Infanticida u myší

Test hypotézy sexuálního soutěžení: samec zvyšuje svůj sexuální úspěch na úkor konkurentů tím, že zabíjí jejich potomstvo a páří se s jejich matkou.

Studie: 114 jedinců 90-denních myší

Funkce připustnosti stavu

v_1 dominance $\begin{cases} 0 & \text{dominantní} \\ 1 & \text{podřízený} \end{cases}$

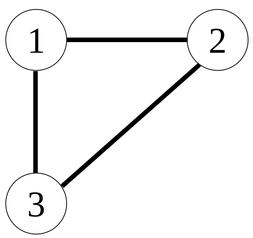
v_2 sexuální zkušenost $\begin{cases} 0 & \text{nemá} \\ 1 & \text{má} \end{cases}$

v_3 chování vůči potomstvu $\begin{cases} 0 & \text{infanticidní} \\ 1 & \text{rodičovské} \\ 2 & \text{nevšímavé} \end{cases}$

v_1	v_2	v_3	N
0	0	0	28
0	0	1	4
0	0	2	2
0	1	0	5
0	1	1	25
0	1	2	3
1	0	0	5
1	0	1	9
1	0	2	8
1	1	0	7
1	1	1	15
1	1	2	3

Infanticida u myší: rekonstrukční analýza

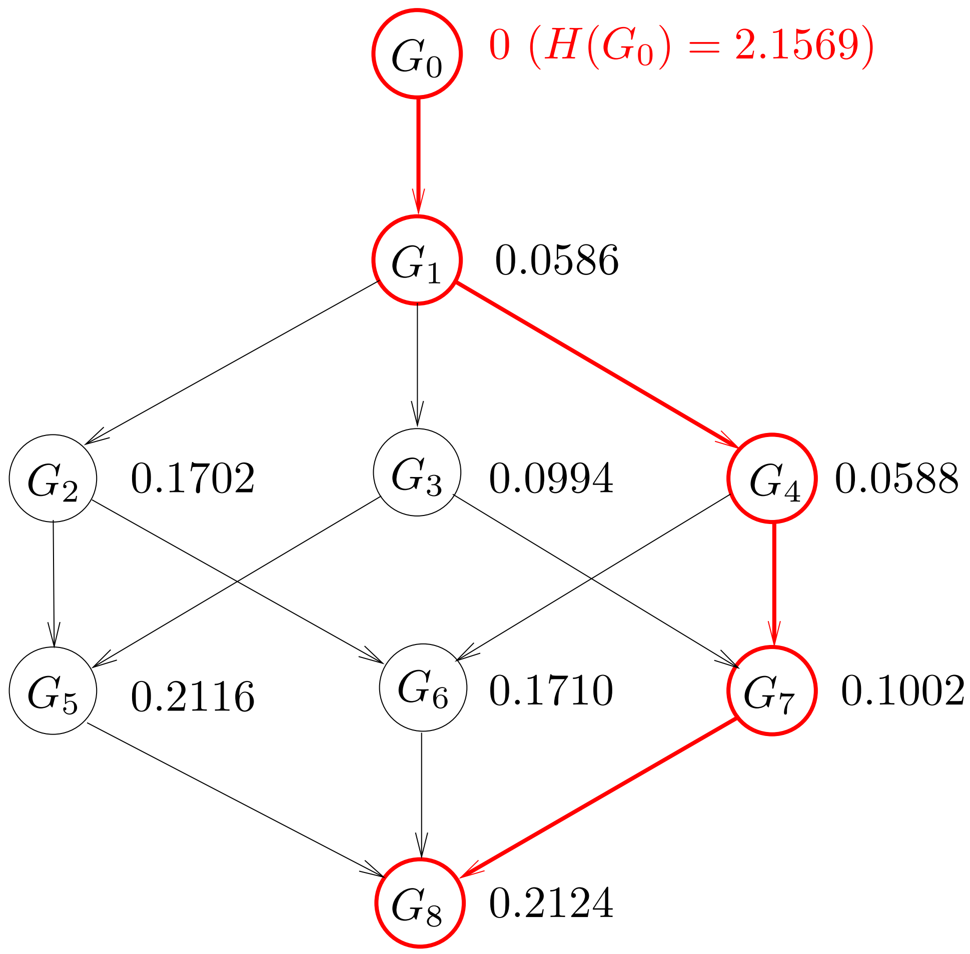
dominance



zkušenost

chování

svaz zjemnění



rekonstrukční hypotézy

$G_0 = 123$

$G_1 = 12|13|23$

$G_2 = 12|13$

$G_3 = 12|23$

$G_4 = 13|23$

$G_5 = 12|3$

$G_6 = 13|2$

$G_7 = 1|23$

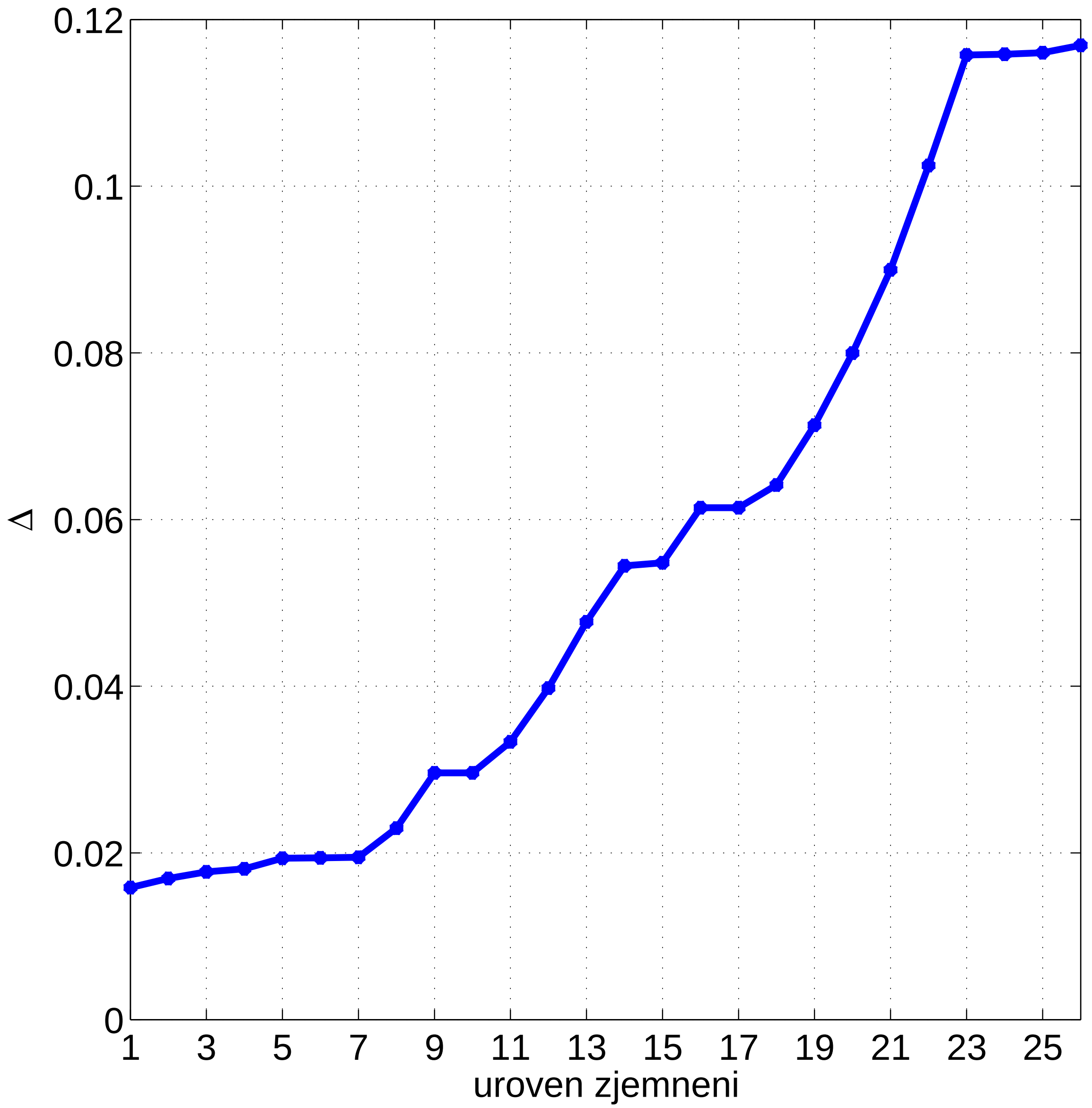
$G_8 = 1|2|3$

Závěry

- chování je ovlivněno oběma faktory, ale zkušenost je dvakrát významnější než dominance (ΔG_6, ΔG_7)
- dominance a zkušenost jsou nezávislé ($\Delta G_3 \sim \Delta G_7$, první vazba, která se rozpadne)

