon},$$

amiből a tétel következik. ■

6.7 Tétel *Legyen κ a*

$$23x^3 - 23x^2 + 9x - 1 = 0 \tag{6.33}$$

egyenlet egyetlen valós megoldása, azaz

$$\kappa = \frac{1}{3} - \frac{2}{3\sqrt[3]{46}}\left(\sqrt[3]{\sqrt{\tfrac{27}{23}}+1} - \sqrt[3]{\sqrt{\tfrac{27}{23}}-1}\right). \tag{6.34}$$

Ekkor

$$\frac{\log_2 s(\mathcal{K}_n)}{n} \longrightarrow \kappa \log_2 \frac{1-\kappa}{2\kappa} + \frac{1-3\kappa}{2}\log_2 \frac{1-\kappa}{1-3\kappa}. \tag{6.35}$$

Jegyezzük meg, hogy $\kappa \approx 0.1770088227\ldots$, és ha λ jelöli a fenti határértéket, akkor $\lambda \approx 0.4056852313\ldots$

Bizonyítás: Becsüljük meg $|\mathcal{K}_n{}^{-1}|$-t:

$$|\{K \in \mathcal{K}_n{}^{-1} \text{ melyre } 1 \in K\}| = \sum_{\substack{t=0 \\ t \equiv n(2)}}^{\lfloor \frac{n}{3} \rfloor} \binom{\frac{n-t}{2}}{t}. \tag{6.36}$$

Ugyanis legyen $K_1 = \{1, a_1, \ldots, a_k\}$ a baloldalon szereplő halmaz egy eleme, ekkor K_1-hez egyértelműen hozzárendelhető az $\Omega = \{1, \ldots, n\}$ alaphalmaz egy $\{1, \ldots, a_1 - 1\}, \{a_1, \ldots, a_2 - 1\}, \ldots, \{a_k, \ldots, n\}$ 2 és 3 nagyságú blokkokra történő felbontása. Fordítva, ha adott egy ilyen egymást követő 2-es, 3-as blokkokra bontás, akkor a blokkok első elemei meghatároznak egy kívánt antikulcsot. (6.36) jobboldalán éppen az ilyen felbontások száma szerepel (ha t a hármas blokkok száma, akkor $\frac{n-t}{2}$ az összes blokkok száma).

$|\mathcal{K}_n{}^{-1}|$ becsléséhez tehát elég (6.36) jobboldalán a legnagyobb tag nagyságrendjét megbecsülni, ugyanis ekkor (6.32) alapján ez ad egy alsó becslést, ennek $\frac{3(n+2)}{6}$ -szorosa pedig felső becslést.

Ehhez írjuk fel az összeg két egymást követő tagjának hányadosát és nézzük meg, mikor lesz ez 1. Egyszerűsítések után:

$$\frac{(\frac{n-t}{2} - t + 3)(\frac{n-t}{2} - t + 2)(\frac{n-t}{2} - t + 1)}{(\frac{n-t}{2} + 1)t(t-1)} = 1.$$

Ezt rendezve és az $x = \frac{t}{n}$ változót bevezetve:

$$23x^3 - (23 + \tfrac{96}{n})x^2 + (9 + \tfrac{68}{n} + \tfrac{112}{n^2})x - (1 + \tfrac{12}{n} + \tfrac{44}{n^2} + \tfrac{48}{n^3}) = 0. \tag{6.37}$$

Mivel a (6.37) baloldalán álló polinomok függvénysorozata minden véges intervallumban egyenletesen tart a (6.33) baloldalán álló polinomhoz, ezért (6.37)-nek elég nagy n esetén, úgy mint (6.33)-nak, egyetlen valós κ_n megoldása van, és minden $\varepsilon > 0$ -hoz létezik n_0, hogy $n > n_0$ esetén $\kappa - \varepsilon \leq \kappa_n \leq \kappa + \varepsilon$.

Most a fentiek alapján $t = \kappa_n n$-et helyettesítve (6.36)-ba

$$\binom{(\frac{1-\kappa}{2} - \varepsilon)n}{(\kappa - \varepsilon)n} \leq |\mathcal{K}_n^{-1}| \leq \frac{n+2}{2} \binom{(\frac{1-\kappa}{2} + \varepsilon)n}{(\kappa + \varepsilon)n}. \tag{6.38}$$

A Stirling-formulát

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\; e^{\frac{\Theta_n}{12n}} \qquad (0 < \Theta_n < 1)$$

alkalmazva

$$\binom{cn}{dn} = \left(\frac{c}{d}\right)^{dn} \left(\frac{c}{c-d}\right)^{(c-d)n} \sqrt{\frac{c}{2\pi n d(c-d)}}\; e^{\frac{\Theta_n}{12n}\left(\frac{1}{c} - \frac{1}{d} - \frac{1}{c-d}\right)} =$$
$$= 2^{n\left(d \log_2 \frac{c}{d} + (c-d)\log_2 \frac{c}{c-d} + o(1)\right)} \qquad (1 \geq c \geq d > 0).$$

