.

Extremal Problems from Coding Perspective (Extremális Problémák Kódolási Nézőpontból)

Habilitációs eljárás tézisei

(E tézisek azonosak a 2009 április 15-én megvédett
fenti című akadémiai doktori értekezés téziseivel)

Ruszinkó Miklós
MTA SZTAKI
Pf. 63, H-1518 Budapest
ruszinko@sztaki.hu

1. Bevezetés

Az értekezés olyan extremális problémákat taglal, amelyek kódolási kérdésekhez kapcsolódnak. A kódok valamilyen adott tulajdonsággal (például páronkénti nagy Hamming távolság) rendelkező n hosszú $0-1$ sorozatok. Ha a sorozatokat egy n elemű alaphalmaz karakterisztikus vektorainak tekintjük, akkor egy halmazrendszert (hipergráfot) kapunk, és a kód tulajdonságai e halmazrendszer tulajdonságaivá fogalmazódnak át. Ilyen módon minden kódolási probléma átfogalmazható hipergráf problémává, és hasonló módon, minden hipergráf probléma átfogalmazható kódolási problémává. Akkor mi a különbség az extremális halmazelmélet és a kódelmélet között? Más a motiváció, a szemlélet. Más kérdéseket tartanak fontosnak a kódkutatók mint az extremális halmazok vizsgálói. Talán a legszembeszökőbb különbség az, hogy az extremális halmazelméletben nagyrészt k-uniform (a halmazrendszer mindegyik tagja k elemű) rendszereket vizsgáltak. Példaként felhozhatnám akár Turán Pál egy híres kérdését: hány hármas adható meg úgy, hogy semelyik négy pont ne tartalmazza mind a négy lehetséges hármast. De ugyanez látszik a Szemerédi Lemma különböző, hipergráfokra vonatkozó verzióiban is: a k halmazméret független az n alaphalmaz mérettől. (Ezt a területet újabban nagyon sokan vizsgálják. A Fields medálos Gowers [30], Tao [52], és sok más kiváló matematikus nagyon szép és fontos eredményeket értek el.) E függetlenség miatt az extremális halmazelméletben általában n-ben *polinomiális* kérdéseket kapnak, hiszen rögzített k esetén $\binom{n}{k} = O(n^k)$ halmaz van. Ezzel szemben, a kódkutatók általában nem tartják fontosnak, hogy egy adott sorozatban pontosan k darab egyes legyen, azaz a halmazok mérete nem konstans. Ezért általában, (mivel mind a 2^n halmaz szóba jöhet) a vizsgált problémák n-ben exponenciális méretűek. Itt a fontosabb extremális kód kérdések szinte kivétel nélkül megoldatlanok. (Alon, Körner és Monti [5] írták, hogy „...all of these problems had one thing in common. Not even the exponential growth rate of the maximum number of n-strings with the required property was known. The breakthrough occurred with cancellative set families when Shearer disproved the corresponding conjecture of Erdős and Katona and this led way to Tolhuizein's beautiful discovery that the Frankl-Füredi bound is tight." Azaz a közel negyven éve vizsgált kancellatív családok esetén a maximális méret exponense a közelmúltban ismertté vált, de más kódok esetén még ezt sem ismerjük. Ezért a jelenlegi állás szerint az a cél, hogy legalább egy durva skálán jó becsléseket kapjunk, azaz becsüljük meg minél élesebben a kitevőt.

Persze az is előfordul, hogy az extremális halmazok vizsgálói exponenciális kérdéseket vizsgálnak, a kódkutatók pedig konstans súlyú kódokat. Ilyenkor bizony előfordul, hogy a két közösség újra felfedezi egymás eredményeit. Erre a jelenségre egy jellemző példa a [48] dolgozatomban

tanulmányozott kérdés, amit nagyon sok szemszögből vizsgáltak.

A kódok kutatása során a két legfontosabb kérdés egyrészt jó tulajdonságokkal rendelkező kódok konstruálása, illetve a maximális méretük meghatározása. A konstrukciók elsősorban lineáris algebrai jellegűek, gyakran használatosak véges testek feletti polinomok, véges geometriák, számelméleti eredmények. Ezek az eszközök általában jól működnek fix, rövid méretű kódok konstruálása során. Azonban az így megadott kódok aszimptotikusan nem viselkednek igazán jól: az ismert konstrukciók szinte kivétel nélkül polinomiális méretűek, még azokban az esetekben is, amikor ismert, hogy létezik exponenciálisan nagy kód. Exponenciálisan nagy alsó határ bizonyítása a fontosabb kódok méretére majdnem mindig valószínűségi módszerekkel történik, ilyen értelemben jó konstrukciók nem ismertek. (Ez alól kivételt képeznek az algebrai geometriai kódok, amelyek nagy kód ABC esetén aszimptotikusan jól viselkednek, viszont a legfontosabb bináris esetben nem érik el a véletlen konstrukciók korlátait. Talán ez az oka, hogy ennek a kezdetben ígéretesnek tűnő módszernek a vizsgálata az utóbbi években alábbhagyott.)

A hagyományosan magyar matematikai szemléletet követve, a disszertáció elsősorban a kódokhoz kapcsolódó aszimptotikus kérdéseket vizsgálja, ezekhez módszereket dolgoz ki.

A vizsgált terület nagyon gazdag, Shannon óta sok kiváló matematikus dolgozott ezeken a kérdéseken. A téma korszerűségét és nagyságát az olyan kiváló szakfolyóiratok is fémjelzik, mint az *IEEE Transactions on Information Theory, Combinatorica, Journal of Combinatorial Theory, Combinatorics, Probability and Computing, Random Structures & Algorithms...*

Az értekezésbe hét angol nyelvű dolgozatom [11, 16, 27, 29, 31, 48, 49] eredményeit válogattam be: igyekeztem úgy összeállítani, hogy megjelenjenek közöttük az elmés, aránylag rövid bizonyítások, amelyekkel régi, nyitott kérdéseket sikerült megválaszolni [16, 27, 29, 48, 49] és a hosszabb, bonyolultabb technikákat alkalmazó [11, 31] gondolatmenetek. A másik szempont az volt, hogy időben aránylag egyenletesen mutassam be munkásságomat, a választott cikkek megjelenési évei 1994 , 1997, 1999, 2001, 2005, 2006, 2007. Azt is fontosnak éreztem, hogy konkrét kód korlátok [16, 27, 29, 48, 49] és használható módszerek [11, 31] kidolgozása egyaránt szerepeljen. Mellékelem társszerzőim egy részének nyilatkozatát arról, hogy a közös dolgozatokhoz lényegesen hozzájárultam.

