4. Gomory-Hu Trees

Cílem této kapitoly je popsat datovou strukturu, která velice kompaktně popisuje minimální st-řezy pro všechny dvojice vrcholů s, t v daném neorientovaném grafu. Tuto strukturu poprvé popsali Gomory a Hu v článku [1].

Zatím umíme nalézt minimální st-řez pro zadanou dvojici vrcholů v neorientovaném grafu v čase $\tau = \mathcal{O}(n^{2/3}m)$ pro jednotkové kapacity, $\mathcal{O}(n^2m)$ pro obecné. Nalézt minimální st-řez pro každou dvojici vrcholů bychom tedy dokázali v čase $\mathcal{O}(n^2\tau)$. Tento výsledek budeme chtít zlepšit.

Značení: Máme-li graf (V, E) a $U \subseteq V$, $\delta(U)$ značí hrany vedoucí mezi U a $\overline{U}$, formálně tedy $\delta(U) = E \cap ((U \times \overline{U}) \cup (\overline{U} \times U))$. Kapacitu řezu $\delta(W)$ budeme značit $d(W)$ a $r(s,t)$ bude kapacita nejmenšího st-řezu.

Pozorování: Minimální řez rozděluje graf jen na dvě komponenty (všimněte si, že pro separátory nic takového neplatí) a každý minimální řez je tím pádem vždy možné zapsat jako $\delta(W)$ pro nějakou množinu $W \subset V$.

Gomory-Hu Tree

Definice: *Gomory-Hu Tree* (dále jen GHT) pro neorientovaný nezáporně ohodnocený graf $G = (V, E)$ je strom $T = (V, F)$ takový, že pro každou hranu $st \in F$ platí: Označíme-li K_1 a K_2 komponenty lesa $T \setminus st$, je $\delta(K_1) = \delta(K_2)$ minimální st-řez. [Pozor, F nemusí být podmnožina původních hran E.]

Další značení: Pro $e \in F$ budeme řezem $\delta(e)$ označovat řez $\delta(K_1) = \delta(K_2)$ a $r(e)$ bude jeho kapacita.

K čemu takový GHT je (existuje-li)? To nám poví následující věta:

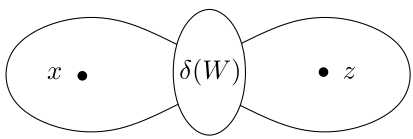
Věta (o využití GHT): Buď T libovolný GHT pro graf G a mějme dva vrcholy s a t. Dále nechť P je cesta v T mezi vrcholy s a t a e je hrana na cestě P s minimálním $r(e)$. Pak $\delta(e)$ je minimální st-řez v G.

Důkaz: Nejprve si dokážeme jedno drobné lemmátko:

Lemmátko: Pro každou trojici vrcholů x, y, z platí, že:

$$r(x,z) \geq \min(r(x,y), r(y,z)).$$

Důkaz: Buď W minimální xz-řez.



Vrchol y musí být v jedné z komponent, Pokud je v komponentě s x, pak $r(y,z) \leq d(W)$, protože $\delta(W)$ je také yz-řez. Pokud v té druhé, analogicky platí $r(x,y) \leq d(W)$. ♡

Zpět k důkazu věty: Chceme dokázat, že $\delta(e)$ je minimální st-řez. To, že je to nějaký řez, plyne z definice GHT. Minimalitu dokážeme indukcí podle délky cesty P:

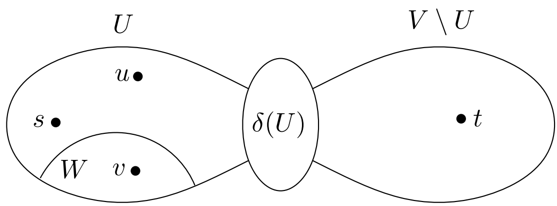
- $|P| = 1$: Hrana e je v tomto případě přímo st, takže i minimalita plyne z definice GHT.
- $|P| > 1$: Cesta P spojuje vrcholy s a t, její první hranu označme sx. Naše právě dokázané lemmátko říká, že $r(s,t) \geq \min(r(s,x), r(x,t))$. Určitě je pravda, že $r(s,x) \geq r(e)$, protože e byla hrana cesty P s nejmenším $r(e)$. To, že $r(x,t) \geq r(e)$, plyne z indukčního předpokladu, protože cesta mezi x a t je kratší než cesta P. Dostáváme tak, že $r(s,t) \geq \min(r(s,x), r(x,t)) \geq r(e)$. ♡

Pokud dokážeme GHT sestrojit, nalézt minimální st-řez pro libovolnou dvojici vrcholů dokážeme stejně rychle jako nalézt hranu s nejmenší kapacitou na cestě mezi s a t v GHT. K tomu můžeme použít například Sleator-Tarjanovy stromy, které tuto operaci dokážou provést v amortizovaném čase $\mathcal{O}(\log n)$, nebo můžeme využít toho, že máme spoustu času na předvýpočet, a minimální hrany si pro každou dvojici prostě přichystat předem. Také lze vymyslet redukci na problém nalezení společného předchůdce vrcholů ve stromě (nebude to GHT) a použít jedno z řešení tohoto problému.

