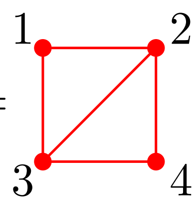
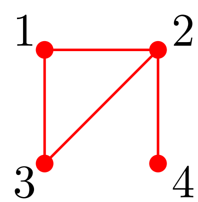
Struktura svazu zjemnění: Systémy o podobném stupni rozkladu

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:

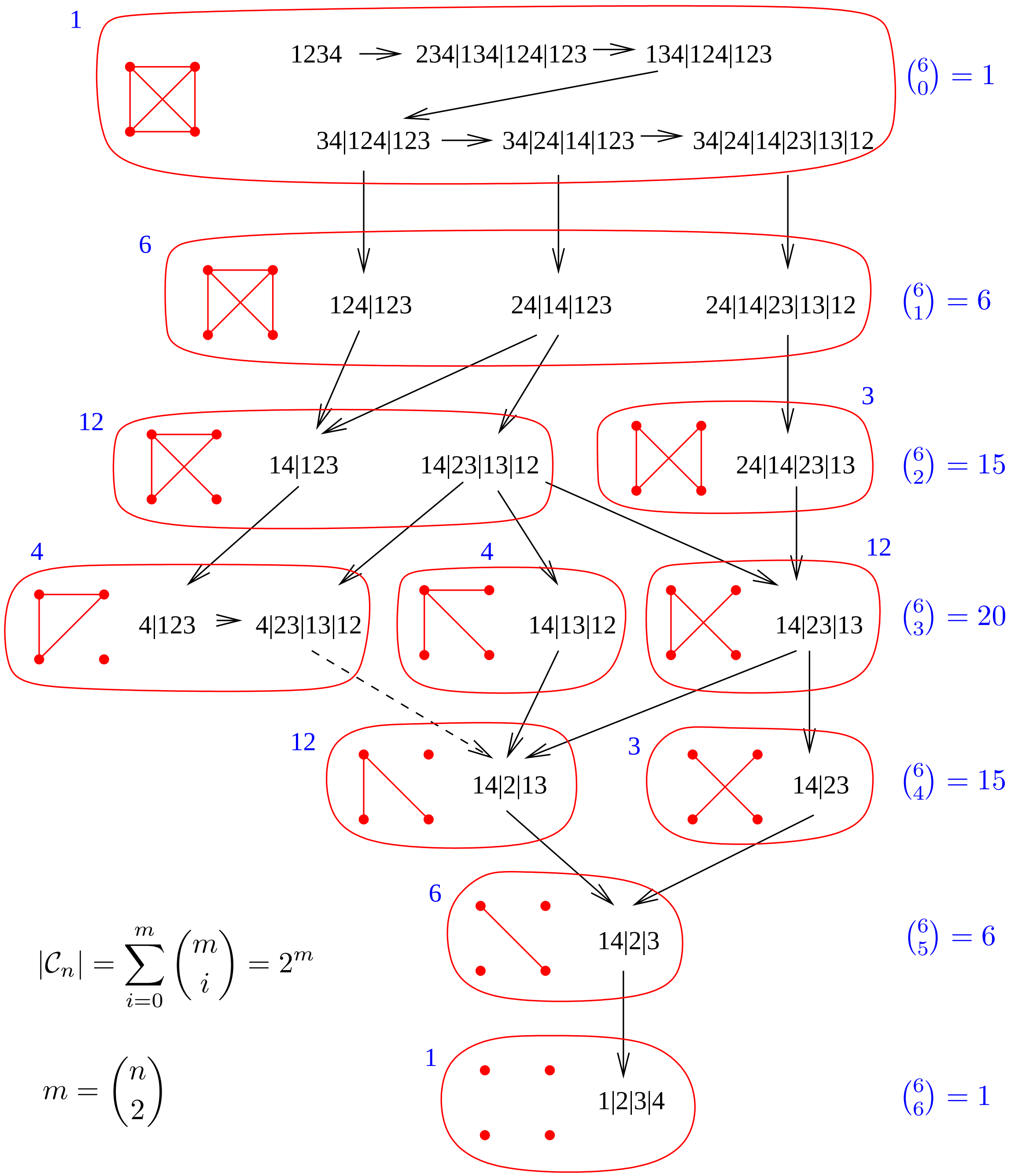
$^1\mathbf{S} = 123|234$, $^2\mathbf{S} = 12|31|234$, $^4\mathbf{S} = 123|24|34$, $^3\mathbf{S} = 12|13|23|34|24$

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
- $^1\mathbf{S}$ je nejméně dekomponovaný mezi všemi rozklady odpovídajícími grafu $\mathcal{R}$: graf je interpretován jako množina klik, této interpretaci budeme říkat **$\mathcal{C}$-hypotéza** clique
- $^3\mathbf{S}$ 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 až na r-ekvivalenci a <u>permutace</u> indexů proměnných značíme $\mathcal{C}_n/i$ (n proměnných)
- Uspořádání ve smyslu zjemnění je v takovém svazu zachováno

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

počet proměnných n

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

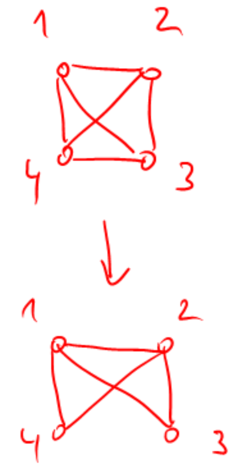
$\log|\mathcal{C}_n| = \binom{n}{2}$

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

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

$\binom{n}{2}\, 2^{\binom{n}{2}}$ celé hledání

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



1234

124 | 123

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

Hierarchizované prohledávání svazu zjemnění

Langdonův algoritmus Knuth

- $\triangleright$ Generuj $\mathcal{C}$-hypotézy včetně permutací indexů ($\mathcal{C}_n$)
  - $\triangleright$ 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.

- volbou masky

kritéria

1. generativní neurčitost = jak dobře systém generuje generované proměnné

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

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 $\Delta(\mathbf{S}^*) = L(\mathbf{S}, \mathbf{S}^*)$
   $\mathbf{S}^*$ – rekonstrukce ze zvolených podsystémů $\mathbf{S}$ – původní systém
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