Řešení úlohy č. 2
Navazování spojení

V tejto úlohe sme pracovali s orientovaným grafom. Skôr než začneme ukazovať postup nášho riešenia, zavedieme si niekoľko pojmov a ukážeme niektoré z vlastností orientovaných grafov, ktoré nás budú zaujímať. Vlastnosti, ktoré si v tejto časti ukážeme sú základné a ak ich už poznáte, tak sa nič nové nedozviete a kľudne túto časť preskočte. V tejto časti zároveň zavedieme značenie, ktoré budeme používať ďalej.

Základné definície a vlastnosti skúmaných grafov

Keďže pracujeme s orientovaným grafom, musíme si ho aj formálne zaviesť.

Definícia 1. *Orientovaný graf G=(V, E) je usporiadaná dvojica, kde V je množina vrcholov grafu a E je množina orientovaných hrán* $E = \{(u, v)|u, v \in V\}$.

Budeme pracovať s niekoľkými funkciami deg, deg_+ a deg_-. Funkcie deg_+, resp. deg_- vrátia pre každý vrchol počet hrán, ktoré vedú do tohto vrcholu, respektíve z tohto vrcholu. Funkcia deg vráti súčet výsledkov týchto dvoch funkcií.

Definícia 2. *Buď orientovaný graf G=(V, E). Každý vrchol* $u \in V$*, pre ktorý platí, že* $deg_+(u) = 0$*, respektíve* $deg_-(u) = 0$*, nazývame zdroj, respektíve stok. Vrchol, pre ktorý platí* $deg(u) = 0$ *nazývame izolovaný vrchol.*

Definícia 3. *Buď orientovaný graf G=(V, E). Postupnosť hrán tohto grafu, ktoré tvoria orientovaný kružnicu nazývame cyklus. Graf, ktorý neobsahuje žiadnu orientovanú kružnicu, nazývame acyklickým grafom.*

Veta 1. *Buď acyklický orientovaný graf G=(V, E). V takomto grafe sa nachádza aspoň jeden zdroj a aspoň jeden stok.*

Dôkaz

Vetu dokážeme sporom. Majme teda acyklický orientovaný graf G=(V, E). Prepokladajme teda, že tento graf neobsahuje žiaden stok. Vyberieme si teda ľubovoľný vrchol a budeme prechádzať graf do hĺbky a budeme si pamätať, ktoré vrcholy sme už navštívili. V grafe neexistuje stok, takže vždy budeme mať nejakého nasledovníka. Vzhľadom na to, že graf má konečný počet vrcholov, nutne nám niekedy musia "dôjsť" a my sa skôr či neskôr vrátime do vrcholu, ktorý sme už navštívili, čím objavíme v našom grafe cyklus. Tu však nastáva spor s naším predpokladom, že graf je acyklický. Dôkaz pre zdroj je ekvivalentný, stačí pri prehľadávaní do hĺbky ísť v opačnom poradí hrán. □

Definícia 4. *Buď orientovaný graf G=(V, E). Ak pre* $\forall u, v \in V$ *existuje cesta z u do v a aj z v do u, povieme, že je graf G silne súvislý.*

Rekapitulácia zadania

Buď acyklický orientovaný graf. Doplňte graf o orientované hrany, tak, aby platilo, že výsledný graf je silne súvislý. Popíšte stratégiu pridávania hrán a odhadnite ich počet. Snažte sa aby tento počet bol čo najmenší.

V rozšírení je cieľom dosiahnuť to isté, ale s grafom o ktorom vieme iba to, že je orientovaný, a teda môže obsahovať cykly.

Pozorovania pre našu úlohu

Nájsť nejaké riešenie pre túto úlohu je jednoduché, stačí pridať orientovanú hranu medzi každými dvoma vrcholmi a jasne vidieť, že existuje cesta zo každého vrcholu do každého vrcholu a naspäť. Hrany však nie sú zadarmo a preto potrebujeme niekoľko pozorovaní, ktoré nám umožnia znížiť počet pridaných hrán.