Az első rész öt dolgozata [16, 27, 29, 48, 49] kódok korlátaival, illetve konstrukciókkal foglalkozik. A második rész (hatodik illetve hetedik cikk, [11, 31]) egy gráf fedési (Hajnal-Szemerédi), illetve Ramsey típusú problémát dolgoz fel. Ezeket az eszközöket kiterjedten használják a kódelméletben. A Hajnal-Szemerédi tételt (és más partíciós tételeket, mint például a Baranyai tételt) jól lehet használni olyankor, amikor kódok méretét becsüljük. Egy friss (idei) példa

erre, hogy Aspnes, Safra és Yin [7] a Hajnal-Szemerédi tételt használják, hogy megoldjanak egy régi, monoton (ranged) hash függvényekre vonatkozó sejtést. A Ramsey elmélet széles körű használatára is sok példát lehetne felhozni: ismert az anti-háromszögmentes gráfok Shannon kapacitása és az $R(3:k)$ (k szín, egyszínű háromszög) Ramsey számok közötti szoros összefüggés [1], de a csatornák uniójának a kapacitására is a becslések Ramsey számokon keresztül adódtak.

2. A fontosabb eredmények ismertetése

Fedés mentes halmazrendszerek ([48] dolgozat)

A szuperponált kódokat Kautz és Singleton [42] vezették be a hatvanas években. Azóta a kérdést a matematika több ágában is vizsgálták, úgymint extremális halmazelméletben (Frankl, Füredi [26]; Erdős, Frankl, Füredi [20, 21]; Füredi [28]; Coppersmith, Shearer [14]), elméleti számítógéptudományban (Hwang, T. Sós [37]; Linial [44]; Szegedy, Vishvanathan [51]; Buhrman, Miltersen, Radhakrishnan, Venkatesh [13]), kódelméletben (Alon, Asodi [2, 3]; Bassalygo, Pinsker [10]; Dyachkov, Rykov [18]), génkutatásban (Csűrös, Milosavljevic [15]; Balding, Torney [8]), és fontos geometriai következményei is vannak (Grünbaum; Erdős, Frankl, Füredi). A teljesség igénye nélkül említének egy geometriai következményt. Grünbaumnak az a sejtése évtizedekig nyitott volt, hogy $\mathbf{R^n}$-ben legfeljebb $2n-1$ pont adható meg úgy, hogy bármely három hegyes szögű háromszöget határoz meg. A szuperponált kódok segítségével ezt exponenciálisan megcáfolták: Erdős, Frankl, Füredi belátták [20], hogy legalább $1,134^n$ ilyen pont van. Számos kapcsolódó cikk született, és a téma szerteágazása folytán előfordult, hogy a matematika valamely ágában dolgozó kutatók újra felfedezték egy másik ágban dolgozó kollégáik eredményeit. A fő kérdés így hangzik:

1. Probléma. *Mekkora $\mathcal{F}$ halmazrendszer választható ki egy n elemű alaphalmazból oly módon, hogy $\forall\ A_0, A_1, \ldots, A_r \in \mathcal{F}$ esetén $A_0 \not\subseteq A_1 \cup A_2 \cup \ldots \cup A_r$?*

Sok egymástól független dolgozatban (Hwang, T. Sós; Kautz, Singleton; Erdős, Frankl, Füredi...) a következő eredmény született.

2. Tétel. *Legyen $f(n,r)$ a fönti tulajdonságnak eleget tevő halmazrendszer maximális mérete. Ekkor alkalmas c_1, c_2 abszolút konstansokkal*

$$2^{c_1 n/r^2} \le f(n,r) \le 2^{c_2 n/r}. \tag{1}$$

Dyachkov és Rykov [18] 1981-ben orosz nyelven publikálták a következő eredményt.

3. Tétel. *(Dyachkov, Rykov 1981)*

$$f(n,r) \leq 2^{c_3 n \log r / r^2}. \tag{2}$$

Látható, hogy a (2) egyenlőtlenség nagyon erős javítása az (1) felső korlátjának. A cikkben leírt bizonyítás viszont meglehetősen hézagos. („Theorem 3 has not been completely proved", írja a hivatalos angol verzió [18]) A [48] cikkem fő eredménye az, hogy adtam egy aránylag egyszerű, tisztán kombinatorikai bizonyítást erre a vitatott eredményre.

4. Tétel. *(Theorem 3.7., [48])*

$$f(n,r) \leq 2^{8n \log r / r^2}. \tag{3}$$

Emellett konstans halmazméret esetén, a Baranyai tételt [9] alkalmazva megjavítottam Dyachkov és Rykov ide vonatkozó korlátait. Erre a dolgozatomra eddig 56 független hivatkozás érkezett, és több könyvben is leközölték. Az eredmény további kutatásokat is inspirált, például Füredi adott később egy még egyszerűbb bizonyítást a tételre, Alon pedig kiterjesztette nagyobb r értékekre és bevezetésre kerültek újabb kódok.

