PQ-stromy a rozpoznávání intervalových grafů v lineárním čase

Permutace s předepsanými intervaly

Označme $[n]$ množinu $\{1, 2, \ldots, n\}$. Mějme permutaci $\pi = \pi_1, \pi_2, \ldots, \pi_n$ množiny $[n]$. Řekneme, že množina $S \subseteq [n]$ tvoří *interval* v permutaci π, pokud existují indexy $i \leq j$ takové, že $S = \{\pi_i, \pi_{i+1}, \ldots, \pi_j\}$. Řekneme, že S je *cyklický interval* v π, pokud existují indexy $i \leq j$ takové, že buď $S = \{\pi_i, \pi_{i+1}, \ldots, \pi_j\}$, nebo $S = \{\pi_j, \pi_{j+1}, \ldots, \pi_n\} \cup \{\pi_1, \ldots, \pi_i\}$.

Uvažme následující dva rozhodovací problémy, které označíme TI a TCI, což jsou zkratky pro 'testování intervalů', resp. 'testování cyklických intervalů'.

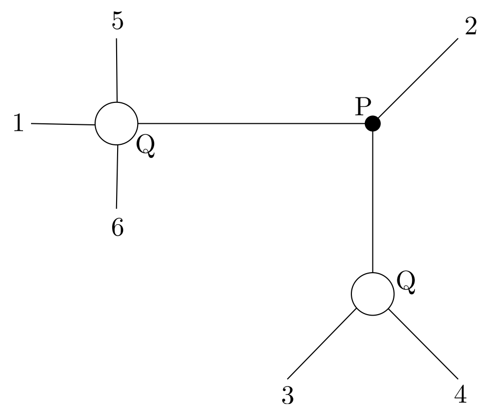
Problém TI: Na vstupu máme číslo $n \in \mathbb{N}$ a k-tici množin S_1, S_2, $\ldots$, $S_k \subseteq [n]$. Otázka je, zda existuje permutace π množiny $[n]$ taková, že každá zadaná množina S_i tvoří interval v π.

Problém TCI: Na vstupu máme opět číslo $n \in \mathbb{N}$ a k-tici množin S_1, S_2,$\ldots$, $S_k \subseteq [n]$. Otázka je, zda existuje permutace π množiny $[n]$ taková, že každá zadaná množina S_i tvoří cyklický interval v π.

Problém TI lze snadno převést na problém TCI, protože pro libovolnou k-tici množin $S_1, S_2, \ldots, S_k \subseteq [n]$ platí, že existuje permutace $[n]$, v níž tyto množiny tvoří intervaly, právě tehdy, když existuje permutace $[n+1]$, v níž tyto množiny tvoří cyklické intervaly. Dále se tedy budeme zabývat pouze řešením problému TCI.

Všimneme si nejprve, že pokud množina S tvoří cyklický interval v permutaci $\pi = \pi_1, \ldots, \pi_n$, tak S tvoří cyklický interval i v libovolném cyklickém posunu π, tj. v libovolné permutaci tvaru $\pi_i, \pi_{i+1}, \ldots, \pi_n, \pi_1, \pi_2, \ldots, \pi_{i-1}$. Proto prohlásíme, že dvě permutace jsou ekvivalentní, jestliže jedna je cyklickým posunem druhé, a třídám této ekvivalence budeme říkat *cyklické permutace*. Cyklickou permutaci si můžeme představit jako způsob, jak přiřadit čísla $1, \ldots, n$ v libovolném pořadí vrcholům pravidelného n-úhelníku, přičemž čísla budeme číst po směru hodinových ručiček, ale nerozlišujeme, které číslo přečteme první.

Pro dané množiny $S_1, \ldots, S_k \subseteq [n]$ označme $\mathrm{Cyc}_n(S_1, \ldots, S_k)$ množinu všech cyklických permutací $[n]$ takových, že každá z množin $S_1, \ldots, S_k$ se v nich vyskytuje jako cyklický interval. Naším cílem při řešení problému TCI je tedy poznat, zda je $\mathrm{Cyc}_n(S_1, \ldots, S_k)$ neprázdná.



Obrázek 1: Příklad PQ-stromu.

PQ-stromy

Pro vyřešení problému Tci využijeme datovou strukturu zvanou PQ-strom, kterou nyní popíšeme. *PQ-strom* řádu n je strom T s následujícími vlastnostmi:

- T má n listů, očíslovaných čísly $1, \ldots, n$.
- Vnitřní vrcholy T jsou dvou typů: P-vrcholy a Q-vrcholy.
- Každý vnitřní vrchol v stromu T má předepsanou cyklickou permutaci svých sousedů popisující, v jakém pořadí vycházejí z vrcholu v hrany.

