

## Report – Zimný semester

Cieľom projektu bolo vymyslieť a implementovať efektívne riešenie nasledujúceho algoritmického problému:

Máme daný orientovaný acyklický graf (DAG)  $G$  s ohodnotenými vrcholmi a vrcholy  $s, t$ . Skóre vrcholu  $u$  označujeme  $r(u)$ . Nech máme cesty  $P_1, P_2$ . Ich skóre je dané ako

$$(1 + \alpha) \sum_{u \in (P_1 \cap P_2)} r(u) + \sum_{u \in (P_1 \cap P_2), r(u) < 0} r(u) + \sum_{u \in (P_1 \cup P_2) - (P_1 \cap P_2)} r(u)$$

Úlohou je nájsť dvojicu ciest z  $s$  do  $t$  s maximálnym skóre.

---

### Report mojej práce počas zimného semestra

1. Dokázal som, že sa dá problém previesť na problém *min-cost flow*.
2. Implementoval som GFA parser grafov na načítanie vstupu.
3. Implementoval som mierne upravený Suurballeov algoritmus, ktorý rieši zadaný problém, ak sú skóre vrcholov nezáporné.

Pri implementácii som ošetril chyby, ktoré by vznikli kvôli práci s necelými číslami. Výsledný program funguje správne pre  $\alpha$  určené 5 desatinnými miestami a vstupný graf musí byť zadaný v GFA formáte, alebo v jednoduchom textovom formáte s explicitne zadanými skóre vrcholov, ktorý som si vymyslel pre účely testovania. To kvôli tomu, že pri GFA grafoch by sa skóre malo vypočítať z dát na vstupe, ale zatiaľ sme sa nedohodli na skórovacej funkcii. Pri GFA formáte teda skóre zatiaľ priradujem náhodne a v letnom semestri program doplním o skórovaciu funkciu.

---

### Transformácia problému na *min-cost flow*

*Min-cost flow* je optimalizačný a rozhodovací problém hľadania najlacnejšieho spôsobu, ako poslať určité množstvo toku cez sieť. Podľa pôvodného grafu a danej  $\alpha$  skonštruujeme sieť, kde cena najlacnejšieho spôsobu, ako poslať tok veľkosti 2, je riešenie pôvodného problému.

Cenu hrany  $(u, v)$  v sieti  $N$  označujeme  $c((u, v))$ . Sieť  $N$  vytvoríme nasledovne:

Z každého vrcholu  $u$  nahradíme štyrmi novými vrcholmi  $u_a, u_b, u_c, u_d$  a hranami  $(u_a, u_c), (u_b, u_d)$ . Ak  $r(u) \geq 0$ , položíme  $c(u_b, u_d) = -\alpha \cdot r(u)$ , inak  $c(u_b, u_d) = -(1 + \alpha) \cdot r(u)$ . Bez ohľadu na  $r(u)$  položíme  $c((u_a, u_c)) = -r(u)$ .

Pre každú hranu  $(u, v)$  z  $G$  následne pridáme hrany  $(u_c, v_a), (u_c, v_b), (u_d, v_a), (u_d, v_b)$  s cenami 0. Každéj hrane vo vytvorenej sieti nastavíme kapacitu 1.

Pre jednoduchosť tiež predpokladajme, že  $r(s) = r(t) = 0$ , teda vrcholy  $s$  a  $t$  nemusíme rozdeľovať na 4, ale ostanú ako jeden vrchol aj v  $N$ .

Dokážeme, že pôvodný problém vieme vyriešiť nájdením **najlacnejšieho toku veľkosti 2 z  $s$  do  $t$**  v sieti  $N$ .

### Označenia

Cenu hrany  $(u, v)$  v sieti  $N$  označujeme  $c((u, v))$ .

Hovoríme, že vrchol zo siete  $N$  korešponduje vrcholu z  $G$ , ak pri konštrukcii  $N$  vznikol z tohto vrcholu. Teda vrcholy  $u_a, u_b, u_c, u_d$  zo siete  $N$  korešpondujú vrcholu  $u$  z  $G$ .

Tiež budeme cestou  $P = u_1, u_2, \dots, u_n$  myslieť takú cestu (v  $G$  alebo v  $N$ ), že existujú hrany  $(s, u_1)$  a  $(u_n, t)$ .