Konstrukce GHT

Nyní se naučíme GHT konstruovat, čímž také rozptýlíme obavy o jejich existenci. Nejprve však budeme potřebovat jedno užitečné lemma s hnusně technickým důkazem:

Hnusně technické lemma (HTL): Buďtež s,t vrcholy grafu (V,E), $\delta(U)$ minimální st-řez a $u \neq v$ dva různé vrcholy z U. Pak existuje množina vrcholů $W \subseteq U$ taková, že $\delta(W)$ je minimální uv-řez. ⟨1⟩

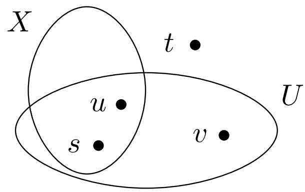


Důkaz: Nechť je $\delta(X)$ minimální uv-řez. BÚNO můžeme předpokládat, že $s \in U$ a $t \notin U$, $u \in X$ a $v \notin X$ a $s \in X$. Pokud by tomu tak nebylo, můžeme vrcholy přeznačit nebo některou z množin nahradit jejím doplňkem.

⟨1⟩ To důležité a netriviální je, že celá W leží v U.

Nyní mohou nastat následující dva případy:

a) $t \notin X$. Tehdy si všimneme, že platí:



$$d(U \cup X) \geq d(U), \qquad (1)$$

$$d(U \cap X) + d(U \cup X) \leq d(U) + d(X) \qquad (2)$$

První nerovnost plyne z toho, že $\delta(U \cup X)$ je nějaký st-řez, zatímco $\delta(U)$ je minimální st-řez. Druhou dokážeme rozborem případů.

Množinu vrcholů si disjunktně rozdělíme na $X \setminus U$, $X \cap U$, $U \setminus X$ a *ostatní*. Každý z řezů vystupujících v nerovnosti (2) můžeme zapsat jako sjednocení hran mezi některými z těchto skupin vrcholů. Vytvoříme tedy tabulku hran mezi čtyřmi označenými skupinami vrcholů a každému řezu z (2) označíme jemu odpovídající hrany. Protože je graf neorientovaný, stačí nám jen horní trojúhelník tabulky. Pro přehlednosti si označíme $L_1 = \delta(U \cap X), L_2 = \delta(U \cup X), P_1 = \delta(U)$ a $P_2 = \delta(X)$.

	$X \setminus U$	$X \cap U$	$U \setminus X$	*ostatní*
$X \setminus U$	—	L_1, P_1	P_1, P_2	L_2, P_2
$X \cap U$		—	L_1, P_2	L_1, L_2, P_1, P_2
$U \setminus X$			—	L_2, P_1
ostatní				—

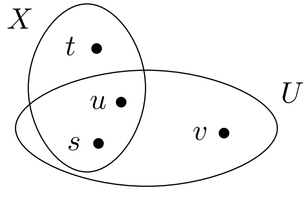
Vidíme, že ke každé hraně řezu na levé straně nerovnosti máme vpravo její protějšek a navíc hrany mezi $U \setminus X$ a $X \setminus U$ počítáme jenom vpravo. Nerovnost (2) tedy platí.

Nyní stačí nerovnosti (2) a (1) odečíst, čímž získáme:

$$d(U \cap X) \leq d(X),$$

což spolu s obrázkem dokazuje, že $\delta(U \cap X)$ je také minimální uv-řez.

b) $t \in X$. Postupovat budeme obdobně jako v předchozím případě. Tentokrát se budou hodit tyto nerovnosti:



$$d(X \setminus U) \geq d(U) \qquad (3)$$

$$d(U \setminus X) + d(X \setminus U) \leq d(U) + d(X) \qquad (4)$$

První platí proto, že $\delta(X \setminus U)$ je nějaký st-řez, zatímco $\delta(U)$ je minimální st-řez, druhou dokážeme opět důkladným rozborem případů.