Struktura Parlamentu

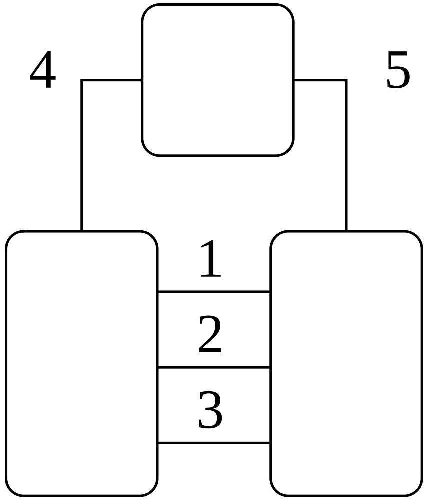
k	G_k	ΔG_k
1	1234|2345|3451|4512|5123	0.0158
2	1234|2345|3451|5123	0.0169
3	1234|2345|5123|451	0.0177
4	1234|2345|5123	0.0181
5	1234|5123|345|452	0.0194
6	1234|5123|345	0.0194
7	1234|5123|45	0.0195
8	5123|45|234|341|412	0.0230
9	45|234|341|412|512|123|235|351	0.0296
10	45|234|341|412|123|235|351	0.0296
11	45|234|341|412|235|351	0.0333
12	45|234|341|412|235|51	0.0398
13	45|234|341|235|51|12	0.0477
14	45|341|235|51|12|42	0.0544
15	45|341|235|51|12	0.0548
16	45|235|51|12|34|41|13	0.0614
17	45|235|51|12|34|13	0.0614
18	45|235|51|12|13	0.0641
19	45|235|51|12	0.0713
20	235|51|12|4	0.0800
21	235|12|4	0.0900
22	235|4|1	0.1025
23	4|1|23|35|52	0.1158
24	4|1|23|35	0.1158
25	4|1|35|2	0.1160
26	4|1|2|3|5	0.1169

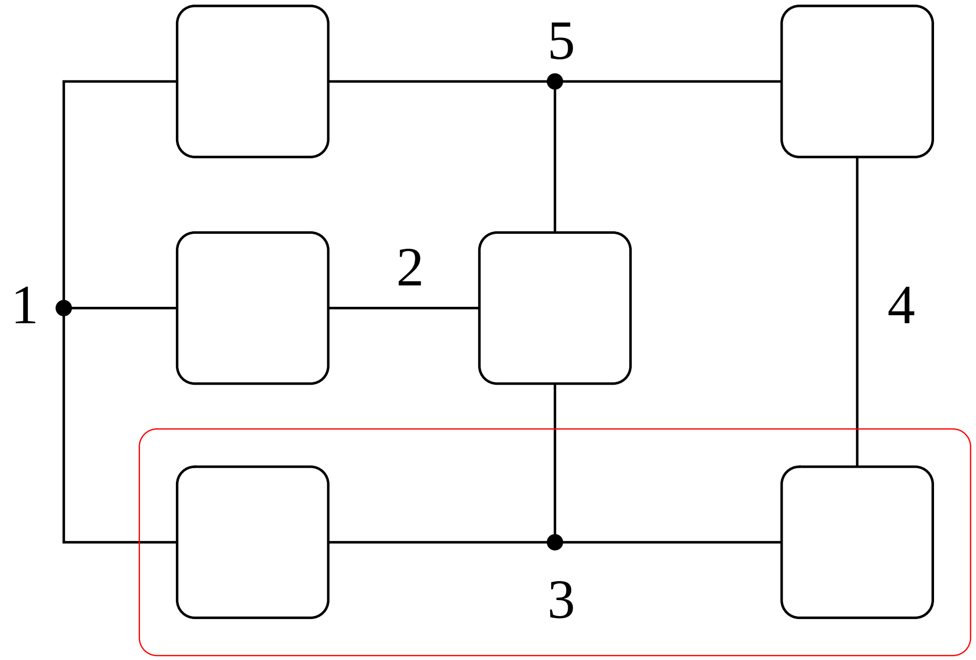


Parlament: interpretace úrovně 7 a 15+17

k	G_k	ΔG_k
7	1234|5123|45	0.0195

k	G_k	ΔG_k
15	45|341|235|51|12	0.0548
17	45|235|51|12|34|13	0.0614





Posibilistická funkce přípustnosti

- posibilita π: stupeň možnosti
- $\pi = 1 \ldots$ událost je možná

x_i elementární nezávislé jevy, ‚hodnoty proměnných z S' , $x_i \in \mathcal{R}(S)$