## Dôkaz

**Ak existuje dvojica ciest  $(P_1, P_2)$  z  $s$  do  $t$  v  $G$  s cenou  $C$ , tak existuje tok veľkosti 2 v  $N$  s cenou  $-C$ .**

Nech  $P_1 = u_1, u_2, \dots, u_n$ ,  $P_2 = v_1, v_2, \dots, v_m$ .

Nech  $P'_1 = u_{1,a}, u_{1,c}, u_{2,a}, u_{2,c}, \dots, u_{n,a}, u_{n,c}$  v sieti  $N$ . Táto cesta zjavne existuje, keďže z každého vrcholu  $u$  vznikli štyri vrcholy  $u_a, u_b, u_c, u_d$ , kde je hrana  $(u_a, u_c)$  v  $N$ , a ak je hrana  $(u, v)$  v  $G$ , tak je v  $N$  hrana  $(u_c, v_a)$ . Cena hrany  $(u_{i,a}, u_{i,c}) = -r(u_i)$  a cena hrany  $(u_{i,c}, u_{(i+1),a}) = 0$ , teda cena tejto cesty v  $N$  je

$$- \sum_{u \in P'_1} r(u).$$

Nech  $P'_2 = v_{1,x}, v_{1,y}, v_{2,x}, v_{2,y}, \dots, v_{m,x}, v_{m,y}$  a ak  $v \in P_1$ , tak  $x = b, y = d$ , inak  $x = a, y = c$ . Táto cesta tiež zjavne existuje, keďže v  $N$  je hrana medzi  $(v_a, v_c)$  aj  $(v_b, v_d)$  a ak  $(u, v)$  je hrana v  $G$ , tak  $(u_x, v_y)$  sú hrany v  $N$  pre  $x \in \{c, d\}$ ,  $y \in \{a, b\}$ . Keďže niektoré hrany v ceste môžu byť typu  $(v_b, v_d)$ , musíme rozlíšiť vrcholy s kladným a záporným skóre. Potom už podobným spôsobom ako pre  $P'_1$  vieme spočítať, že cena tejto cesty je

$$- \sum_{u \in (P_2 - P_1)} r(u) - \alpha \cdot \sum_{u \in (P_1 \cap P_2), r(u) \geq 0} r(u) - (1 + \alpha) \cdot \sum_{u \in (P_1 \cap P_2), r(u) < 0} r(u).$$

Z konštrukcie  $P'_2$  je zrejmé, že je vrcholovo disjunktná s  $P'_1$ , teda spolu tvoria validný tok. Kapacita každej hrany je jednotková, teda cesty  $P'_1, P'_2$  tvoria tok veľkosti 2.

Po spojení výsledkov tak dostaneme cenu toku

$$-(1 + \alpha) \sum_{u \in (P_1 \cap P_2)} r(u) - \sum_{u \in (P_1 \cap P_2), r(u) < 0} r(u) - \sum_{u \in (P_1 \cup P_2) - (P_1 \cap P_2)} r(u),$$

čo je presne  $-C$ .

**Ak je  $C$  cena najlacnejšieho toku veľkosti 2 v  $N$ , tak existuje dvojica ciest  $(P_1, P_2)$  z  $s$  do  $t$  s cenou  $-C$ .**

Je známe, že ak existuje vo váženej sieti s celočíselnými kapacitami tok nejakej veľkosti, tak existuje aj najlacnejší tok tejto veľkosti, kde každá hrana má celočíselný tok ([zdroj](#)).

V našom prípade, ak existuje tok veľkosti 2, tak existuje aj najlacnejší tok veľkosti 2, kde tok na každej hrane je 0 alebo 1. Bez ujmy na všeobecnosti predpokladajme, že takýto najlacnejší tok s cenou  $C$  sme zvolili.

Keďže tok je veľkosti 2 a kapacita každej hrany je práve 1, tak zo zdroja  $s$  idú práve dve hrany  $(s, u_1), (s, v_1)$  s jednotkovým tokom. Zo zákonov zachovania toku a toho, že každá hrana má kapacitu 1, vyplýva, že musí existovať práve jedna postupnosť vrcholov  $P'_1 = u_1, u_2, \dots, u_k, t$  taká, že pre všetky  $u_i$  platí  $f(u_i, u_{i+1}) = 1$ . A tiež jednoznačne určená postupnosť  $P'_2 = v_1, v_2, \dots, v_l, t$ .

