Drsná matematika III — 7. přednáška
Grafy a algoritmy: základní pojmy a úvahy

Jan Slovák

Masarykova univerzita
Fakulta informatiky

30. 10. 2006

Obsah přednášky

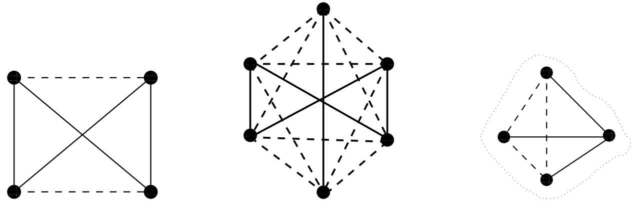
Kde je dobré číst?

- Jiří Matoušek, Jaroslav Nešetřil, Kapitoly z diskrétní matematiky, Univerzita Karlova v Praze, Karolinum, Praha, 2000, 377 s.
- Petr Hliněný, Teorie grafů, studijní materiály, http://www.fi.muni.cz/~hlineny/Vyuka/GT/
- Bill Cherowitzo, Applied Graph Therory, Lecture Notes, http://www-math.cudenver.edu/~wcherowi/courses/m4408/m4408f.html

Example

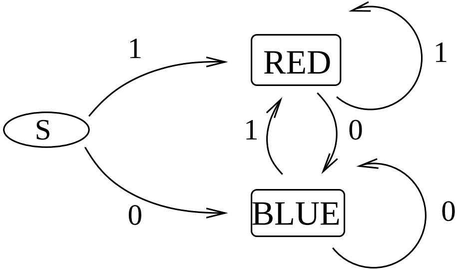
Na večírku se někteří návštěvníci po dvojicích znají a jiné dvojice se naopak neznají. Kolik lidí musíme pozvat, abychom zaručili, že se alespoň tři hosté budou buď navzájem znát nebo neznát?

Představíme situaci pomocí obrázku. Puntíky nám představí jednotlivé hosty, plnou čarou spojíme ty dvojice, které se znají, čárkovanou ty ostatní. Naše tvrzení pak zní: při jakém počtu puntíků vždy nejdeme trojúhelník, jehož strany jsou buď všechny plné nebo všechny čárkované?



Na levém obrázku takový trojúhelník není, uprostřed je. Pravý naznačuje důkaz, že existuje vždy, když počet hostů bude alespoň šest.

Představme si krabičku, která požírá jeden bit za druhým podle toho, jestli dveřmi zrovna prošel muž nebo žena – jednička nechť označuje třeba ženu. Přitom svítí buď modře nebo červeně podle toho, zda byl poslední bit nula nebo jednička (a bodle barvy světla tedy poznáme, zda je za dveřmi muž nebo žena.
Znázorníme funkci schématem:



Třetí uzel, ze kterého pouze vychází dvě šipky naznačuje start před prvním zaslaným bitem.
Takovým schématům se říká *konečný automat*.

Zachytíme společné rysy předchozích příkladů.

Definition

Grafem $G = (V, E)$ rozumíme množinu V jeho *vrcholů* spolu s podmnožinou E množiny $\binom{V}{2}$ všech dvouprvkových podmnožin ve V. Prvkům E říkáme *hrany grafu*. Vrcholům ve hraně $e = \{v, w\}$, $v \neq w$, říkáme *hraniční vrcholy* hrany e. O hranách, které mají daný vrchol v za hraniční říkáme, že z vrcholu v *vycházejí*. *Orientovaným grafem* $G = (V, E)$ rozumíme množinu V jeho vrcholů spolu s podmnožinou $E \subset V \times V$. Prvnímu z vrcholů definujících hranu $e = (v, w)$ říkáme *počáteční vrchol hrany*, druhému pak *koncový vrchol*.