posibilita	pravděpodobnost
definiční vlastnosti	
$\pi\colon\ \exp \mathcal{R}(S) \rightarrow \langle 0, 1 \rangle$	$p\colon\ \exp \mathcal{R}(S) \rightarrow \langle 0, 1 \rangle$
$\pi(\{\}) = 0$	$p(\{\}) = 0$
$\pi\big(\mathcal{R}(S)\big) = 1$	$p\big(\mathcal{R}(S)\big) = 1$
$\pi\Big(\bigcup_i x_i\Big) = \max_{x_i \in \cancel{\mathcal{R}(S)}\ Q} \pi(x_i)$ $\quad Q \subseteq \mathcal{R}(S)$	$p\Big(\bigcup_i x_i\Big) = \sum_{x_i \in \cancel{\mathcal{R}(S)}\ Q} p(x_i)$
odhad z dat	
$\pi(x_i) \approx \dfrac{N(x_i)}{\max_{x_j \in \mathcal{R}(S)} N(x_j)}$	$p(x_i) \approx \dfrac{N(x_i)}{\sum_{x_j \in \mathcal{R}(S)} N(x_j)}$

Posibilistická míra neurčitosti

$$X_\alpha = \{x_i,\ \pi(x_i) \geq \alpha\} \qquad \alpha\text{-řez}$$

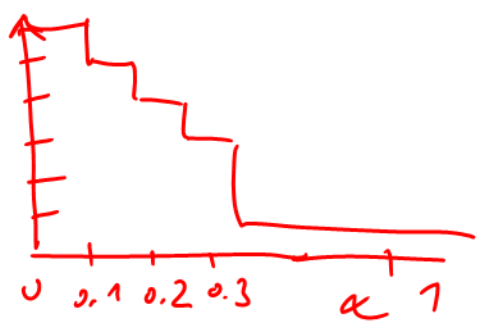
$$L = \{\alpha,\ \exists x_i\colon \pi(x_i) = \alpha\} \cup \{0\} \qquad \text{množina } \alpha\text{-hladin}$$

U-míra

$$U(\pi) = \sum_{j=2}^{k} (\alpha_j - \alpha_{j-1}) \log |X_{\alpha_j}| = \int_0^1 \log |X_\alpha| \, d\alpha$$

---

x_i	1	2	3	4	5	6
$\pi(x_i)$	0.1	0.0	0.3	1.0	0.3	0.2



příklad

j	α_j	X_{α_j}						$\lvert X_{\alpha_j}\rvert$
1	0.0	×	×	×	×	×	×	6
2	0.1	×		×	×	×	×	5
3	0.2			×	×	×	×	4
4	0.3			×	×	×		3
5	1.0				×			1

$$U = \underbrace{0.1 \cdot \log 5}_{j=2} + \underbrace{0.1 \cdot \log 4}_{j=3} + \underbrace{0.1 \cdot \log 3}_{j=4} + \underbrace{0.7 \cdot \log 1}_{j=5}$$

Posibilistické spojení

Posibilistická marginální funkce

$$\pi(x) = \max_y \pi(x, y)$$

Posibilistická generativní neurčitost

$$U(G \mid \bar{G}) = U(G, \bar{G}) - U(\bar{G})$$

Posibilistické spojení

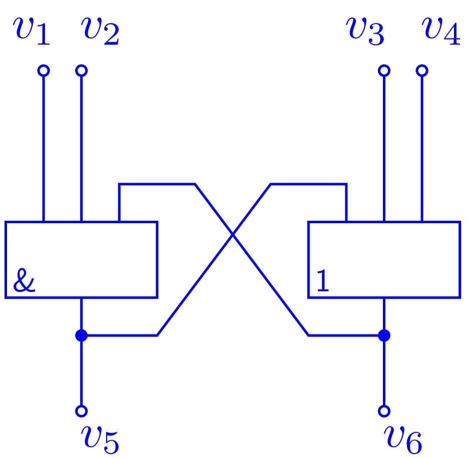
$$({}^1\pi * {}^2\pi)(a, b, c) = \min\left({}^1\pi(a, b),\ {}^2\pi(b, c)\right)$$

není nutná iterativní spojovací procedura

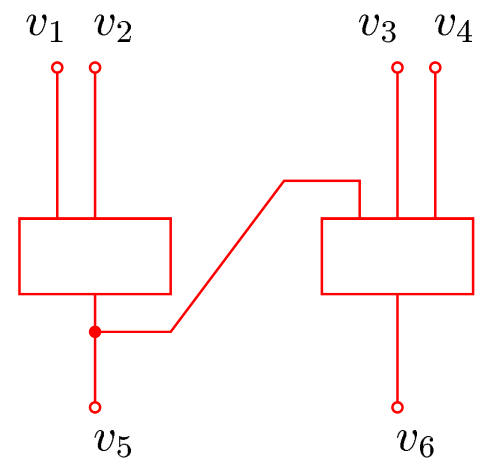
Rekonstrukční chyba

$$\Delta(S, S^*) = L(\pi, \pi^*) = \int_0^1 \log \frac{|X_\alpha(\pi^*)|}{|X_\alpha(\pi)|}\, d\alpha$$