Výpočet lower boundy

Skôr ako však budeme rozoberať tieto pozorovania, odvodíme si minimálny počet hrán, ktoré musíme pridať, aby sme túto úlohu splnili. Zavedieme si predtým značenie:

- m - počet zdrojov, ktoré nie sú izolované vrcholy
- n - počet stokov, ktoré nie sú izolované vrcholy
- r - počet izolovaných vrcholov

Vidíme, že sa samostatne pozeráme na izolované vrcholy a na zdroje a stoky, napriek tomu, že podľa definície sú aj izolované vrcholy zdrojmi a aj stokmi. Takéto orwellovské delenie zdrojov a stokov na rovné a rovnejšie zavádzame pretože nám uľahčia niektoré kroky pri popise našej stratégie pridávanie hrán a odhadu počtu pridaných hrán. Zatiaľ teda to delenie berte ako fakt, ktorý sa nám zíde neskôr. Keď budeme odteraz hovoriť o rýdzich zdrojoch, respektíve rýdzich stokoch, budeme myslieť tie, ktoré nie sú izolovanými vrcholmi. Ak spomenieme stoky a zdroje, myslíme tie podľa pôvodnej definície.

Nutnou podmienkou toho, aby existovala cesta medzi každým párom vrcholov je to, aby do každého vrcholu viedla aspoň jedna hrana a zároveň aby aspoň jedna viedla aj z každého vrcholu. Takto môžeme ľahko spočítať koľko potrebujeme pridať vstupných hrán a koľko výstupných. Vstupných potrebujeme pridať $m+r$ a výstupných $n+r$, takže celkovo potrebujeme pridať aspoň $max(m+r, n+r)$ hrán. Získali sme teda lower boundu a vieme, že s menším počtom pridaných hrán nevieme získať riešenie.

Spôsoby prepojení

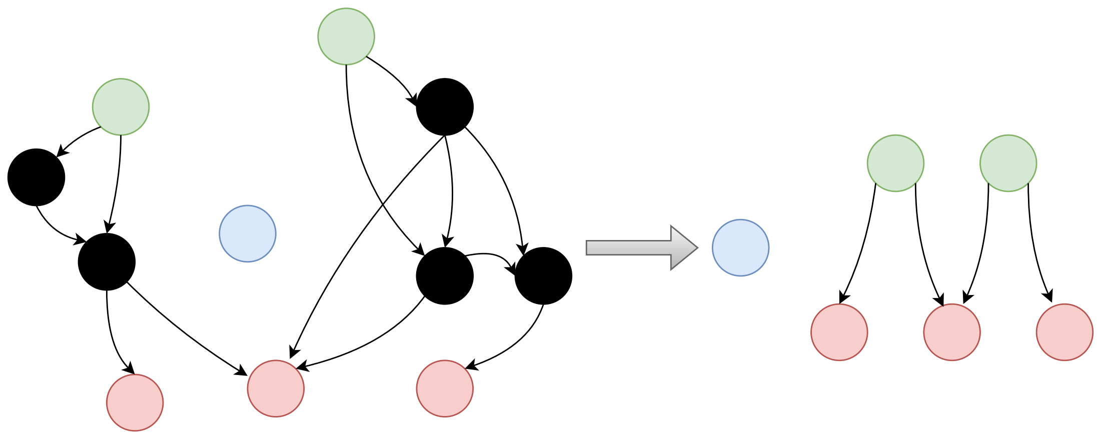
Ukážeme si, aké hrany nám stačí pridávať. Dá sa ukázať, že pridanú hranu sa nám vždy oplatí viesť iba zo stoku do zdroja. Dá sa vidieť, že ak by sme viedli hranu z vrcholu, ktorý nie je stokom, tak v prípade, že túto hranu budeme viesť z jeho nasledovníka namiesto nášho vrcholu, neprídeme o žiaden pár vrcholov medzi ktorými takto zanikne cesta. Postupne teda prídeme až do stoku. Ekvivalentne dokážeme pre zdroje. Pre izolované vrcholy je pozorovanie samozrejmé, keďže tie musíme určite napojiť na zvyšok grafu.