Euklideszi szuperponált kódok ([29] dolgozat)

Az euklideszi szuperponált kódokat (ESZK) 1988-ban Ericson és Györfi [25] vezette be. Valós egységvektoroknak egy $\mathcal{F}$ halmaza $\mathbf{R^n}$-ben ESZK, ha a különböző, legfeljebb r-es vektorösszegek távolsága legalább d, azaz ha $\mathcal{F} \ni \mathcal{A} \neq \mathcal{B} \in \mathcal{F}$ és $|\mathcal{A}| \leq r$, $|\mathcal{B}| \leq r$, akkor $d_E(\sum_{\mathbf{v}\in\mathcal{A}} v,\ \sum_{\mathbf{v}\in\mathcal{B}} v) \geq d$, ahol d_E az euklideszi távolság. (Látható, hogy az ESZK szorosan kapcsolódik egy klasszikus geometriai problémához, nevezetesen nagy ponthalmazok egyenletes elhelyezése a gömbhéjon.) Egy ESZK $f(n,r,d)$ maximális méretére Ericson és Györfi belátták, hogy

5. Tétel. *(Ericson, Györfi 1988)*

$$2^{n \log r/(4r)} \leq f(n,r,d) \leq 2^{n \log r / r}. \tag{4}$$

A két korlát (4)-ben exponenciálisan távol van egymástól, és sok próbálkozás történt erős eszközök (például Levenstein polinomok) bevetésével, hogy ezeket közelebb hozzák egymáshoz. Azonban csak polinomiális javítások születtek és az első exponenciálisan jobb korlátra több, mint tíz éven keresztül várni kellett, amikor is Füredi Zoltánnal a [29] dolgozatban beláttuk, hogy

6. Tétel. *(Theorem 3.2., [29])*

$$f(n,r,d) \leq 2^{n \log r/(2r)}. \tag{5}$$

A bizonyítás egyszerű, lényegében megmutatjuk, hogy az r-es vektorösszegek nagy hányada egy $r^{1/2}$ sugarú gömb belsejében lesz, és utána térfogati becsléseket adunk. Az eredményeinket Milman [46] egy klasszikus tételét alkalmazva sikerült kiterjeszteni tetszőleges n dimenziós normált vektorterekre, azaz ha egy $d_{\mathcal{N}}$ normával vizsgáljuk a fenti kérdést, akkor igaz a következő.

7. Tétel. *(Theorem 4.1., [29])*

$$f_{\mathcal{N}}(n,r,d) = 2^{\Theta(n \log r/r)}. \tag{6}$$

Egy felhasználót azonosító és diszjunkt szuperponált kódok ([16] dolgozat)

A fedés mentes halmazrendszerek (szuperponált kódok) rendelkeznek azzal a Sidon típusú tulajdonsággal, hogy bármely $1 \leq k, \ell \leq r$ értékekre

$$\{A_1, A_2, \ldots, A_k\} \neq \{B_1, B_2, \ldots, B_\ell\}$$

esetén

$$\bigcup_{i=1}^{k} A_i \neq \bigcup_{j=1}^{\ell} B_j.$$

Ezért jól használhatóak azonosításra: a vevő az elküldött halmazok uniójából (bitenkénti vagy) be tudja azonosítani magukat az unióban résztvevő halmazokat. Fölmerült az a kérdés, hogy lehetséges-e nagyobb kódot konstruálni, ha nem az összes az unióban részt vevő halmazt akarjuk meghatározni, de legalább egyet igen. Ezt a tulajdonságot ragadják meg az egy felhasználót azonosító (single user tracing, SUT) halmazrendszerek.

8. Definíció. *Egy $\mathcal{F} \subseteq 2^{[n]}$ halmazrendszer r-SUT, ha minden olyan $\mathcal{F}_1, \ldots, \mathcal{F}_k \subseteq \mathcal{F}$ választás esetén, ahol $1 \leq |\mathcal{F}_i| \leq r$, abból, hogy*

$$\bigcup_{A \in \mathcal{F}_1} A = \bigcup_{A \in \mathcal{F}_2} A = \ldots = \bigcup_{A \in \mathcal{F}_k} A$$

következik, hogy $\cap_{i=1}^{k} \mathcal{F}_i \neq \emptyset$. Azaz (ekvivalens módon) létezik olyan $\phi \colon 2^{[n]} \mapsto \mathcal{F}$ SUT függvény, hogy $\forall \mathcal{F}' \subseteq \mathcal{F}$, amelyre $1 \leq |\mathcal{F}'| \leq r$, $\phi(\cup_{A \in \mathcal{F}'} A) \in \mathcal{F}'$.

Legyen $g(n,r)$ egy r-SUT halmazrendszer maximális mérete.

9. Tétel. *(Theorem 3.4, [16]) Léteznek olyan $c_1, c_2 > 0$ konstansok, hogy*

$$\frac{c_1}{r^2} \leq \limsup_{n\to\infty} \frac{\log g(n,r)}{n} \leq \frac{c_2}{r} \tag{7}$$

A felső korlátot Alon, Fachini és Körner [4] eredményeit felhasználva bizonyítottuk be. Később Alon és Asodi [2] belátták, hogy ez éles. Ugyanezt a problémát általánosítottuk Laczay Bálint szerzőtársammal [43] és erre Alon és Asodi reflektált egy újabb dolgozatban [3].

Egy nagyon régi, máig megoldatlan (Sidon típusú) kérdés a következő. (Weakly union free families, WUF)

10. Probléma. *Mekkora $\mathcal{F}$ halmazrendszer választható ki egy n elemű alaphalmazból oly módon, hogy $A \cup B \neq C \cup D$ bármely* **négy** *különböző $A, B, C, D \in \mathcal{F}$-re.*

A WUF halmazrendszereket általánosítottuk a következő módon.

11. Definíció. *$\mathcal{F} \subseteq 2^{[n]}$ diszjunktan r-szuperponált, ha*

$$\bigcup_{i=1}^{k} A_i \neq \bigcup_{j=1}^{\ell} B_j \tag{8}$$

következik abból, hogy

$$\{A_1, A_2, \ldots, A_k\} \cap \{B_1, B_2, \ldots, B_\ell\} = \emptyset$$

minden $1 \leq k, \ell \leq r$; $A_1, A_2, \ldots, A_k, B_1, B_2, \ldots, B_\ell \in \mathcal{F}$.

Legyen $h(n,r)$ egy ilyen halmazrendszer maximális mérete. Látszólag az ilyen halmazrendszerek nagyon hasonlítanak a szuperponált kódokra, azonban aszimptotikus viselkedésük teljesen más:

12. Tétel. *(Theorem 4.3, [16])*

$$\frac{1}{2r} \leq \limsup_{n\to\infty} \frac{\log h(n,r)}{n} \leq \Big(\frac{1}{2} + o(1)\Big)\frac{\lg r}{r}. \tag{9}$$

Az alsó korlátot véletlen módszerrel, a felsőt pedig második momentum és térfogati meggondolások segítségével láttuk be.