Diagnostika obvodů



předpokládaný systém



diagnostifikovaný systém

není pozorován stav a

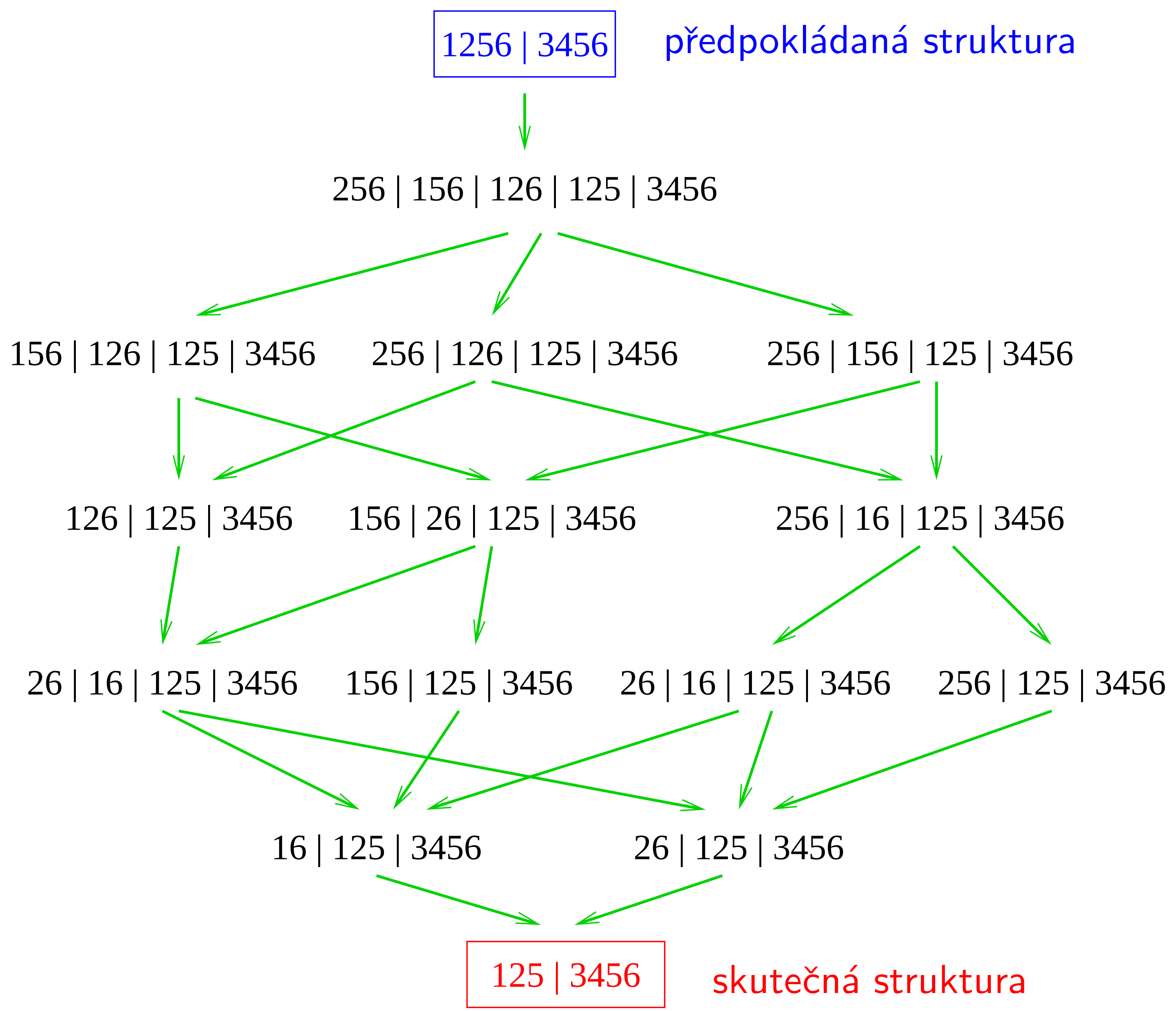
pozorované stavy
$(\pi(x_i) = 1)$

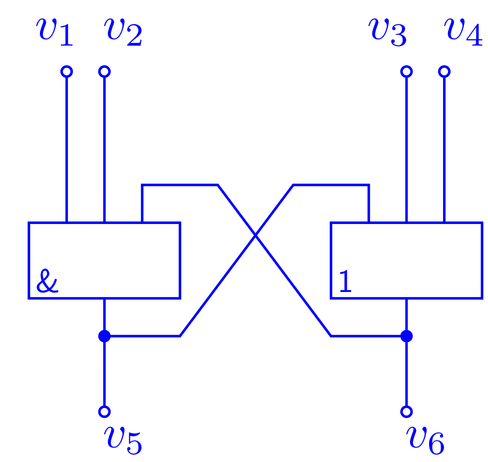
	$\bar{G}$				G	
	v_1	v_2	v_3	v_4	v_5	v_6
	0	0	0	0	0	0
	0	0	0	1	0	1
	0	0	1	0	0	1
	0	0	1	1	0	1
	0	1	0	0	0	0
	0	1	0	1	0	1
	0	1	1	0	0	1
	0	1	1	1	0	1
	1	0	0	0	0	0
	1	0	0	1	0	1
	1	0	1	0	0	1
	1	0	1	1	0	1
~~$a =$~~	~~1~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0~~
$b =$	1	1	0	0	1	1
	1	1	0	1	1	1
	1	1	1	0	1	1
	1	1	1	1	1	1

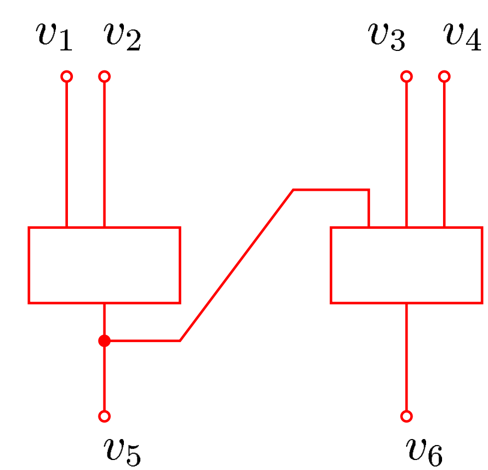
správná rekonstrukční hypotéza $1256 \mid 3456 = G_A$

- $\Delta G_A > 0 \Rightarrow$ obvod je vadný — nutná rekonstrukční analýza pro předchůdce G_A
- $\Delta G_A = 0$ nedokazuje plnou funkcionalitu — nutná rekonstrukční analýza pro následníky G_A

Podgraf s nulovou rekonstrukční chybou



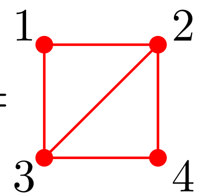




Pozn: $\Delta(1256 \mid 346) = 3$

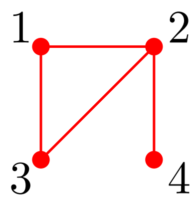
Struktura svazu zjemnění: Systémy o podobném stupni rozkladu

Sestrojíme graf $\mathcal{R}$, jehož uzly jsou proměnné a hrana mezi dvěma uzly je přítomna právě když obě proměnné jsou v jednom podsystému.

$\mathcal{R} =$ 

reprezentuje:

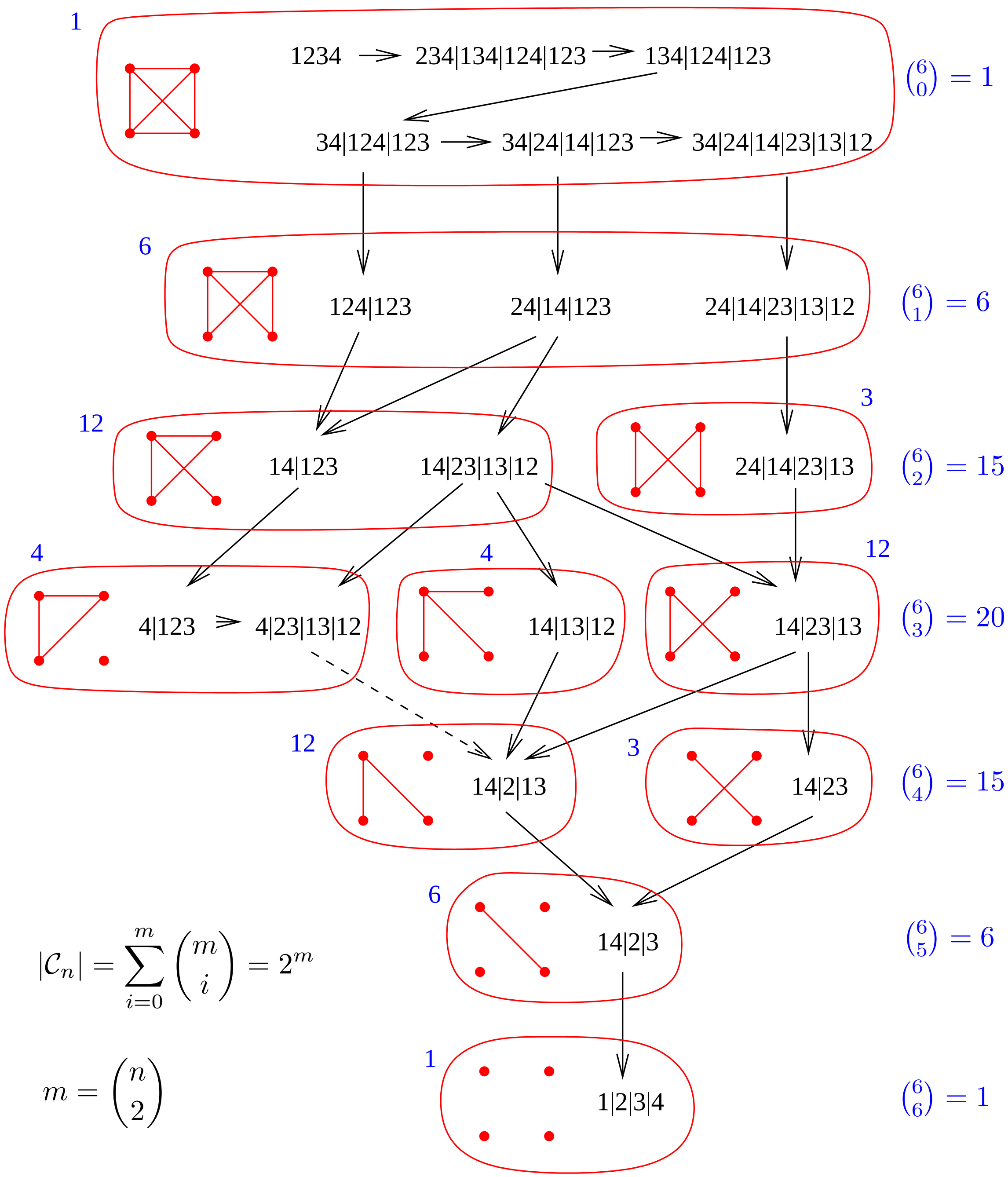
$G_1 = 123|234,\ G_2 = 12|31|234,\ G_3 = 123|24|34,\ G_4 = 12|13|23|34|24$

Ale rozklad $123|24$ by měl 

Pozorování

- graf $\mathcal{R} = (V_n, E)$ generuje třídu ekvivalence rozkladů, budeme je nazývat **r-ekvivalence**, kde V_n – množina n proměnných systému, E – množina dvojic proměnných, které jsou obě v nějakém podsystému
- G_1 je nejméně dekomponovaný mezi všemi rozklady odpovídajícími grafu $\mathcal{R}$: graf je interpretován jako množina klik, této interpretaci budeme říkat **$\mathcal{C}$-hypotéza** clique
- G_4 je nejvíce dekomponovaný, této interpretaci $\mathcal{R}$ budeme říkat **$\mathcal{P}$-hypotéza** pair
- $\mathcal{C}$-hypotéza a $\mathcal{P}$-hypotéza jsou kanonickými reprezentanty třídy ekvivalence r
- $\mathcal{P}$-hypotéza je zjemněním $\mathcal{C}$-hypotézy (existuje do ní cesta ve svazu zjemnění)

Příklad: Svaz $\mathcal{C}_4/i$



$$|\mathcal{C}_n| = \sum_{i=0}^{m} \binom{m}{i} = 2^m$$

$$m = \binom{n}{2}$$

- Svaz zjemnění pro n proměnných až na r-ekvivalenci značíme $\mathcal{C}_n$
- $\binom{6}{k}$ je počet kombinací odebrání k hran
- Potom

$$|\mathcal{C}_n| = 2^{\binom{n}{2}}$$

- Tento svaz až na permutace indexů proměnných značíme $\mathcal{C}_n/i$
- Uspořádání ve smyslu zjemnění je v takovém svazu zachováno