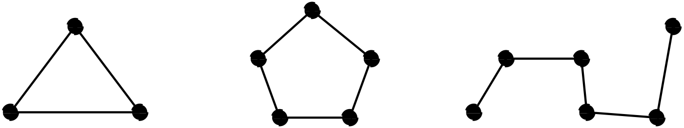
Definition (... pokračování)

Hrana e *vychází* ze svého počátečního vrcholu a *vchází* do koncového. U orientovaných hran mohou být koncový a počáteční vrchol totožný, hovoříme pak o *smyčce*.
Sousední hrany grafu jsou ty, které sdílí hraniční vrchol, u *sousedních hran orientovaného grafu* musí být vrchol pro jednu koncový a pro druhou počáteční.
Naopak, sousední vrcholy jsou ty, které jsou hraničními pro tutéž hranu.

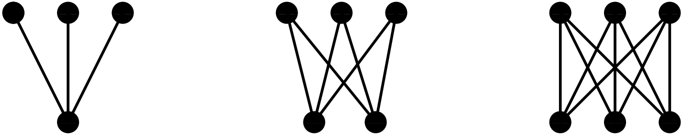
Jistě umíme grafy popisovat pomocí relací, viz. první kapitola prvního semestru. Jednoduchým věcem se dá říkat i složitě: V prvním příkladu pracujeme na stejné množině hostů se dvěmi komplementárními symetrickými a antireflexními relacemi, ve druhém pak jde o příklad dvou antisymetrických relací na třech prvcích. Nenechte se také zmást novým významem slova *graf*, pro který jsme již měli význam u funkcí. (Ve skutečnosti není podobnost věcně vzdálená.)

Grafy jsou pěkným příkladem kompromisu mezi přirozeným sklonem k „přemýšlením v obrázcích“ a přesným matematickým vyjadřováním. Obecný jazyk teorie grafů nám v konkrétních úlohách také umožňuje přidávat informace o vrcholech nebo hranách. Můžeme tak např. „obarvit“ vrcholy podle příslušnosti objektů k několika disjunktním skupinám nebo můžeme označit hrany několika různými hodnotami apod. Existence hrany mezi vrcholy různých barev může naznačit „konflikt“. Např. když modré a červené uzly představují pánskou a dámskou část večírku, pak hrana mezi vrcholy různých barev může znamenat potenciální nevhodnost sdílení pokoje pro přenocování.
Náš první příklad v předchozím odstavci můžeme tedy chápat jako graf s obarvenými hranami. Dokázané tvrzení v této řeči zní: *V grafu $K_n = (V, \binom{V}{2})$ s n vrcholy a se všemi možnými hranami obarvenými na dvě barvy je vždy alespoň jeden trojúhelník z hran o stejné barvě, pokud je počet vrcholů alespoň šest.*

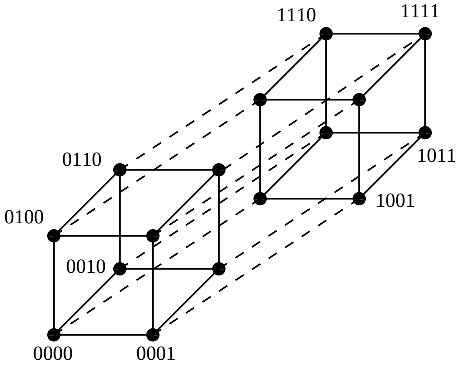
Grafu se všemi možnými hranami říkáme *úplný graf*. Značíme symbolem K_n, kde n je počet vrcholů grafu. Graf K_4 a K_5 jsme již viděli, K_3 je trojúhelník, K_2 je úsečka.
Cesta je graf, kde existuje uspořádání vrcholů $(v_0, \ldots, v_n)$ takové, že $E = \{e_1, \ldots, e_n\}$, kde $e_i = \{v_{i-1}, v_i\}$, pro všechny $i = 1, \ldots, n$. Hovoříme o *cestě délky n* a značíme ji P_n. Pokud cestu upravíme tak, že poslední a první vrchol splývají, dostaneme *kružnici délky n* a značíme C_N. Na obrázku jsou $K_3 = C_3$, C_5 a P_5



Úplný *bipartitní graf* vznikne tak, že vrcholy si obarvíme dvěmi barvami a pak přidáme všechny hrany, které spojí vrcholy různých barev. Značíme jej $K_{m,n}$, kde m a n jsou počty vrcholů s jednotlivými barvami. Na obrázku je vidět $K_{1,3}$, $K_{2,3}$ a $K_{3,3}$.