Azonosító kódok ([27] dolgozat)

Az azonosító kódokat Karpovsky, Chakrabarty, Levitin [40] vezették be és az elmúlt tíz évben több, mint száz cikk jelent meg ebben a kérdésben. Egy $G = (V, E)$ gráfban legyen

$N[v] = N(v) \cup \{v\}$, a v pont zárt szomszédsága. Egy fix $C \subseteq V$ ponthalmazra és minden legfeljebb ℓ elemű $X \subseteq V$ ponthalmazra legyen

$$I(X, C) := \bigcup_{x \in X} N[x] \cap C.$$

Ha minden $I(X, C)$ különböző, akkor azt mondjuk, hogy C szeparálja, és ha mindegyik $I(X, C) \neq \emptyset$, akkor C lefedi a G gráf legfeljebb ℓ elemű ponthalmazait. Azt mondjuk, hogy C egy kód, mely azonosítja a G gráf legfeljebb ℓ elemű részhalmazait (ℓ-ID kód), ha mindkét fenti tulajdonság teljesül. (Az 1-ID kódokat egyszerűen ID kódoknak fogjuk nevezni.)

Az ID kódok vizsgálata a számítógéphálózatok teszteléséből ered. Ha a G gráf minden pontja egy processzor, és ezek az E élek mentén vannak összekötve, akkor egy C ℓ-ID kód a következőt teszi lehetővé. Ha minden C-beli processzor lefuttat egy teszt programot önmagán és a szomszédain és legfeljebb ℓ helyről jön hibajelzés, akkor a hibás processzorok X halmazát meg lehet állapítani.

Eredendően azt a két kérdést vizsgálták ebben a témában, hogy milyen gráfoknak van ID kódjuk, illetve speciális gráfokban (például rács) mi a legkisebb elemszámú C ID kód. Ezeket a kérdéseket az Erdős-Rényi féle $G_{n,p}$ véletlen gráf modellben vizsgáltuk. Legyen $q = p^2 + (1-p)^2$. A következő tétel a véletlen gráfok minimális ID kódjának a $c(G_{n,p})$ méretét határozza meg majdnem biztosan.

13. Tétel. *(Theorem 1, [27]) Legyen* $p, (1-p) \geq 4 \log\log n / \log n$. *Ekkor* $\mathcal{G}(n,p)$ *majdnem minden gráfjára* $c(G_{n,p}) \sim \frac{2\log n}{\log(1/q)}$, *azaz, bármely* $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\left(\left| c(G_{n,p}) \cdot \left(\frac{2\log n}{\log(1/q)} \right)^{-1} - 1 \right| \geq \epsilon \right) = 0.$$

14. Következmény. *(Corollary 1, [27])* $c(G_{n,1/2}) \sim 2\log_2 n$, *majdnem biztosan.*

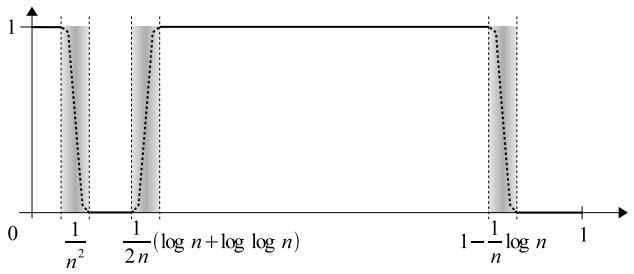
Pontosan meghatároztuk, hogy milyen $p = p(n)$ valószínűségek esetén van (majdnem biztosan) a $G_{n,p}$ véletlen gráfnak ID kódja.

15. Tétel. *(Theorem 5, [27]) Bármely* $\epsilon > 0$ *esetén:*

- *ha* $p = o(n^{-2})$, *akkor* $\mathcal{G}(n,p)$*-ben majdnem minden gráfnak van ID kódja (és majdnem biztosan ez egyértelműen az egész ponthalmaz),*
- *ha* $pn^2 \to +\infty$ *és* $p \leq \frac{1}{2n}(\log n + (1-\epsilon)\log\log n)$, *akkor* $\mathcal{G}(n,p)$*-ben majdnem minden gráfnak nincs ID kódja,*

- *ha* $\frac{1}{2n}\left(\log n + (1+\epsilon)\log\log n\right) \leq p \leq 1 - \frac{1}{n}\left(\log n + \epsilon \log\log n\right)$, *akkor* $\mathcal{G}(n,p)$*-ben majdnem minden gráfnak van ID kódja,*

- *ha* $p \geq 1 - \frac{1}{n}\left(\log n - \epsilon \log\log n\right)$, *akkor* $\mathcal{G}(n,p)$*-ben majdnem minden gráfnak nincs ID kódja.*



1. ábra. Az ID kód létezésének a küszöb valószínűségei. A függőleges tengelyen a $\Pr(G_{n,p}$-nek van ID kódja) valószínűségek aszimptotikus értékei, a vízszintes tengelyen a $p(n)$ értékek vannak.

Erdős és Rényi tételeit használva a Theorem 6,7,8 tételekben ([27]) egészen pontosan leírtuk, hogyan viselkednek a valószínűségek a küszöböknél, azaz, amikor p az 1. ábra valamelyik leárnyékolt részében van.

Az ℓ-ID kódok esetében beláttuk a következőt.

16. Tétel. *(Theorem 9, [27]) Legyen* ϵ *olyan, hogy* $n^\epsilon \to +\infty$, p *konstans,* $p \neq 0, 1$. *Ekkor* $\mathcal{G}(n,p)$*-ben majdnem minden gráfnak létezik*

$$|C| \leq \frac{2(\ell+\epsilon)\log n}{\log(1/q_\ell)}$$

méretű C ℓ*-ID kódja, ahol* $q_\ell = 1 - \min\{p, 2p(1-p)\}(1-p)^{\ell-1}$.