Počet $\mathcal{G}$-hypotéz a $\mathcal{C}$-hypotéz

n	1	2	3	4	5	6	7
$\lvert\mathcal{G}_n\rvert$	1	2	9	114	6894	7785062	2414627396434
$\lvert\mathcal{C}_n\rvert$	1	2	8	64	1024	32768	2097152
$\lvert\mathcal{G}_n/i\rvert$	1	2	5	20	180	?	?
$\lvert\mathcal{C}_n/i\rvert$	1	2	4	11	34	156	1044

$\mathcal{G}_n$ — svaz zjemnění rekonstrukčních hypotéz

$\mathcal{C}_n$ — svaz zjemnění až na ekvivalenci r, $\quad |\mathcal{C}_n| = 2^{\binom{n}{2}}$

$\mathcal{C}_n/i$ — svaz zjemnění až na permutace indexů

$\mathcal{C}_n/i$ — svaz zjemnění až na ekvivalenci r a permutace indexů

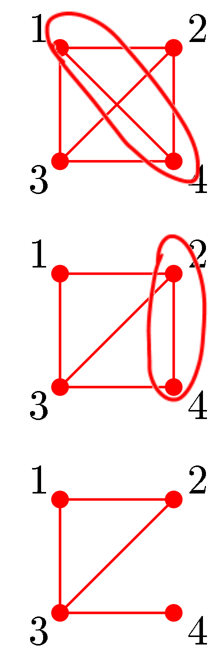
Prohledávání svazu zjemnění $\mathcal{C}$-hypotéz

Vstup: $\mathcal{C}$-hypotéza G_i, graf $r(G_i)$

Výstup: Všechna zjemnění $\mathcal{C}$-hypotézy v $\mathcal{C}$ svazu

Opakuj 1–3 pro všechny hrany grafu $r(G_i)$:

1. Vyjmi hranu (a, b) z grafu.
2. Každý podsystém ${}^kS \in G_i$ obsahující a i b rozděl na dva podsystémy ${}^kS \setminus \{a\}$ a ${}^kS \setminus \{b\}$ a nahraď jimi ${}^kS$.
3. Odstraň z výsledku redundantní podsystémy



1234

234 | 123

234 | 123

34 | ~~23~~ | 123

34 | 123

⋮

1 | 2 | 3 | 4

Hierarchizované prohledávání svazu zjemnění

▷ Generuj $\mathcal{C}$-hypotézy včetně permutací indexů ($\mathcal{C}_n$) Langdonův algoritmus, viz [Knuth 1997]

▷ Generuj $\mathcal{G}$-hypotézy uvnitř každé třídy ekvivalence r

nutná modifikace základní zjemňovací procedury: podmínka $|{}^iS| \geq 2$ se nahradí podmínkou $|{}^iS| > 2$ (pak zůstaneme v třídě $r(G)$)

Shrnutí

PA Identifikace parametrů generativního systému z datového systému.

$\boxed{\bullet \text{ volbou masky}}$

kritéria

1. generativní neurčitost = jak dobře systém generuje generované proměnné

$$H(G \mid \bar{G}) = H(S) - H(\bar{G}), \quad H(S) = H(G, \bar{G})$$

2. složitost = počet vzorkovacích proměnných

Algoritmus: prohledávání grafu

použití a vlastnosti

- nutná fáze modelování dynamických systémů
- existuje celá množina řešení

poznámky

- neparametrická metoda odhadu hustoty psti
- další možnosti: směs gausiánů, n nejbližších sousedů

Shrnutí

PB Zjednodušení modelu generativního systému.

- vylučováním proměnných
- redukcí rozlišení

kritéria

1. generativní neurčitost — jako v **PA**
2. složitost = počet stavů s nenulovou pravděpodobností

Algoritmus: prohledávání grafu

použití a vlastnosti

- redukce velikosti modelu před identifikací struktury
- existuje celá množina řešení

Shrnutí

PC Identifikace struktury systému.

- postupným zjemňováním dekompozice

kritéria

1. rekonstrukční neurčitost hypotézy G $\Delta(G) = L(p, p^*)$

   S^* – rekonstrukce a G s fcí připustnosti p^*

   S – původní systém s fcí připustnosti p
2. míra dekompozice $=$ počet podsystémů
3. statistická významnost na hladině α

použití a vlastnosti

- redukce dimenze popisu
- analýza kritických vazeb a závislostí
- dolování znalostí z dat
- diagnostika systémů
- volba deskriptivních příznaků pro rozpoznávání
- existuje celá množina řešení

Pozn • princip maxima entropie