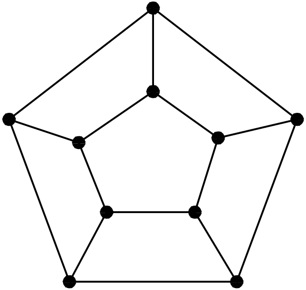


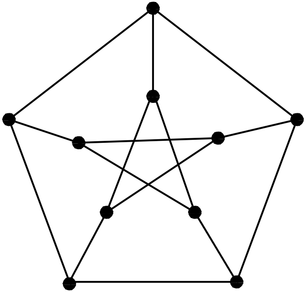
Hyperkostka H_n v dimenzi n vznikne tak, že vrcholy jsou všechna čísla $0, \ldots, 2^n - 1$. Hrany spojí právě ta čísla, která se v zápisu v dvojkové soustavě liší v právě jednom bitu. Na obrázku níže je H_4 a popis vrcholů je naznačen.



Přímo z definice vyplývá, že hyperkostku v dané dimenzi vždy dostaneme tak, že vhodně spojíme hranami dvě hyperkostky o jednu dimenzi menší. Na obrázku je naznačeno spojení dvou H_3 čárkovanými hranami. Samozřejmě ale můžeme tímto způsobem

Cyklický žebřík CL_n s $2n$ vrcholy je složen propojením dvou kopií kružnice C_n tak, že hrany spojí odpovídající vrcholy dle pořadí. Tzv. *Petersenův graf* je sice docela podobný CL_5, ale ve skutečnosti je to nejjednoduší *uv*vyvraceč nesprávných úvah – graf, na němž se vyplatí testovat tvrzení, než je začneme dokazovat.





Definition

Pro grafy grafy $G = (V, E)$ a $G' = (V', E')$ budeme za morfismus $f : G \to G'$ považovat zobrazení $f_V : V \to V'$ mezi množinami vrcholů takové, že je-li $e = \{v, w\}$ hrana v E, pak $e' = \{f(v), f(w)\}$ musí být hranou v E'. V dalším textu nebudeme ve značení odlišovat morfismus f a zobrazení f_V. Zároveň pak takové zobrazení f_V určuje i zobrazení $f_E : E \to E'$, $f(e) = e'$, kde e a e' jsou jako výše.
Pro orientované grafy je definice shodná, jen pracujeme s uspořádanými dvojicemi $e = (v, w)$ v roli hran.

Všimněme si, že tato definice znamená, že pokud $f(v) = f(w)$ pro dva různé vrcholy ve V, pak mezi nimi nesměla být hrana. U orientovaných grafů, taková hrana je přípustná, pokud je na společném obrazu smyčka.

Speciálním případem je morfismus libovolného grafu G do úplného grafu K_m. Takový morfismus je ekvivalentní vybranému obarvení vrcholů grafu V pomocí m různých jmen uzlů K_m tak, že stejně obarvené uzly nejsou spojeny hranou. Hovoříme v tomto případě o *barvení grafu* pomocí m barev.

Definition

V případě, že je morfismus $f : G \longrightarrow G'$ bijekcí na vrcholech takovou, že i f^{-1} je morfismem, hovoříme o *izomorfismu* grafů.

Izomorfní grafy se liší pouze různým pojmenováním vrcholů. Snadno umíme načrtnout až na izomorfismus všechny grafy na málo vrcholech (třeba třech nebo čtyřech). Obecně jde ale o nesmírně složitý kombinatorický problém a i rozhodnutí o konkrétních dvou daných grafech, zda jsou izomorfní je obecně mimořádně obtížné.

Jednoduchými a mimořádně užitečnými příklady morfismů grafů jsou pojmy *cesta*, *sled* a *kružnice* v grafu:

Definition

Cestou délky n v grafu G rozumíme morfismus $p : P_n \to G$ takový, že p je injektivní zobrazení (tj. všechny obrazy vrcholů $v_0, \ldots, v_n$ z P_n jsou různé). *Sled délky n* v grafu G je jakýkoliv morfismus $s : P_n \to G$ (tj. v obrazu se mohou opakovat vrcholy).

