

Samoopravné kódy

Alexandr Kazda

Univerzita Karlova

15. dubna 2020

- Generující matice

$$M = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ I_{12} & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- Matice D začíná za sloupcem ze skoro samých jedniček

Proč D nemá řádek váhy 10 ani dva řádky s průnikem 1

$$M = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 1 & \dots & 1 \\ 0 & 1 & 0 & \dots & 0 & 1 & & & \\ 0 & 0 & 1 & \dots & 0 & 1 & & D & \\ & & & \ddots & 0 & 1 & & & \\ 0 & 0 & 0 & \dots & 1 & 1 & & & \end{pmatrix}$$

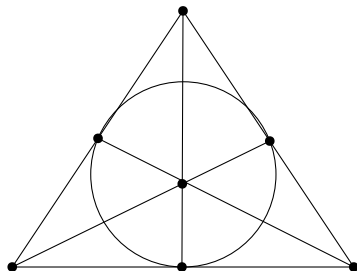
- Značme m_1, m_2, \dots řádky M
- Lepší argument (díky za návrh!): Necht' třeba ~~druhý~~**první** řádek D má váhu 10...
- pak $m_1 + m_2 = (1, 1, 0, \dots, 0, 1, [10 \text{ nul a } 1 \text{ jednička}])$ má váhu 4, spor
- Necht' součet ~~druého a třetího~~**prvního a druhého** řádku D má váhu 10...
- pak $m_1 + m_2 + m_3 = (1, 1, 1, 0, \dots, 0, 0, [10 \text{ nul a } 1 \text{ jednička}])$ má váhu 4, spor

$$M = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 1 & \dots & 1 \\ 0 & 1 & 0 & \dots & 0 & 1 & & & \\ 0 & 0 & 1 & \dots & 0 & 1 & & D & \\ & & & \ddots & 0 & 1 & & & \\ 0 & 0 & 0 & \dots & 1 & 1 & & & \end{pmatrix}$$

- D je 11×11 matice, jejíž každý řádek má 6 jedniček a dva řádky mají „průnik“ přesně 3
- Sloupce D : **vrcholy**, řádky D : **bloky** $B_1, \dots, B_{11} \subset \{1, 2, \dots, 11\}$
- Minule bylo: Každá dvojice vrcholů leží v přesně 3 blocích
- Potom systém B_1, \dots, B_{11} bude 2 -($11, 6, 3$) design
- Záleží na pořadí řádků D ? Ne! Permutace řádků 2 až 11 a a stejná permutace sloupců 2 až 11 zachová „tvar“ M

- (Jednoduchý) $2-(v, k, \lambda)$ design je systém k -prvkových podmnožin (bloků) B_1, B_2, \dots množiny $V = \{1, 2, \dots, v\}$ takový, že každá dvojice vrcholů leží v přesně λ blocích
- Náš systém B_1, \dots, B_{11} je $2-(11, 6, 3)$ design
- Víme více: Počet našich bloků = počet vrcholů; tomu se říká čtvercový design

Příklad: Fanova rovina



- Čtvercový $2-(7, 3, 1)$ design
- 7 vrcholů, 7 bloků
- Každé dva body určují blok (přímku)
- Fanova rovina je častý (proti)příklad v kombinatorice a geometrii
- Příklad použití designu v návrhu pokusů: Máme 7 léků (vrcholů), 7 pokusných pacientů (bloků); otestujeme interakce každé dvojice léků

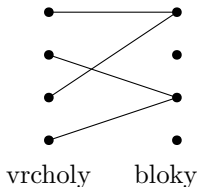
Každý design je regulární

- Regulární: Každý vrchol leží ve stejném počtu bloků

Věta

Bud' $\{B_1, \dots, B_b\}$ 2 -(v, k, λ) design na V . Pak každý vrchol $z V$ leží v $\lambda(v-1)/(k-1)$ blocích.

- Bud' a vrchol, který leží v r blocích
- Uvažme bipartitní graf s partitami $V \setminus \{a\}$, $\{B_1, \dots, B_b\}$.
- Hrany $\{i, j\}$ pro $\{a, i\} \subset B_j$:



- Spočteme počet hran dvěma způsoby: $r(k-1) = \lambda(v-1)$

Design \mathcal{D} určený D

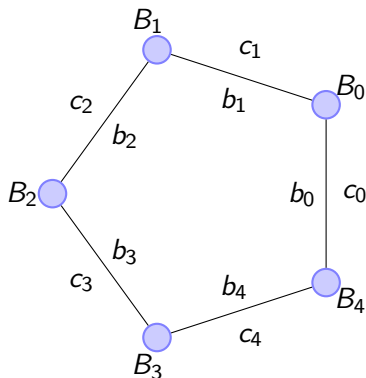
- Velikost bloků 6
- Každá dvojice vrcholů je ve 3 blocích
- Každý vrchol je v $\lambda(v-1)/(k-1) = 3 \cdot 10/5 = 6$ blocích (není náhoda, že to vyšlo stejně jako velikost bloku)
- Průnik každých dvou bloků má velikost 3 (ani tohle není náhoda)
- PIE: Sjednocení každých dvou různých bloků má velikost $6 + 6 - 3 = 9$
- Chceme určit jednoznačnost designu \mathcal{D} až na permutace vrcholů