Tím, že u každého vnitřního vrcholu PQ-stromu je předepsáno cyklické pořádí jeho sousedů, získáme jednoznačné cyklické pořadí listů stromu, tj. cyklickou permutaci množiny $[n]$, kterou označíme $\pi(T)$. Pro strom na Obrázku 1 je to například permutace 152436.

Řekneme, že dva PQ-stromy T a T' jsou *ekvivalentní*, pokud T lze převést na T' posloupností následujících dvou typů operací:

- Zvolíme libovolný P-vrchol a cyklickou permutaci jeho sousedů nahradíme libovolnou jinou cyklickou permutací.
- Zvolíme libovolný Q-vrchol a cyklickou permutaci jeho sousedů nahradíme zrcadlově obrácenou cyklickou permutací.

Každému stromu T pak můžeme přiřadit množinu cyklických permutací

$$R_T = \{\pi(T'),\, T' \text{je ekvivalentní s } T\}.$$

Řekneme, že strom T *reprezentuje* množinu cyklických permutací R_T. Strom z Obrázku 1 reprezentuje množinu osmi cyklických permutací: 152436, 152346, 162435, 162345, 154326, 153426, 164325 a 163425.

Z následující věty plyne, že Tci lze řešit v lineárním čase:

Věta 1. *Pro zadané n a zadané množiny $S_1, S_2, \ldots, S_k \subseteq [n]$ lze v čase $O(n + k + \sum_{i=1}^{k} |S_i|)$ buď zjistit, že $\mathrm{Cyc}_n(S_1, \ldots, S_k)$ je prázdná, nebo zkonstruovat PQ-strom T, pro nějž platí $R_T = \mathrm{Cyc}_n(S_1, \ldots, S_k)$.*

Částečný důkaz Věty 1. Nebudeme zde probírat celý důkaz této věty, pouze si předvedeme, jak se dá příslušný PQ-strom zkonstruovat. Z popisu tohoto postupu bude zřejmé, že konstrukci lze provést v polynomiálním čase. Pro lineární odhad na časovou složitost by bylo potřeba použít poměrně netriviální amortizaci jednotlivých operací, což dělat nebudeme.

Pro zadané množiny $S_1, \ldots, S_k$ zkonstruujeme postupně posloupnost PQ-stromů $T_0, T_1, \ldots, T_k$, kde T_i bude reprezentovat množinu $\mathrm{Cyc}_n(S_1, \ldots, S_i)$ a T_0 bude reprezentovat množinu všech cyklických permutací množiny $[n]$.

Strom T_0 vyrobíme snadno: obsahuje jeden vnitřní vrchol typu P, s nímž sousedí všech n listů.

Nyní předpokládejme, že už jsme pro nějaké $i \geq 0$ zkonstruovali strom T_i reprezentující $\mathrm{Cyc}_n(S_1, \ldots, S_i)$. Popišme, jak tento strom modifikovat na strom T_{i+1} reprezentující $\mathrm{Cyc}_n(S_1, \ldots, S_{i+1})$. Označme $S := S_{i+1}$ a $\overline{S} := [n] \setminus S$. Nechť e je libovolná hrana stromu T_i a nechť T' je jedna ze dvou komponent grafu $T_i - e$, tedy T' je nějaký podstrom stromu T_i. Řekneme, že T' je *plný*, pokud všechny listy T_i patřící do T' reprezentují prvky S, naopak T' je *prázdný*, pokud všechny tyto listy reprezentují prvky $\overline{S}$, a konečně T' je *smíšený*, pokud na svých listech má prvky S i $\overline{S}$. Řekneme, že hrana e je *smíšená*, pokud obě komponenty $T_i - e$ jsou smíšené podstromy.

Nechť E_S je množina všech smíšených hran T_i. Všimněme si, že hrany E_S tvoří souvislý podgraf T_i, neboť mám-li dvě smíšené hrany e, e', tak i každá hrana na cestě mezi e a e' je smíšená.