Ezt (6.38)-ba helyettesítve

$$2^{n\left(\kappa \log_2 \frac{1-\kappa}{2\kappa} + \frac{1-3\kappa}{2}\log_2 \frac{1-\kappa}{1-3\kappa} - \varepsilon'\right)} \leq |\mathcal{K}_n{}^{-1}| \leq$$
$$\leq 2^{n\left(\kappa \log_2 \frac{1-\kappa}{2\kappa} + \frac{1-3\kappa}{2}\log_2 \frac{1-\kappa}{1-3\kappa} + \varepsilon'\right)},$$

amiből a 6.6 Tétel alapján a tétel következik. ■

Hivatkozási jegyzék

[1] W.W. Armstrong, Dependency structures of database relationship, *Information Processing 74* (North Holland, Amsterdam, 1974) 580-583.

[2] F.E. Benett, Lisheng Wu, On minimum matrix representation of closure operations, *Discrete Appl. Math.* **26** (1990) 25-40.

[3] B. Bollobás, On generalized graphs, *Acta Math. Hungar.* **16** (1965) 447-452.

[4] G. Burosch, J. Demetrovics, G.O.H. Katona, The poset of closures as a model of changing databases, *Order* **4** (1987) 127-142.

[5] E.F. Codd, A relational model of data for large shared data banks, *Comm. ACM* **13** (1970) 377-387.

[6] J. Demetrovics, On the equivalence of candidate kays with Sperner systems, *Acta Cybernet.* **4** (1979) 247-252.

[7] J. Demetrovics, Z.Füredi, G.O.H. Katona, Minimum matrix representation of closure operations, *Discrete Appl. Math.* **11** (1985) 115-128.

[8] J. Demetrovics, Gy. Gyepesi, A note on minimal matrix representation of closure operations, *Combinatorica* **3** (1983) 177-180.

[9] J. Demetrovics, G. Hencsey, L.O. Libkin, I.B. Muchnik, On the interaction between closure operations and choice functions with applications to relational databases, *Acta Cybernetica* **10** (1992) 129-140.

[10] J. Demetrovics, G.O.H. Katona, Extremal combinatorial problems in relational database, in *Fundamentals of Computation Theory 81, Proc. of the 1981 International FCT-Conference,* Szeged, Hungary, 1981, Lecture Notes in Computer Science 117 (Springer, Berlin 1981) 110-119.

[11] J. Demetrovics, G.O.H. Katona, A.Sali, Design Type Problems Motivated by Database Theory, preprint

[12] J. Demetrovics, Son Hua Nam, Closures and Sperner families, *Bolyai Society Mathematical Studies,* **3** Extremal Problems for Finite Sets (Ed. P. Frankl, Z. Füredi, Gy. Katona, D. Miklós) (Visegrád, Hungary, 1991) 199-204.

[13] P. Erdős, J. Surányi Válogatott fejezetek a számelméletből, *Polygon Könyvtár* 193-195.

[14] Z. Füredi, Perfect error-correcting databases, *Discrete Appl. Math.* **28** (1990) 171-176.

[15] B. Ganter, H.-D.O.F. Gronau, R.C. Mullin, On orthogonal double covers of K_n, *Ars Combinatorica* **37** (1994) 209-221.

[16] H.-D.O.F. Gronau, R.C. Mullin, preprint

[17] G.O.H. Katona, Combinatorial and algebraic results for database relations, *Database Theory - ICDT '92,* Berlin, 1992, Lecture Notes in Comput Science, 646 (Ed. J. Biskup, R. Hull) (Springer Verlag, Berlin, 1992) 1-20.

[18] T. Kővári, V.T. Sós, P. Turán, On a prblem of K. Zarankiewicz, *Colloq. Math.* **3** (1959) 50-57.

[19] D. Lubell, A short proof of Sperner's lemma, *J. Combinat. Theory* **1** (1966) 299.

[20] L.D. Meshalkin, A generalization of Sperner's theorem on the number of subsets of a finite set (in Russian), *Teor. Veroyatnost. i Primenen.* **8** (1963) 219-220.

[21] E. Sperner, Ein Satz über Untermengen einer endlichen Menge, *Math Z.* **27** (1928) 544-548.

[22] K. Yamamoto, Logarithmic order of free distributive lattices, *J. Math. Soc. Japan* **6** (1954) 347-357.