- Princip inkluze a exkluze: Bud' A, B, C bloky designu \mathcal{D}

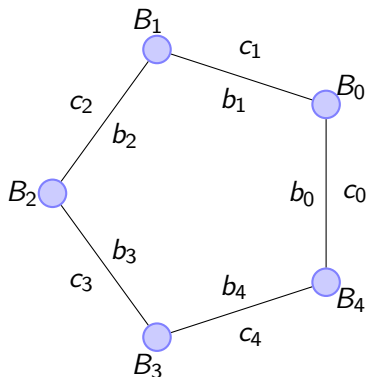
$$|A \cup B \cup C| = 3 \cdot 6 - 3 \cdot 3 + |A \cap B \cap C| = 9 + |A \cap B \cap C| \leq 11$$

- Tedy $A \cap B \cap C$ má nejvýš 2 prvky
- Pokud $|A \cap B \cap C| = 2$, tak $A \cup B \cup C$ obsahuje všechny vrcholy
- Přeznačíme bloky a vrcholy jako ve skriptech (Věta 15.4.1)
- $V = \{a, b_0, b_1, \dots, b_4, c_0, c_1, \dots, c_4\}$
- Búno $a \in A, B_0, B_1, B_2, B_3, B_4$
- Búno $A = \{a, b_0, b_1, b_2, \dots, b_4\}$

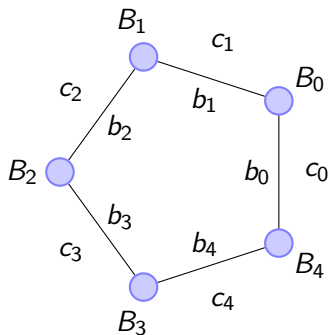
- Pro $i = 0, 1, 2, 3, 4$ máme $|A \cap B_i| = 3$, z toho jeden prvek je a
- Může být $A \cap B_i = A \cap B_j$ pro $i \neq j$? Ne, pak by bylo $|A \cap B_i \cap B_j| = |A \cap B_i| = 3$, spor s předchozím slajdem
- Definujme graf G na $\{b_0, b_1, \dots, b_4\}$; hrany $A \cap B_i \setminus \{a\}$
- Máme celkem 5 hran
- Každé b_i leží spolu s a v celkem 3 blocích: A a dvou množinách B_i , tedy všechny stupně vrcholů G jsou 2
- Graf s 5 vrcholy, 5 hranami a vrcholy stupně 2 je pěticyklus



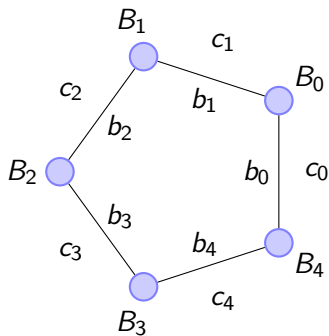
- Graf s 5 vrcholy, 5 hranami a vrcholy stupně 2 je pěticyklus
- Búno je $A \cap B_i = \{a, b_i, b_{i+1}\}$ (indexy modulo 5)
- Víme, že $B_{i-1} \cap B_i$ tvoří vrcholy a, b_i a jeden další bod mimo $\{a, b_0, \dots, b_4\}$ – označme ho c_i
- Jsou c_0, c_1, c_2, c_3, c_4 po dvou různé?



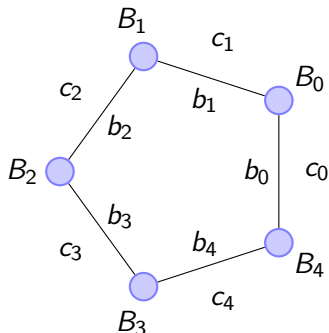
- Nechť $c_1 = c_2$. Pak $B_1 \cap B_2 \cap B_0 = \{a, c_1\}$, tedy $|B_1 \cup B_2 \cup B_0| = 11$
- Ale pak $b_4 \in B_1$ nebo $b_4 \in B_2$ nebo $b_4 \in B_0$
- Pak $A \cap B_1$ nebo $A \cap B_2$ nebo $A \cap B_0$ má 4 prvky, což nejde



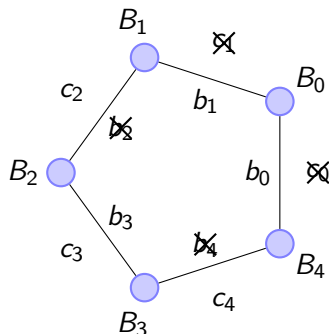
- Nechť $c_1 = c_3$. Pak $c_1 \in B_1, B_2$ a $B_1 \cap B_2 = \{a, b_2, c_2\}$, tedy $c_1 = c_2 = c_3$ a spor jako na předchozím slajdu
- Ostatní případy $c_i = c_j$ jsou symetrické