Stačí nám teda ukázať, že existuje cesta z každého zdroju do každého stoku a späť a toto bude určite platiť aj pre zvyšok grafu.

Popis postupu

Budeme teda pracovať iba so zdrojmi, stokmi a izolovanými vrcholmi. Náš graf si zjednodušíme na graf v ktorom sa budú nachádzať iba zdroje, stoky a izolované vrcholy. Hrana medzi zdrojom a stokom sa bude nachádzať v tomto novom grafe v prípade, že v pôvodnom grafe existuje cesta zo zdroju do stoku medzi ktoré chceme hranu pridať.

Odteraz teda pracujeme iba s týmto zjednodušeným grafom. Naše pridávanie hrán bude pozostávať z niekoľkých krokov. Najskôr si však vytvoríme páry rýdzich zdrojov a rýdzich stokov tak, aby platilo:

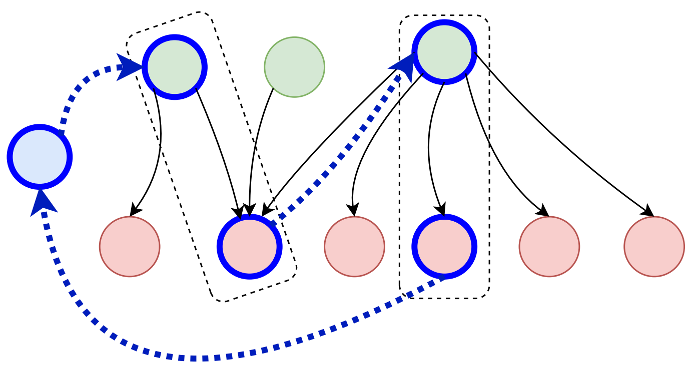


Ukážka zjednodušeného grafu. Zelenou sú označené rýdze zdroje, červenou rýdze stoky a modrou izolované vrcholy. Čierne sú všetky ostatné vrcholy, ktoré sa v našom zjednodušenom grafe nebudú nachádzať.

- rýdzi zdroj každého páru vie v pôvodnom grafe dosiahnuť odpovedajúci rýdzi stok
- každý rýdzi zdroj, ktorý sa nenachádza v žiadnom páre vie dosiahnuť nejaký rýdzi stok, ktorý sa v páre nachádza
- každý rýdzi stok, ktorý sa nenachádza v žiadnom páre je dosiahnuteľný nejakým rýdzim zdrojom, ktorý sa v páre nachádza

Takéto párovanie vždy existuje. Rýchlo popíšeme naivný spôsob ako takéto párovanie získať. Majme rýdzi zdroj, ktorý nie je súčasťou žiadneho páru. Ak vedie z neho hrana do rýdzeho stoku, ktorý sa už v nejakom páre nachádza, spĺňa tento rýdzi zdroj podmienky a my pokračujeme ďalej. Ak vedie hrana do rýdzeho stoku, ktorý sa už v nejakom páre nachádza, tak sú tiež splnené podmienky. Ekvivalentne pracujeme so stokmi.

Následne vieme zabezpečiť silnú súvislosť pridaním niekoľkých skupín hrán. V prvom kroku vytvoríme z párov a izolovaných vrcholov jeden cyklus. To dosiahneme správnym prepojením stokov na zdroj iného páru, respektíve na izolovaný vrchol. Celkový počet hrán, ak počet párov označíme p, ktoré takto pridáme je $p + r$. Týmto máme teda zabezpečené, že existuje cesta medzi všetkými zdrojmi, stokmi a izolovanými vrcholmi na tejto kružnici.



Ukážka párovanie na zjednodušenom grafe, ktoré spĺňa naše podmienky. Páry sú vyznačené obdĺžnikmi. Modré hrany vyznačujú hrany pridané v prvom kroku. Vrcholy s modrým okrajom sú tie, ktoré tvoria kružnicu vytvorenú v prvom kroku.