Z konštrukcie vidno, že každý vrchol v sieti okrem  $s$  a  $t$  má *in-degree* alebo *out-degree* najviac 1, teda prítok aj odtok z každého vrcholu je  $\leq 1$ . Keďže na  $P'_1$  aj  $P'_2$  máme jednotkový tok, vidíme, že  $P'_1$  a  $P'_2$  musia byť vrcholovo disjunktné.

Z konštrukcie siete je zrejmé, že vrcholy v  $G$  korešpondujúce k postupnosti  $P_1 = u_1, u_3, \dots, u_{k-1}$  tvoria cestu v  $G$ . Taktiež  $P_2 = v_1, v_3, \dots, v_{l-1}$  tvoria cestu v  $G$ . Ukážeme, že tieto cesty majú cenu opačnú k cene nami zvoleného toku.

Z konštrukcie siete je zrejmé, že na ľubovoľnej ceste od  $s$  sú v párnych vzdialenostiach hrany s cenou 0 (tie, ktoré vznikli z hrán v  $G$ ) a na nepárnych tie, ktoré vznikli zo „rozštvorenia“ vrcholov. Pri vyjadrovaní ceny toku teda stačí počítat súčet cien hrán v množinách  $E_1 = \{(u_1, u_2), (u_3, u_4), \dots, (u_{k-1}, u_k)\}$ ,  $E_2 = \{(v_1, v_2), \dots, (v_{l-1}, v_l)\}$ .

Pre každú z týchto hrán jej incidentné vrcholy korešpondujú k rovnakému vrcholu v  $G$ . Budeme preto ďalej hovoriť, že hrane korešponduje vrchol v  $G$ . Každá hrana má teda svoj korešpondujúci vrchol. Môžeme teda postupne pre vrcholy spočítať ceny hrán, ktoré im korešpondujú, a dostaneme takto súčet hrán v najlacnejšom toku.

Zjavne  $P_1 \cup P_2$  je množina vrcholov, ktoré stačí uvažovať, keďže iba tie korešpondujú aspoň jednej hrane v  $E_1$  alebo  $E_2$ . Taktiež platí, že každej hrane v  $E_1$  korešponduje iný vrchol v  $G$ , keďže  $u_1, u_3, \dots, u_{k-1}$  tvoria cestu (na nej sa vrcholy neopakujú), a teda k vrcholu z  $G$  môže korešpondovať najviac jedna hrana z  $E_1$  a najviac jedna z  $E_2$ .

Spočítajme ceny pre hrany, ktoré korešpondujú vrcholom v  $P_1 \cap P_2$ . Nech  $w \in P_1 \cap P_2$ . Potom existujú dvojice korešpondujúcich vrcholov k  $w$  v  $P'_1$  aj  $P'_2$ , a teda aj hrana v  $E_1$  aj  $E_2$ . Nech sú tieto hrany  $(u_1, u_2) \in E_1$  a  $(v_1, v_2) \in E_2$ . Z toho, že  $P'_1$  a  $P'_2$  sú vrcholovo disjunktné, vyplýva, že tieto hrany nie sú rovnaké.

Jediné možné korešpondujúce hrany k  $w$  v celej sieti sú  $(w_a, w_c)$  a  $(w_b, w_d)$ . Z toho vyplýva, že sme museli použiť obe z nich.

Ak  $r(w) \geq 0$ , ich spoločná cena je

$$c((w_a, w_c)) + c((w_b, w_d)) = -(1 + \alpha) \cdot r(w).$$

Ak  $r(w) < 0$ , ich spoločná cena je

$$-(2 + \alpha) \cdot r(w).$$

Spočítajme ceny pre hrany, ktoré korešpondujú vrcholom v  $(P_1 \cup P_2) - (P_1 \cap P_2)$ . Nech  $w \in P_1 - P_2$ . Vieme, že  $w \notin P_2$ , teda neexistuje korešpondujúca hrana v  $E_2$ . Vieme, že  $w \in P_1$ , teda existuje korešpondujúca hrana k  $w$  v  $E_1$  a tá je jediná.