- Máme $V = \{a, b_0, b_1, \dots, b_4, c_0, c_1, \dots, c_4\}$
- Máme $\{a, b_i, b_{i+1}, c_i, c_{i+1}\} \subset B_i$; jaký prvek chybí?
- Nemůže to být b_j ani c_{i-1}, c_{i+2} , tedy zbývá c_{i+3}
- Dokázali jsme $B_i = \{a, b_i, b_{i+1}, c_i, c_{i+1}, c_{i+3}\}$



- Dokázali jsme $B_i = \{a, b_i, b_{i+1}, c_i, c_{i+1}, c_{i+3}\}$
- Zbývá 5 bloků C_0, C_1, C_2, C_3, C_4 neobsahujících a
- Dvojice $\{b_3, c_3\}$ leží v B_2, B_3 a (búno) v C_3
- Protože $B_2 \cap B_3 \cap C_3$ obsahuje dva prvky, tak $b_0, b_1 \in B_2 \cup B_3 \cup C_3$
- Tedy $C_3 = \{b_3, c_3, b_0, b_1, ?, ?\}$



- $C_3 = \{b_3, c_3, b_0, b_1, ?, ?\}$
- Přitom $b_0, b_1 \notin B_2, B_3$, tedy $b_0, b_1 \in C_3$
- $A \cap C_3 = \{b_3, b_0, b_1\}$, tedy $b_2, b_4 \notin C_3$
- $B_0 \cap C_3 = \{b_0, b_1, c_3\}$, tedy $c_0, c_1 \notin C_3$
- Zbývá $C_3 = \{b_3, c_3, b_0, b_1, c_2, c_4\}$

- Symetricky $C_i = \{b_i, b_{i+2}, b_{i-2}, c_{i-1}, c_i, c_{i+1}\}$
- Určili jsme všech 11 bloků systému \mathcal{D}
- Vše je až na pořadí vrcholů jednoznačné, tedy matice D je jednoznačná
- Víme, že \mathcal{G}_{24} je $[24, 12, 8]$ -kód
- Tím pádem je $[24, 12, 8]$ -kód vždy ekvivalentní \mathcal{G}_{24}

- Jak vypadá kódování pro dlouhé bloky?
- Nekonečné rodiny kódů: Totální, paritní, opakovací Hammingovy, Reed-Solomonovy, Reed-Mullerovy, BCH, QR kódy...
- Kdy použít které?
- Zhruba: Zvolím si abecedu a prahovou hodnotu p , pro danou velikost abecedy chceme z kódů s **relativní vzdáleností** $d/n > p$ zvolit kód s největší hustotou k/n
- Singletonův odhad $k \leq n + 1 - d < n + 1 - np = n(1 - p) + 1$, tj. horní mez na hustotu je asi $1 - p$

- Totální kódy: Hustota 1; rel. vzdálenost $1/n \rightarrow 0$
- Paritní kódy: Hustota $1 - 1/n \rightarrow 1$; rel. vzdálenost $2/n \rightarrow 0$
- Opakovací kódy: Hustota $1/n \rightarrow 0$; rel. vzdálenost 1
- Hammingovy kódy: Hustota asi $(n - \log n)/n \rightarrow 1$; rel. vzd. $3/n \rightarrow 0$
- Reed-Solomonovy: Volme $q = 2^m, k = 2^{m-17}$ Kódujme \mathbb{F}_{2^m} v binární abecedě, aby to bylo fér, tedy $n = m2^m$
- Hustota: $m2^{m-17}/m2^m = 2^{-17}$
- Rel. vzdálenost $(2^m - k)/m2^m = (1 - 2^{-17})/m \rightarrow 0$

Asymptotické odhady [na přednášce jsme tento slajd přeskočili]

- Reed-Mullerovy kódy: Máme parametry m, r . Chceme $2^{m-r} > p$
- Pro fixní r a velké m máme hustotu $\leq 2\binom{m}{r}/2^m \leq 2m^r/2^m \rightarrow 0$
- BCH (pesimisticky, pro fixní q): Hustota
 $q^m - (m+1)\delta - 1/q^m - 1 = 1 - (m+1)\delta/(q^m - 1) = 1 - (m+1)d/n$
(pro velké m a konstantní jde relativní vzdálenost k nule)
- QR kódy (pesimisticky): Hustota $(p+1)/2p \rightarrow 1/2$; rel. vzd.
 $d/n = \sqrt{p}/p \rightarrow 0$

- Existuje vůbec rodina kódů, která by měla pro $n \rightarrow \infty$ jak k/n , tak d/n odražená od 0?!
- Je pravda, že pokud je relativní vzdálenost kódu $>$ pst chyby, tak je chybně dekódovaných slov „málo“?
- A co je to „málo“ chyb?
- A jak vlastně modelovat komunikační kanál?