Egy optimális kód véletlen hozzáféréshez ([49] dolgozat)

Az úgynevezett többszörös felhasználású ütközéses csatornák vizsgálata szintén a hatvanas években kezdődött: szép eredményeket értek el Pippenger; Abramson; A, Györfi, Massey; Tsybakov; Capetanakis, hogy csak néhányat említsek. A témakör kérdései szorosan kapcsolódnak Erdős, Rényi és Lindström számelméleti eredményeihez. A modell a következő. Korlátlan sok, egymással nem kommunikáló felhasználó „csomagokat" küldhet egy közös csatornán a $[t, t+1)$ $(t = 0, 1, 2 \ldots)$ időintervallumokban egy központba. Egy $[t, t+1)$-ben küldött csomag pontosan

akkor jut el a központba, ha $[t, t+1)$-ben más csomagot nem küldtek. Ha mégis, akkor a csomagok ütköznek és mindegyik elvész. Mindegyik felhasználó még a $t+1$-edik időpillanat előtt megtudja (visszacsatolás), hogy $[t, t+1)$-ben hány csomag küldésével próbálkoztak. Ha ez a szám 0, akkor $[t, t+1)$-ben senki sem küldött, ha 1, akkor egyvalaki küldött és a csomag sikeresen átért, ha $k \geq 2$, akkor többen próbálkoztak, és minden küldemény elveszett. (A vizsgált csatornamodellek abban különböznek egymástól, hogy mi a visszacsatolás. Így egy másik intenzíven tanulmányozott eset abban különbözik a fentitől, hogy ha $[t, t+1)$-ben legalább két csomagot küldtek, akkor a visszacsatolás csak egy (ütközés) szimbólum.)

A csomagok λ paraméteres Poisson folyamat szerint generálódnak. Pontosabban, legyen $x_1 \leq x_2 \leq x_3 \ldots$ egy véletlen folyamat, ahol x_i az i-edik csomag keletkezésének a (véletlen) időpontja. Ekkor az $(x_{i+1} - x_i)$ különbségek független valószínűségi változók, minden i-re azonosan

$$\mathbf{P}\{x_{i+1} - x_i > x\} = e^{-\lambda x}$$

eloszlással. Egy csomag δ késésén a keletkezése és a sikeres átküldése közötti időt értjük. Egy adott f protokoll késése

$$D_f^{(\lambda)} := \limsup_{i \to \infty} \mathbf{E}(\delta_i), \tag{10}$$

ahol $\mathbf{E}()$ a várható érték és hatékonysága (*throughput*)

$$R_f = \sup\{\lambda : D_f^{(\lambda)} < \infty\}. \tag{11}$$

A legfontosabb kérdés itt az, hogy mi a λ-k $\mathcal{A}$ protokollokon vett szuprémuma (csatorna kapacitás),

$$C = \sup\{R_f : \ f \in \mathcal{A}\}, \tag{12}$$

amelyre minden generálódott csomag véges várható késéssel eljut a központba. Vegyük észre, hogy ha $\lambda = 1 + \varepsilon > 1$, akkor a $[0, t)$-ben várhatóan $(1+\varepsilon)t$ csomag keletkezik, így még ha mindegyik időrést sikeresen ki is használjuk, εt csomag nem kerül küldésre. Ezért a várható késés t növekedésével minden korláton túl nő, azaz a kapacitás, $C \leq 1$. Tsybakov 1980-ban [54] belátta, hogy $C \geq 0.533$. Meglepő eredményként, 1981-ben Pippenger [47] belátta, hogy ebben a modellben a kapacitás, $C = 1$. A bizonyításában megad egy protokollt, amely Sidon típusú mátrixok létezésén alapszik. Ilyen mátrixok konstruálására azonban eljárás máig sem ismert. (Ezek a mátrixok Erdős és Rényi, illetve Lindström „coin weighing" eredményeinek az erősítését foglalják magukban.) Egy konkrét protokoll megadásához a kódkutatók Pippenger bizonyításában az egyetlen véletlen elem, a mátrix megkonstuálásával próbálkoztak. Ezek nem

vezettek eredményre, és ez a központi probléma (nevezetesen egy konkrét protokoll megadása) több mint másfél évtizeden keresztül nyitott volt. A [49] cikkben ezt a problémát oldjuk meg Lindström számelméleti eredményeit felhasználva.

17. Tétel. *(Theorem 5.1., [49]) A [49] dolgozatban ismertetett f protokoll hatékonysága $R_f = 1$, azaz f optimális.*

A Hajnal-Szemerédi tétel véletlen gráfokban ([11] dolgozat)

A klasszikus Hajnal-Szemerédi tétel [35] azt mondja ki, hogy ha egy n pontú gráfban a minimális fok legalább $(1-1/t)n$, akkor a gráf tartalmaz $\lfloor n/t \rfloor$ pontdiszjunkt K_t klikket. (K_t faktort, oszthatóságot feltételezve.) Az Erdős-Rényi féle $G_{n,m}$ (m az élszám) véletlen gráf modellben a K_t faktor létezésének a küszöb élszámát Erdős és Rényi kezdték el vizsgálni és belátták, hogy $t = 2$ esetén ez $m = n \log n/2 + cn$. A $t = 3$ esetről Erdős a következőket írta ([6], Appendix B): „How many edges are needed in a graph of $3n$ vertices to be able to cover the vertices by n vertex disjoint triangles? The correct answer will be probably $n^{4/3}$, but perhaps a little more ... the lack of analogs of Tutte's theorem may cause a serious trouble." A jóslat nagyon pontosnak bizonyult, Joel Spencer néhány éve ezt a kérdést egy, a Carnegie Mellon egyetemen tartott előadásában, a véletlen gráfok egyik legnehezebb nyitott kérdésének nevezte. Sőt, aránylag egyszerűen belátható, hogy ahhoz, hogy mindegyik pont szerepeljen legalább egy K_t-ben, szükséges (nagyságrendileg) $\binom{n}{2}(\log n)^{1/\binom{t}{2}} n^{-2/t}$ véletlen él. (A K_3 esetre ez $n^{4/3}(\log n)^{1/3}$, azaz a 'perhaps a little more' elenyésző, $(\log n)^{1/3}$.) A kérdést nagyon sok élenjáró matematikus vizsgálta, így Alon, Yuster; Krivelevich; Ruciński... Alon és Yuster belátták, hogy K_3 faktor létezéséhez elegendő $O((n^3 \log n)^{1/2})$ él. Krivelevich ezt a becslést $n^{7/5}$-re javította, és általánosabban bebizonyította, hogy a $p = O(n^{-2t/(t-1)(t+2)})$ küszöb valószínűség elegendő K_t faktorhoz. Az általánosan elfogadott, nagyon régóta nyitott sejtés az, hogy a szükséges feltétel adja ki az igazságot, azaz $O(\binom{n}{2}(\log n)^{1/\binom{t}{2}} n^{-2/t})$ véletlen él elegendő K_t faktorhoz. Ezt a sejtést sikerült belátni a [11] dolgozatban egy aszimptotikusan elenyésző hibataggal: ha olyan K_t *pontfedést* nézünk, ahol $o(n)$ pontra megengedjük, hogy legalább két K_t-ben szerepeljen ($(K_t, 2)$-fedés, majdnem partíció!), akkor az állítás igaz.