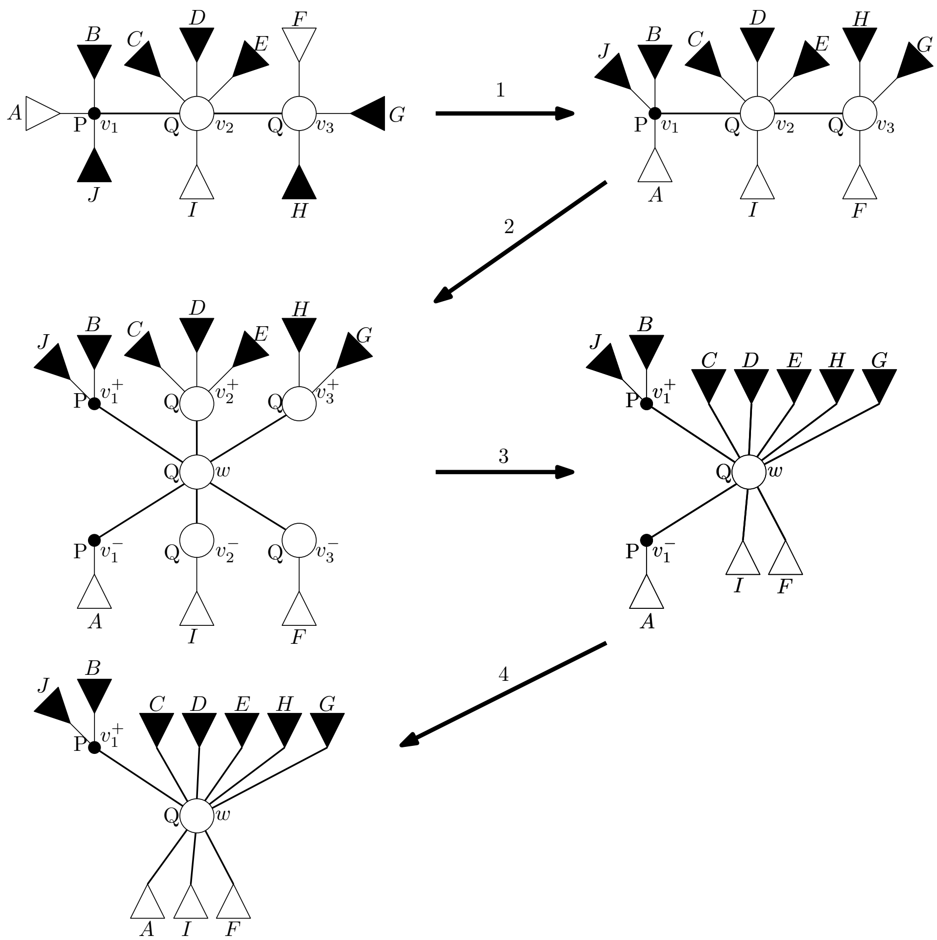
Všimněme si také, že pokud je některý vrchol stromu T_i incidentní s alespoň třemi smíšenými hranami, tak množina $S = S_{i+1}$ není cyklický interval v žádné cyklické permutaci reprezentované stromem T_i, a tedy $\mathrm{Cyc}_n(S_1, \ldots, S_{i+1}) = \emptyset$.

Odteď předpokládejme, že každý vrchol T_i je incidentní s nejvýše dvěma smíšenými hranami. Tedy buď je E_S prázdná, nebo tvoří cestu v T_i. Případ, kdy E_S je prázdná, je jednodušší a přenecháme ho čtenářům jako cvičení. Zabývejme se nyní situací, kdy E_S tvoří cestu P. Označme vrcholy této cesty $v_1, v_2, \ldots, v_m$ v pořadí, v jakém leží na cestě E_S. Všimněme si, že všechny hrany vycházející z vrcholů $v_1, \ldots, v_m$, které nepatří do cesty E_S, vedou buď do plného nebo do prázdného podstromu.

Představme si nyní, že cesta E_S je nakreslená vodorovně, s vrcholy označenými $v_1, \ldots, v_m$ zleva doprava. Se stromem T_i provedeme postupně následující úpravy (viz. Obrázek 2):

1. Nahradíme T_i ekvivalentním stromem T_i' takovým, že v každém vrcholu v_j budou hrany vedoucí do plných podstromů nad cestou E_S a hrany vedoucí do prázdných podstromů pod ní. Pokud takový ekvivalentní strom neexistuje, tak S_{i+1} netvoří cyklický interval v žádné permutaci reprezentované stromem T_i a můžeme skončit.

2. Každý vrchol v_j rozdělíme na dva nové vrcholy v_j^+ a v_j^- stejného typu jako



Obrázek 2: Postup, jak ze stromu T_i vytvořit strom T_{i+1}. Vyplněné trojúhelníčky odpovídají plným podstromům, prázdné trojúhelníčky odpovídají prázdným podstromům.

v_j, kde v_j^+ bude mít za sousedy plné podstromy sousedící s v_j, a v_j^- naopak ty prázdné. Cyklické pořadí sousedů bude stejné jako u v_j. Zároveň cestu E_S nahradíme jedním novým Q-vrcholem w, který bude mít za sousedy vrcholy $v_1^+, v_2^+, \dots, v_m^+, v_m^-, \dots, v_1^-$, v tomto cyklickém pořadí. V cyklickém pořadí sousedů u vrcholů v_j^+ a v_j^- bude w na té pozici, na které byly u v_j hrany E_S.

3. Pokud některý vrchol v_j byl Q-vrchol, tak zkontrahujeme hranu mezi w a v_j^+, jakož i hranu mezi w a v_j^-. To znamená, že tam, kde je k w připojen vrchol v_j^+ (nebo v_j^-), budou místo něj k w připojeny vrcholy sousedící s v_j^+ (resp. v_j^-).