Sled si můžeme představit jako dráhu „přičinlivého ale tápajícího“ poutníka z uzlu $f(v_0)$ do uzlu f_{v_n}. Poutník se totiž v žádném uzlu nezastaví, ale klidně se po cestě grafem vrací do uzlů nebo i dokonce po hranách, kterými dříve šel. Cesta je naopak průchod grafem z počátečního uzlu $f(v_0)$ do koncového $f(v_n)$ bez takových zbytečných oklik.

Obrazy cest i sledů jsou příkladem tzv. *podgrafů*, ne však stejným způsobem:

Definition

Graf $G' = (V', E')$ je podgrafem v grafu $G = (V, E)$, jestliže $V' \subset V$, $E' \subset E$.

Speciální příklady: Uvažujme graf $G = (V, E)$ a nějakou podmnožinu $V' \subset V$. *Indukovaný podgraf* je graf $G' = (V', E')$, kde $e \in E$ patří i do E' právě, když oba krajní vrcholy hrany e patří do V'. Podgraf $G' = (V, E')$ je takový graf, který má stejnou množinu vrcholů jako G, ale jeho množina hran E' je libovolnou podmnožinou. Obecný případ je kombinací těchto dvou.

Theorem

Každý obraz homomorfismu (tj. obraz jak vrcholů tak hran) tvoří podgraf.

Podgraf, který je homomorfním obrazem cesty nazýváme také cestou. Je zřejmé, každá taková cesta o $n \geq 2$ vrcholech v grafu vzniká právě dvěma způsoby jako homomorfní obraz P_n, které se liší v počátečním a koncovém uzlu. Naopak, jestliže obraz sledu obsahuje k uzlů, můžeme obecně pro $n > k$ najít nepřeberně způsobů, jak takový obraz obdržet.
Kružnice v grafu G je injektivním homomorfním obrazem grafu C_n v G. Všimněte si, že sama kružnice C_n je také homomorfním obrazem cesty P_n, kdy první a poslední bod cesty zobrazíme do téhož vrcholu a zvolíme orientaci cesty.

Neizomorfních grafů nemůže být méně než

$$k(n) = \frac{2^{\binom{n}{2}}}{n!}.$$

Pro $n \to \infty$ lze přímo dovodit

$$\log_2 k(n) = \frac{1}{2}n^2 - O(n \log_2 n).$$

Můžeme to nepřesně formulovat tak, že velká většina všech možných grafů bude po dvou neizomorfní.

Izomorfní grafy se od sebe liší pouze přejmenováním vrcholů. Proto musí mít stejné všechny číselné charakteristiky, které se přešíslováním vrcholů nemění. Jednoduché údaje tohoto typu můžeme dostat sledováním počtů hran vycházejících z jednotlivých vrcholů.

Definition

Pro vrchol $v \in V$ v grafu $G = (V, E)$ říkáme, že jeho *stupeň* je k, jestliže v E existuje k hran, jejichž hraničním vrcholem v je. Píšeme v takovém případě

$$\deg v = k.$$

Skóre grafu G s vrcholy $V = (v_1, \dots, v_n)$ je posloupnost

$$(\deg v_1, \deg v_2, \dots, \deg v_n)$$

Pro izomorfní grafy se jejich skóre může lišit pouze permutací hodnot. Pokud tedy porovnáme skóre grafů setříděné podle velikosti hodnot, pak různá skóre zaručují neizomorfnost grafu. Naopak ale snadno najdeme příklad grafů se stejným skóre, které izomorfní být nemohou, např. $G = C_3 \cup C_3$ má skóre $(2, 2, 2, 2, 2, 2)$, stejně jako C_6. Zjevně ale izomorfní nejsou, protože v C_6 existuje cesta délky 5, která v druhém grafu být nemůže.
Jaká skóre mohou grafy mít?
Protože každá hrana vychází ze dvou vrcholů, musí být v celkovém součtu skóre započtena každá hrana dvakrát. Proto platí

$$\sum_{v \in V} \deg v = 2|E|.$$

Zejména tedy musí být součet všech hodnot skóre sudý.