Nech je táto hrana  $(u_1, u_2)$ . Ukážeme sporom, že  $(u_1, u_2) = (w_a, w_c)$ . Pre spor predpokladajme, že  $(u_1, u_2) = (w_b, w_d)$ . Nech je  $p_1$  predchodca  $w_b$  na ceste toku z  $s$  do  $t$  a  $p_2$  nasledovník  $w_d$  na tejto ceste. Ak  $r(w) \geq 0$ ,  $c(w_b, w_d) = -\alpha \cdot r(w)$ , inak  $c(w_b, w_d) = -(1 + \alpha) \cdot r(w)$ . To je v oboch prípadoch  $> c(w_a, w_c) = -r(w)$ .

Teda existuje hrana  $(w_a, w_c)$ , kde  $c((w_a, w_c)) < c((w_b, w_d))$ , existuje aj hrana  $(p_1, w_a)$ , kde  $c((p_1, w_a)) = 0$ , a hrana  $(w_c, p_2)$ , kde  $c((w_c, p_2)) = 0$ . Všetky tieto hrany majú nulový tok a kapacitu 1, teda jednotkový tok, ktorý máme na popísanej ceste z  $p_1$  do  $p_2$ , môžeme nahradiť novým lacnejším tokom. To je v rozpore s tým, že náš tok je najlacnejší možný.

Z toho vyplýva, že hrana, ktorá korešponduje nejakému vrcholu  $w$  z tejto množiny, má cenu  $-r(w)$ .

Teda spolu dostávame cenu

$$-(1 + \alpha) \sum_{u \in (P_1 \cap P_2)} r(u) - \sum_{u \in (P_1 \cap P_2), r(u) < 0} r(u) - \sum_{u \in (P_1 \cup P_2) - (P_1 \cap P_2)} r(u),$$

čo je presne opačná cena ciest  $P_1, P_2$ .

Z dokázaných tvrdení vyplýva, že **najlacnejší tok veľkosti 2** v skonštruovanej sieti  $N$  z  $s$  do  $t$  bude zodpovedať **najväčšiemu možnému skóre dvojice ciest** v grafe  $G$ .

## Efektívny algoritmus na riešenie problému, kde $r(u) \geq 0$ pre všetky $u$

Pre nezáporné skóre vrcholov vieme použiť základné idey [Suurballeovho](#) algoritmu. Upravená verzia algoritmu vyzerá nasledovne:

1. Transformujeme daný graf na sieť  $N$  vyššie popísaným spôsobom.

2. Keďže sieť  $N$  je acyklická, tak štandardným dynamickým programovaním nájdeme v sieti  $N$  dĺžku najkratšej cesty z  $s$  do každého iného vrcholu siete. Dĺžku najkratšej cesty z  $s$  do  $u$  budeme ďalej označovať  $d(s, u)$ .
3. Hranám priradíme nové ceny tak, že pre každú hranu v  $N$   $(u, v)$  položíme  $c'((u, v)) = c((u, v)) + d(s, u) - d(s, v)$ . Tak ako v Suurballeovom algoritme, aj tu bude platiť, že  $c'((u, v)) \geq 0$  pre každú hranu  $(u, v)$ .
4. Vytvoríme reziduálnu sieť  $N_r$  podľa najkratšej cesty z  $s$  do  $t$ , podobne ako pri Ford–Fulkersonovej metóde. Ceny nových reziduálnych hrán budú opačné k pôvodným cenám. Reziduálne hrany ale pribudnú iba na najkratšej ceste z  $s$  do  $t$  a pre každú hranu  $(u, v)$  tejto cesty platí, že  $c'(u, v) = 0$ , teda ceny hrán v  $N_r$  budú nezáporné.
5. Nájdeme najkratšiu zlepšujúcu cestu z  $s$  do  $t$  v reziduálnej sieti  $N_r$  pomocou Dijkstrovho algoritmu a podľa nej upravíme tok. Nový tok veľkosti 2 určuje dvojicu ciest, ktoré už jednoznačne určujú dvojicu ciest v pôvodnom grafe  $G$ , ktoré sú riešením pôvodného problému.

Časová zložitosť tohto algoritmu je  $O((n + m) \log m)$ .

Implementácia spolu s výsledným funkčným programom: [odkaz na GitHub](#)