4. Pokud ve vzniklém stromě existuje P-vrchol nebo Q-vrchol stupně 1, smažeme ho. Pokud existuje P-vrchol nebo Q-vrchol stupně 2, nahradíme ho hranou spojující jeho dva sousedy.

Strom vytvořený výše zmíněnými kroky označíme T_{i+1}. Dá se dokázat, že tento strom reprezentuje přesně množinu $\mathrm{Cyc}_n(S_1, \dots, S_{i+1})$. Formální důkaz tohoto faktu ovšem dělat nebudeme. □

Poznamenejme, že v literatuře se kromě nezakořeněných PQ-stromů, které jsme si zde představili, často užívá i zakořeněná verze PQ-stromů, která funguje obdobně, ale slouží k reprezentaci obyčejných permutací a intervalů, nikoliv cyklických permutací a cyklických intervalů. Popis konstrukce zakořeněných PQ-stromů je více technický, proto jsme se soustředili na nezakořeněnou verzi.

Rozpoznávání intervalových grafů

Díky PQ-stromům lze problém Tci, a tedy i Ti, řešit v lineárním čase. Nyní ukážeme, jak toto využít k získání lineárního algoritmu pro rozpoznávání intervalových grafů. Pomůže nám následující ekvivalentní charakterizace intervalových grafů.

Věta 2. *Graf $G = (V, E)$ je intervalový právě tehdy, když jeho maximální kliky lze uspořádat do posloupnosti $Q_1, Q_2, \dots, Q_t$ tak, že pro každý vrchol $x \in V$ tvoří kliky obsahující vrchol x interval v této posloupnosti. (Maximální klikou zde rozumíme kliku maximální vzhledem k inkluzi, tj. nikoliv nutně největší.)*

Důkaz. $\Rightarrow$: Nechť G je intervalový graf se zadanou intervalovou reprezentací, která každému vrcholu x přiřazuje interval I_x. Potom pro každou maximální kliku Q v grafu G existuje reálné číslo α_Q takové, že Q obsahuje přesně vrcholy x, pro něž platí $\alpha_Q \in I_x$. Uspořádáme-li maximální kliky zleva doprava podle hodnot α_Q, dostaneme uspořádání splňující pravou stranu ekvivalence.

$\Leftarrow$: Naopak, nechť platí pravá strana ekvivalence, tj. máme dané příslušné pořadí maximálních klik $Q_1, \dots, Q_t$. Nechť x je libovolný vrchol. Dle předpokladů existují indexy $i \le j$ takové, že x patří právě do klik $Q_i, Q_{i+1}, \dots, Q_j$. Přiřaďme vrcholu x interval $I_x = [i, j]$. Snadno ověříme, že toto je intervalová reprezentace grafu G. □

Věta 3. *Intervalové grafy lze rozpoznávat v lineárním čase.*

Důkaz. Mějme graf $G = (V, E)$ s n vrcholy a m hranami. Jak už víme z předchozích přednášek, lze v lineárním čase ověřit, že G je chordální, a případně zkonstruovat jeho perfektní eliminační schéma $x_1, \ldots, x_n$. Protože intervalové grafy jsou podtřída chordálních grafů, předpokládejme, že G je chordální a mějme jeho PES.

Naším cílem je ověřit podmínku na uspořádání maximálních klik z Věty 2.

Označme N_i množinu vrcholů, která obsahuje vrchol x_i spolu se všemi jeho levými sousedy, jinými slovy:

$$N_i := \{x_i\} \cup \{x_j,\, j < i \,\&\, \{x_j, x_i\} \in E\}.$$

Z vlastností PES plyne, že každá množina N_i je klika, navíc snadno nahlédneme, že všechny maximální kliky patří mezi $N_1, \ldots, N_n$.

Není také těžké poznat, že některá klika N_i není maximální. Ostrá inkluze $N_i \subsetneq N_j$ platí právě tehdy, když x_i je levý soused x_j a počet levých sousedů x_i je roven počtu levých sousedů x_j ležících vlevo od x_i. Na základě této vlastnosti v čase $O(n + m)$ najdeme seznam všech maximálních klik grafu G.

Následně pro každý vrchol x_i najdeme množinu těch maximálních klik, které obsahují x_i. Počet takovýchto klik je nejvýše o jedna větší než počet pravých sousedů x_i, takže celková velikost těchto množin je $O(m + n)$.

Otestovat existenci posloupnosti klik z Věty 2 tedy odpovídá řešení instance problému Tɪ o velikosti $O(n + m)$. Jelikož lze problém Tɪ vyřešit v lineárním čase, máme lineární algoritmus pro rozpoznávání intervalových grafů. $\square$