Označme $L_1 = \delta(U \setminus X), L_2 = \delta(X \setminus U), P_1 = \delta(U)$ a $P_2 = \delta(X)$ a vytvořme tabulku:

	$X \setminus U$	$X \cap U$	$U \setminus X$	*ostatní*
$X \setminus U$	—	L_2, P_1	L_1, L_2, P_1, P_2	L_2, P_2
$X \cap U$		—	L_1, P_2	P_1, P_2
$U \setminus X$			—	L_1, P_1
ostatní				—

Stejně jako v předchozím případě nerovnost (4) platí. Odečtením (4) a (3) získáme:

$$d(U \setminus X) \leq d(X),$$

z čehož opět dostaneme, že $\delta(U \setminus X)$ je také minimální uv-řez. ♡

Nyní se konečně dostáváme ke konstrukci GHT. Abychom mohli používat indukci, zavedeme si trochu obecnější GHT.

Definice: Mějme neorientovaný graf (V, E). *Částečný Gomory-Hu Tree* (alias ČGHT) pro podmnožinu vrcholů $R \subseteq V$ je dvojice $((R, F), C)$, kde (R, F) je strom a množina $C = \{C(r) \mid r \in R\}$ je rozklad množiny vrcholů V. Tento rozklad nám říká, k jakým vrcholům ČGHT máme přilepit které vrcholy původního grafu. Navíc musí platit, že:

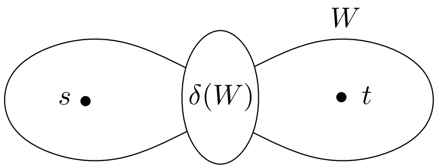
1. $\forall r : r \in C(r)$, neboli každý vrchol ČGHT je přilepen sám k sobě, a
2. $\forall st \in F : \delta\left(\bigcup_{r \in K_1} C(r)\right) = \delta\left(\bigcup_{r \in K_2} C(r)\right)$ je minimální st-řez, kde K_1 a K_2 jsou komponenty $(R, F) \setminus st$.

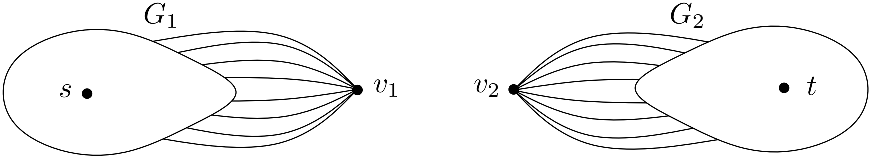
Věta (o existenci ČGHT): Buď (V, E) neorientovaný nezáporně ohodnocený graf. Pro každou podmnožinu vrcholů R existuje ČGHT.

Důkaz: Dokážeme indukcí podle velikosti množiny R.

- $|R| = 1$: ČGHT má jediný vrchol $r \in R$ a $C(r) = V$.
- $|R| > 1$: Najdeme dvojici vrcholů $s, t \in R$ takovou, že jejich minimální st-řez $\delta(W)$ je nejmenší možný. Nyní vytvoříme graf G_1 z grafu G kontrahováním všech vrcholů množiny W do jednoho vrcholu, který označíme v_1, a vytvoříme graf G_2 z G kontrahováním všech vrcholů z $\overline{W}$ do jednoho vrcholu v_2.[2]

[2] Proč to děláme „tak složitě“ a přidáváme do G_1 vrchol v_1? Na první pohled to přeci vypadá zbytečně. Problém je v tom, že i když dle HTL leží všechny minimální řezy oddělující vrcholy z W v množině vrcholů W, *hrany* těchto řezů celé v podgrafu indukovaném W ležet nemusí. K těmto řezům totiž patří i hrany, které mají ve W jenom jeden konec. Proto jsme do G_1 přidali v_1 – do něj vedou všechny zajímavé hrany, které mají ve W jeden konec. Tím *zajímavé* myslíme to, že z každého vrcholu $w \in W$ vede do v_1 *nejlevnější* hrana, která z něj vedla do množiny $V \setminus W$, případně žádná, pokud do této množiny žádná hrana nevedla.





Dále vytvoříme množiny vrcholů $R_1 = R \cap \overline{W}$ a $R_2 = R \cap W$. Dle indukčního předpokladu (R_1 i R_2 jsou menší než R) existuje ČGHT $T_1 = ((R_1, F_1), C_1)$ pro R_1 na G_1 a $T_2 = ((R_2, F_2), C_2)$ pro R_2 na G_2.
Nyní vytvoříme ČGHT pro původní graf. Označme r_1 ten vrchol R_1, pro který je $v_1 \in C_1(r_1)$, a obdobně r_2. Oba ČGHT T_1 a T_2 spojíme hranou $r_1 r_2$, takže ČGHT pro G bude $T = ((R_1 \cup R_2, F_1 \cup F_2 \cup r_1 r_2), C)$. Třídy rozkladu C zvolíme tak, že pro $r \in R_1$ bude $C(r) = C_1(r) \setminus \{v_1\}$ a pro $r \in R_2$ bude $C(r) = C_2(r) \setminus \{v_2\}$ [odebrali jsme vrcholy v_1 a v_2 z rozkladu C].
Chceme ukázat, že tento T je opravdu ČGHT. C je určitě rozklad všech vrcholů a každé $r \in C(r)$ z indukčního předpokladu, takže podmínka 1 je splněna. Co se týče podmínky 2, tak:

- pro hranu $r_1 r_2$ je $\delta(W)$ určitě minimální $r_1 r_2$-řez, protože řez mezi s a t je současně i $r_1 r_2$-řezem a je ze všech možných minimálních řezů na R nejmenší,
- pro hranu $e \neq r_1 r_2$ je $\delta(e)$ z indukce minimální řez na jednom z grafů G_1, G_2. Tento řez také přesně odpovídá řezu v grafu G, protože v G_1 i v G_2 jsme počítali s hranami vedoucími do v_1, v_2 a protože jsme ČGHT napojili přes vrcholy, k nimž byly v_1 a v_2 přilepeny.
  HTL nám navíc říká, že existuje minimální řez, který žije pouze v příslušném z grafů G_1, G_2, takže nalezený řez je minimální pro celý graf G. ♡

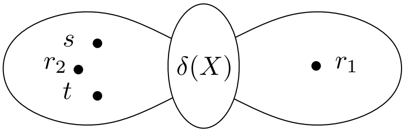
Nyní víme, že GHT existují, a také víme, jak by se daly konstruovat. Nicméně nalezení vrcholů s, t tak, aby byl minimální st-řez nejmenší možný, je časově náročné. Proto si poslední větu ještě o něco vylepšíme.

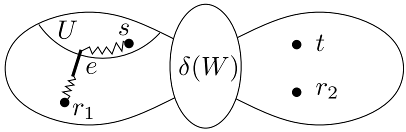
Vylepšení věty o existenci ČGHT: Na začátku důkazu není nutné hledat vrcholy s a t takové, aby byl minimální st-řez nejmenší možný. Stačí zvolit *libovolné* vrcholy $s, t \in R$ a zvolit $\delta(W)$ jako minimální st-řez.

Důkaz: Nejprve si uvědomme, proč jsme v předchozím důkazu potřebovali, aby byl $\delta(W)$ nejmenší ze všech možných st-řezů. Bylo to jenom proto, že jsme jím v ČGHT nakonec separovali vrcholy r_1 a r_2 a potřebovali jsme záruku, aby byl $\delta(W)$ opravdu minimální r_1r_2-řez. Nyní musíme ukázat, že námi nalezený st-řez $\delta(W)$ je také minimálním r_1r_2-řezem.

Pro spor tedy předpokládejme, že nějaký r_1r_2-řez $\delta(X)$ má menší kapacitu než $\delta(W)$. Navíc vezměme ten případ, kdy se to stalo „poprvé“, takže pro každé menší R je všechno v pořádku (to můžeme, protože pro $|R| = 1$ všechno v pořádku bylo).

Protože $\delta(W)$ je minimální st-řez a $\delta(X)$ má menší kapacitu, $\delta(X)$ nemůže separovat s a t. Přitom ale separuje r_1 a r_2, takže musí separovat buď s a r_1, nebo t a r_2. BÚNO nechť X separuje s a r_1.





Podívejme se nyní na ČGHT T_1 (víme, že ten je korektní) a nalezněme v něm nejlevnější hranu e na cestě spojující s a r_1. Tato hrana definuje řez $\delta(U)$, což je minimální sr_1-řez, podle HTL i v celém G. Protože $\delta(X)$ je sr_1-řez, je $d(U) \leq d(X) < d(W)$. Teď si stačí uvědomit, že $v_1 \in C(r_1)$, takže $\delta(U)$ separuje nejenom s a r_1, ale také s a v_1. Tím pádem ale separuje také s a t. To je spor, protože $d(U) < d(W)$, a přitom $\delta(W)$ měl být minimální. ♡

Teď už dokážeme GHT konstruovat efektivně – v každém kroku vybereme dva vrcholy s a t, nalezneme v čase $\mathcal{O}(\tau)$ minimální st-řez a výsledné komponenty s přidanými v_1, v_2 zpracujeme rekurzivně. Celou výstavbu tedy zvládneme v čase $\mathcal{O}(n\tau)$, čili $\mathcal{O}(n^{5/3}m)$ pro neohodnocené grafy.

Literatura

[1] R. E. Gomory and T. C. Hu. Multi-terminal network flows. *Journal of SIAM*, 9(4):551–570, 1961.