18. Tétel. *(Theorem 1, [11]) Legyen $m = \binom{n}{2}((t-1)!(\log n + c_n))^{1/\binom{t}{2}} n^{-2/t}$. Ekkor*

$$\lim_{n\to\infty} \Pr(G_{n,m} \text{ tartalmaz } (K_t,2)\text{-fedést}) = \begin{cases} 0 & c_n \to -\infty \\ e^{-e^{-c}} & c_n \to c \\ 1 & c_n \to \infty \end{cases} \tag{13}$$

A tételt egy kicsit általánosabb formában láttuk be és ebből következik, hogy a kétszer fedett pontok száma $o(n)$. A több mint 20 oldalas bizonyítás sok technikai elemet tartalmaz. A dolgozat fő eredményén kívül még a következő módon általánosítjuk a Hajnal-Szemerédi tételt. Egy $G = (V, E)$ gráf rendelkezik a $\mathcal{C}_{H,k}$ tulajdonsággal, ha léteznek olyan $H_1, \ldots, H_r$ részgráfjai, amik eleget tesznek a következő tulajdonságoknak.

P1. $H_i \cong H, \quad$ minden $i = 1, \ldots, r,$

P2. $\cup_{i=1}^{r} V(H_i) = V,$

P3. $E(H_i) \cap E(H_j) = \emptyset, \quad$ minden $i \neq j$, és

P4. $\forall\, v \in V$ legfeljebb k darab H_i-ben van benne.

Speciálisan, egy $G = (V, E)$ gráf $\mathcal{C}_{t,k}$ tulajdonságú, ha a pontjai lefedhetők K_t klikkekkel oly módon, hogy minden pontot legfeljebb k-szor fedünk. Legyen

$$f(n, t, k) = \max\{d : \exists G, \text{ hogy } \delta(G) = d, |V(G)| = n, \text{ és } G \notin \mathcal{C}_{t,k}\},$$

azaz a minimális $f(n, t, k)+1$ fokszám már garantálja a $\mathcal{C}_{t,k}$ tulajdonságot. (Vegyük észre, hogy ebben az összefüggésben a Hajnal-Szemerédi tétel az $f(n, t, 1) + 1$ meghatározása.) Ekkor igaz a következő.

19. Tétel. *(Theorem 4, [11]) Legyen $t \geq 3$, $k \geq 2$, $n \geq 6t^2 - 4t$ és*

$$n = q[(t-1)k + 1] + r \quad \text{ahol} \quad 1 \leq r \leq (t-1)k + 1.$$

Ekkor

$$n - qk - \left\lceil \frac{r}{t-1} \right\rceil \leq f(n, t, k) \leq n - qk - \left\lceil \frac{r}{t-1} \right\rceil + 1\ .$$

A 19. Tételből következik, hogy

$$f(n, t, k) = \left\lfloor \frac{[(t-2)k + 1]n}{(t-1)k + 1} \right\rfloor + c, \tag{14}$$

ahol $c \in \{0, 1, 2\}$. A háromszög fedések esetét sikerült teljesen pontosan megoldani:

20. Tétel. *(Theorem 5, [11]) Legyen $n \geq 6$ és $k \geq (n-1)/2$. Ekkor*

$$f(n, 3, k) = \left\lceil \frac{n}{2} \right\rceil.$$

Ugyanebben a cikkben belátjuk, hogy éldiszjunkt pontfedést már a Komlós-Sárközy-Szemerédi [41] tételben megfogalmazottnál kisebb minimális fokszám is biztosít abban az esetben, ha H csúcs-színkritikus. (Theorem 6, 62. oldal).

Egy Ramsey típusú eredmény ([31] dolgozat)

A nyolcvanas évek végén fogalmazta meg Gyárfás ($r = 2$ esetben Lehel) a következő Ramsey típusú sejtést. Akárhogy színezzük K_n éleit r színnel, a ponthalmaz partícionálható legfeljebb r darab pontdiszjunkt egyszínű körre. (Konstrukció mutatja, hogy r körre szükség van.) A sejtés attól izgalmas, hogy az osztályok száma független a pontszámtól, csak a színek számától függ! Már az $r = 2$ eset sem egyszerű, Luczak, Rödl és Szemerédi [45] bizonyították be 14 oldalon komoly technikai eszközök bevetésével. Erdős, Gyárfás és Pyber [22] belátták, hogy $cr^2 \log r$ körre a ponthalmaz mindig felbontható. A [31] dolgozatban ezt a becslést sikerült jelentősen megjavítani, így, most már az alsó és felső korlát között csak egy $\log r$-es faktor maradt.

21. Tétel. *(Theorem 1, [31]) Minden $r \geq 2$ egészhez létezik olyan $n_0 = n_0(r)$, hogy ha $n \geq n_0$ és K_n éleit színezzük r színnel, akkor K_n pontjai felbonthatók $100r \log r$ egyszínű pontdiszjunkt körre.*

A majdnem húsz oldalas bizonnyítás során használjuk a Szemerédi Lemmát, a Luczak által bevezetett egyszínű összefüggő párosítások módszerét[1], Madernek egy tételét és valószínűségi (nagy eltérések) meggondolásokat. Lényegében a redukált gráfban keresünk egy nagy egyszínű összefüggő párosítást, aminek a pontjait kivesszük a gráfból. Ezután a maradék ponthalmazból „mohón" választunk ki egyszínű köröket. Azokat a pontokat, amiket nem fedtünk le a „mohó" körökkel, a nagy egyszínű összefüggő párosítás és e pontok közötti páros gráf éleit használva fűzzük fel körökre. A bizonyításnak több nehéz pontja van. Az egyik, hogy az eljárás során a redukált gráfban talált nagy egyszínű összefüggő párosítás sérül, és ezt rendbehozni nem egyszerű. Ugyanakkor újabb, páros gráfok színezésére vonatkozó Ramsey típusú tételek bizonyítására is szükség volt. Ezeket összeszedve és kiegészítve egy következő dolgozatban [32] is publikáltuk.