Ešte stále nám však ostalo niekoľko zdrojov a stokov, ktoré nie sú súčasťou žiadneho páru. V druhom kroku vytvoríme hrany od zvyšných stokov do zvyšných zdrojov ľubovoľným spôsobom ale bez opakovania, teda vlastne znovu vytvoríme páry. Toto budeme robiť až kým nám nedojdú buď zdroje alebo stoky. Všetky vrcholy takto prepojené vlastne napojíme na veľkú kružnicu vytvorenú v prvom kroku, teda budú z nej dosiahnuteľné a zároveň ju budú vedieť dosiahnuť, teda budú existovať potrebné cesty. Toto bude platiť vďaka našej druhej a tretej podmienke pri vytváraní párov.

Celkový počet hrán, ktoré v tomto kroku pridáme je $min(m-p, n-p)$.

Na záver nám zostali už iba zdroje alebo iba stoky. Tie teraz iba jednou hranou napojíme na ľubovoľný vrchol, ktorý už máme vyriešený. Smer pridanej hrany samozrejme zodpovedá, či pracujeme so stokom alebo zdrojom. Počet hrán, ktoré takýmto spôsobom pridáme v tomto kroku je $max(m-p, n-p) - min(m-p, n-p)$.

Použitý počet hrán

Na zistenie počtu použitých hrán nám stačí iba sčítať počet pridaných hrán v každom kroku a tento výraz zjednodušiť:

$$p + r + min(m-p, n-p) + max(m-p, n-p) - min(m-p, n-p)$$

$$p + r + max(m-p, n-p)$$

$$max(m, n) + r$$

Vidíme, že počet hrán, ktorý sme pridali sa rovná dokázanému najmenšiemu možnému počtu hrán. Náš postup teda pre každý orientovaný acyklický graf vráti optimálne riešenie.

Odhad časovej zložitosti

Náš postup sa skladá z niekoľkých krokov, ktoré sme nepopisovali priamo, iba sme popísali podproblémy, ktoré potrebujeme vyriešiť. Prvý potrebný krok je detekcia zdrojov a stokov a zjednodušenie na graf zložený iba z nich. Detekciu zdrojov a stokov a následné vytvorenie popísaného zjednodušeného grafu vieme spraviť upraveným DFS v čase $O(|V| + |E|)$. Vytvoriť páry zdrojov a stokov podľa našich podmienok vieme taktiež spraviť pomocou opakovaného spúšťania DFS na zjednodušenom grafe, prípadne vieme použiť aj nejaký iný greedy prístup. Podstatné pre nás je však, že to zvládame aj spolu s pridaním potrebných hrán v prvom kroku robiť v čase $O(|V| + |E|)$. Druhý a tretí krok sú triviálne a budú oba trvať $O(|V|)$. Celková náročnosť pre základné znenie problému je teda $O(|V| + |E|)$.

Riešenie pre rozšírenie

Pri rozšírení využijeme z veľkej časti naše pôvodné riešenie. V pôvodnom grafe nájdeme podgrafy, v ktorých existuje cesta z každého vrcholu do každého iného vrcholu a naspäť. Tieto sa nazývajú silno súvislé komponenty. Naše predspracovanie bude spočívať v kondenzácii takýchto komponent do jedného vrcholu a následne práca s takýmto grafom. Výsledkom takejto kondenzácie bude nutne acyklický orientovaný graf na ktorý môžeme použiť náš predchádzajúci postup.

Našťastie už existujú algoritmy ako takúto kondenzáciu vykonať, napríklad Tarjanov algoritmus[1], ktorý identifikuje silno súvislé komponenty v $O(|V| + |E|)$ čase. Celková asymptotická náročnosť nášho riešenie sa teda nezvýši.

[1] Viac si môžete prečítať napríklad v knihe Průvodce labyrintem algoritmů

Prehľad bodovania

Bodovaná oblasť	Počet bodov
Korektnosť postupu	1 bod
Práca iba so zdrojmi a stokmi	1 bod
Hrany vedené iba zo stokov do zdrojov	1 bod
Ukážka lower boundy počtu hrán	1 bod
Počet hrán postupu	2 body
Zložitosť postupu	2 body
Riešenie rozšírenia	2 body
Spolu	10 bodov