3. Köszönetnyilvánítás

Köszönöm kollégáimnak a sok felvetett szép problémát. Noga Alon, Alan Frieze, Füredi Zoltán, Gyárfás András, Györfi László, Simonovits Miklós, T. Sós Vera, Szemerédi Endre, Tardos Gábor (és még sokan mások) mindig szivesen tanítottak, bíztattak és támogattak; köszönöm mindnyájuknak.

[1] Ezt a módszert a [32, 33, 34] dolgozatokban is sikeresen alkalmaztuk

Hivatkozások

[1] N. Alon, Graph Powers, *Contemporary Combinatorics,* (B. Bollobas, ed.), Bolyai Society Mathematical Studies, Springer 2002, 11-28.

[2] N. Alon, V. Asodi, Tracing a single user, *European Journal of Combinatorics,* 27(8) (2006), 1227-1234.

[3] N. Alon, V. Asodi, Tracing many users with almost no rate penalty, *IEEE Transactions on Information Theory,* 53(1) (2007), 437-439.

[4] N. Alon, E. Fachini, J. Körner, Locally thin set families, *Combinatorics, Probability and Computing*, 9 (2000), 481-488.

[5] N. Alon, J. Körner, A. Monti, String quartets in binary, *Combinatorics, Probability and Computing*, 9 (2000), 381-390.

[6] N. Alon, J. H. Spencer, *The probabilistic method,* Wiley-Interscience [John Wiley & Sons] (2000).

[7] J. Aspnes, M. Safra, Y. Yin, Ranged Hash Functions and the Price of Churn, *Proceedings of SODA 2008*, 1066-1075.

[8] D.J. Balding, D.C. Torney, Optimal pooling designs with error detection, *Journal of Combinatorial Theory, Series A,* 74(1), 1996, 131-140.

[9] Zs. Baranyai, On the factorization of the complete uniform hypergraph, *Infinite and Finite Sets,* (A. Hajnal, R. Rado and V. T. Sós, eds.), North-Holland, Amsterdam, 91-108.

[10] L.A. Bassalygo and M.S. Pinsker, Limited multiple-access of a nonsynchronous channel (in Russian), *Problems of Information Transmission,* 19 (1983), 92-96.

[11] T. Bohman, A. Frieze, M. Ruszinkó, L. Thoma, Vertex covers by edge disjoint cliques, *Combinatorica*, Vol. 21(2) (2001), 171-197.

[12] B. Bollobás, *Random Graphs,* Cambridge University Press (2001).

[13] H. Buhrman, P.B. Miltersen, J. Radhakrishnan, S. Venkatesh, Are bitvectors optimal? *Proceedings of STOC'00,* pp. 449-458, (A teljes változat: *Siam Journal on Computing,* 31(6) (2002), 1723-1744.)

[14] D. Coppersmith, J.B. Shearer, New bounds for union-free families of sets, *Electronic Journal of Combinatatorics,* 5 (1998), #R39.

[15] M. Csűrös, A. Milosavljevic, A Pooled Genomic Indexing (PGI): Analysis and design of experiments, *Journal of Computational Biology,* 11(5), (2004), 1001-1021.

[16] M. Csűrös, M. Ruszinkó, Single-user tracing and disjointly superimposed codes, *IEEE Transactions on Information Theory,* 51(4) (2005), 1606-1611.

[17] D. Danev, Some constructions of superimposed codes in Euclidean spaces, *Discrete Applied Mathematics*, 128(1) (2003), 85-101.

[18] A. G. Dyachkov and V.V. Rykov, Bounds on the length of disjunctive codes, *Problemy Peredaci Informacii,* 18(3), (1982), 7-13. (orosz nyelven). Angol fordításban: 1983 Plenum Publishing Corporation, 166-171.

[19] P. Erdős, L. Lovász, *Problems and results on 3-chromatic hypergraphs and some related questions,* in Infinite and Finite Sets (to Paul Erdős on his 60th birthday), 609-627, North-Holland, Amsterdam (1975).

[20] P. Erdős, P. Frankl and Z. Füredi, Families of finite sets in which no set is covered by the union of two others, *Journal of Combinatorial Theory, Series A*, 33(2) (1982), 158-166.

[21] P. Erdős, P. Frankl and Z, Füredi, Families of finite sets in which no set is covered by the union of r others, *Israel Journal of Mathematics,* 51(1-2), 79-89.

[22] P. Erdős, A. Gyárfás, and L. Pyber, Vertex coverings by monochromatic cycles and trees, *Journal of Combinatorial Theory, Series B*, 51 (1991), 90-95.

[23] P. Erdős, A. Rényi, On the evolution of random graphs, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5 (1960), 17-61.

[24] P. Erdős, A. Rényi, On the evolution of random graphs, *Bulletin de l'Institut I nternational de Statistique,* 38(4) (1961), 343-347.

[25] T. Ericson and L. Györfi, Superimposed Codes in R^n, *IEEE Transactions on Information Theory,* 34(4) (1988), 877-880.

[26] P. Frankl and Z. Furedi, Union-free Hypergraphs and Probability Theory, *European Journal of Combinatorics,* 5 (1984), 127-131.

[27] A. Frieze, R. Martin, J. Moncel, M. Ruszinkó, C. Smyth, Codes identifying sets of vertices in random networks, *Discrete Mathematics,* 307(9-10) (2007), 1094-1107.

[28] Z. Füredi, On r-cover-free families, *Journal of Combinatorial Theory, Series A,* 73(1) (1996), 172–173.

[29] Z. Füredi, M. Ruszinkó, An improved upper bound of the rate of Eucledian superimposed codes, *IEEE Transactions on Information Theory,* 45(2) (1999), 799-802.

[30] Gowers, W. T. Quasirandomness, counting and regularity for 3-uniform hypergraphs, *Combinatorics, Probability and Computing*, 15(1-2) (2006), 143-184.

[31] A. Gyárfás, M. Ruszinkó, G. Sárközy, E. Szemerédi, An improved bound for the monochromatic cycle partition number, *Journal of Combinatorial Theory, Series B,* 96(6) (2006), 855-873.

[32] A. Gyárfás, M. Ruszinkó, G. Sárközy, E. Szemerédi, One-sided coverings of colored complete bipartite graphs, *Algorithms and Combinatorics, Topics in Discrete Mathematics*, Volume Dedicated to Jarik Nesetril on the Occasion of his 60th Birthday, (2006), ISBN-10 3-540-33698-2, pp. 133-144.

[33] A. Gyárfás, M. Ruszinkó, G. Sárközy, E. Szemerédi, Three-color Ramsey numbers for paths, *Combinatorica*, 27(1) (2007), 35-69.

[34] A. Gyárfás, M. Ruszinkó, G. Sárközy, E. Szemerédi, Tripartite Ramsey numbers for paths, *Journal of Graph Theory*, 55(2) (2007), 164-174.

[35] A. Hajnal, E. Szemerédi, Proof of a conjecture of Erdős, *Comb. Theory and Appl. II,* Colloq. Math. Soc. J. Bolyai 4, North-Holland, 1970, 601-623.

[36] I. Honkala, T. Laihonen, S. Ranto, On strongly identifying codes, *Discrete Mathematics,* 254 (2002), 191-205.

[37] F. K. Hwang and V.T. Sós, Non adaptive hypergeometric group testing, *Studia Sci. Math. Hungar.*, 22 (1987), 257-263.

[38] S. Janson, Poisson approximation for large deviations, *Random Structures and Algorithms,* 1 (1990), 221-229.

[39] S. Janson, New versions of Suen's correlation inequality, *Random Structures Algorithms* 13(3-4) (1998), 467-483.

[40] M. G. Karpovsky, K. Chakrabarty, L. B. Levitin, On a new class of codes for identifying vertices in graphs, *IEEE Transactions on Information Theory,* 44(2) (1998), 599-611.

[41] J. Komlós, G. N. Sárközy, E. Szemerédi, Proof of the Alon - Yuster conjecture, *Discrete Mathematics,* 235(1) (2001), 255-269.

[42] W. H. Kautz, R. R. Singleton, Nonrandom binary superimposed codes, *IEEE Transformations on Information Theory,* 10(4) (1964), 363-377.

[43] B. Laczay, M. Ruszinkó, Multiple User Tracing Codes, *2006 IEEE International Symposium on Information Theory,* July 9-14, Seattle, USA, ISBN 1-4244-0367-7, pp. 1900-1904.

[44] N. Linial, Locality in Distributed Graph Algorithms, *SIAM Journal on Computing,* 21(1) (1992), 193-201.

[45] T. Luczak, V. Rödl, E. Szemerédi, Partitioning two-colored complete graphs into two monochromatic cycles, *Probability, Combinatorics & Computing*, 7 (1998), 423-436.

[46] V. D. Milman, Almost Euclidean quotient spaces of subspaces of finite dimensional normed spaces, *Proceedings of the American Mathematical Society,* 94 (1985), 445-449.

[47] N. Pippenger, Bounds on the performance of protocols for a multiple access broadcast channel, *IEEE Transactions on Information Theory*, 27(2) (1981), 145-151.

[48] M. Ruszinkó, On the upper bound of the size of the r-cover-free families, *Journal of Combinatorial Theory, Series A,* 66(2) (1994), 302-310.

[49] M. Ruszinkó, P. Vanroose, How an Erdős-Rényi type search approach gives an explicit code construction of rate 1 for random access with multiplicity feedback, *IEEE Transactions on Information Theory,* 43(1) (1997), 368-373.

[50] W.-C. S. Suen, A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph, *Random Structures and Algorithms* 1(2) (1990), 231-242.

[51] M. Szegedy, S. Vishwanathan: Locality based graph coloring, *STOC,* 1993, 201-207

[52] T. Tao, A variant of the hypergraph removal lemma. *Journal of Combinatorial Theory, Series A,* 113(7) (2006), 1257-1280.

[53] Ludo M. G. M. Tolhuizen, New rate pairs in the zero-error capacity region of the binary multiplying channel without feedback, *IEEE Transactions on Information Theory,* 46(3) (2000), 1043-1046.

[54] B. S. Tsybakov, Resolution of a conflict of known multiplicity, *Problemy Peredachi Informatsii*, 16(2) (1980), 69-82.

Az értekezés a következő dolgozatokon alapul

[48] M. Ruszinkó, On the upper bound of the size of the r-cover-free families, *Journal of Combinatorial Theory, Series A,* 66(2) (1994), 302-310.

[29] Z. Füredi, M. Ruszinkó, An improved upper bound of the rate of Eucledian superimposed codes, *IEEE Transactions on Information Theory,* 45(2) (1999), 799-802.

[16] M. Csűrös, M. Ruszinkó, Single-user tracing and disjointly superimposed codes, *IEEE Transactions on Information Theory,* 51(4) (2005), 1606-1611.

[27] A. Frieze, R. Martin, J. Moncel, M. Ruszinkó, C. Smyth, Codes identifying sets of vertices in random networks, *Discrete Mathematics,* 307(9-10) (2007).

[49] M. Ruszinkó, P. Vanroose, How an Erdős-Rényi type search approach gives an explicit code construction of rate 1 for random access with multiplicity feedback, *IEEE Transactions on Information Theory,* 43(1) (1997), 368-373.

[11] T. Bohman, A. Frieze, M. Ruszinkó, L. Thoma, Vertex covers by edge disjoint cliques, *Combinatorica*, 21(2) (2001), 171-197.

[31] A. Gyárfás, M. Ruszinkó, G. Sárközy, E. Szemerédi, An improved bound for the monochromatic cycle partition number, *Journal of Combinatorial Theory, Series B,* 96(6) (2006), 855-873.