Následující věta je vlastně o návod, jak pro dané skóre buď zjistit, že graf s takovým neexistuje nebo takový graf sestrojit.

Theorem (Algoritmus na sestrojení grafu s daným skóre)

Pro libovolná přirozená čísla $0 \leq d_1 \leq \cdots \leq d_n$ *existuje graf* G *na* n *vrcholech s těmito hodnotami skóre tehdy a jen tehdy, když existuje graf se skóre*

$$(d_1, d_2, \ldots, d_{n-d_n} - 1, d_{n-d_n+1} - 1, \ldots, d_{n-1} - 1)$$

na $n-1$ *vrcholech.*

U orientovaných grafů rozlišujeme *vstupní stupeň* $\deg_+ v$ vrcholu v a *výstupní stupeň* $\deg_- v$.

Grafy jsou jazykem, ve kterém často formulujeme algoritmy. Rozumíme tím postup, kdy v nějakém orientovaném grafu přecházíme z uzlu do uzlu podél hran a přitom zpracováváme informace, které jsou určeny a ovlivněny výsledkem předchozích operací, uzlem, ve kterém se zrovna nacházíme, a hranou, kterou jsme do uzlu vstoupili. Při zpracování informace se zároveň rozhodujeme, kterými výstupními hranami budeme pokračovat a v jakém pořadí. Pokud je graf neorientovaný, můžeme všechny hrany považovat za dvojice hran orientované opačnými směry.

Abychom mohli algoritmy realizovat pomocí počítače, je třeba umět uvažovaný graf efektivně zadat. Uvedeme dva příklady:

Definition

Hranový seznam (Edge List). Graf $G = (V, E)$ si v něm reprezentujeme jako dva seznamy V a E propojené ukazately tak, že každý vrchol ukazuje na všechny z něj vycházející hrany (případně také na všechny do něj vcházející hrany u orientovaných grafů) a každá hrana ukazuje na svůj počáteční a koncový vrchol.

Je vidět, že pamět potřebná na uchování grafu je v tomto případě $O(|V| + |E|)$, protože na každou hranu ukazujeme právě dvakrát a na každý vrchol ukazujeme tolikrát, kolik je jeho stupeň a součet stupňů je také roven dvojnásobku počtu hran. Až na konstantní násobek jde tedy stále o optimální způsob uchovávání grafu v paměti.

Definition (Matice sousednosti grafu)

Uvažme (neorientovaný) graf $G = (V, E)$, zvolme uspořádání jeho vrcholů $V = (v_1, \dots, v_n)$ a definujme matici $A_G = (a_{ij})$ nad $\mathbb{Z}_2$ (tj. zaplněnou jen nulami a jedničkami) takto:

$$a_{ij} = \begin{cases} 1 & \text{jestliže je hrana } e_{ij} = \{v_i, v_j\} \text{ v } E \\ 0 & \text{jestliže není hrana } e_{ij} = \{v_i, v_j\} \text{ v } E \end{cases}$$

Matici A_G nazýváme matice souslednosti grafu G.

Jak vypadají matice grafů z příkladů na začátku této přednášky?

Při nejjednodušším způsobu uchovávání matic v poli je zadání grafu pomocí matice souslednosti velice neefektivní metoda. Potřebuje totiž vždy $O(n^2)$ místa v paměti. Pokud je ale v grafu málo hran, dostáváme tzv. řídkou matici se skoro všemi prvky nulovými. Existují ovšem postupy, jak tyto řídké matice uchovávat v paměti efektivněji.

Definition (Základní operace nad grafem)

- *odebrání hrany*
- *přidání hrany*
- *přidání vrcholu*
- *odebrání vrcholu*
- *dělení hrany nově přidaným vrcholem*

Jak se projeví tyto operace v našich reprezentacích?

Jednoduchou aplikací maticového počtu je tvrzení:

Theorem

Nechť $G = (V, E)$ je graf s uspořádanými vrcholy $V = (v_1, \dots, v_n)$ a maticí souslednosti A_G. Označme $A_G^k = (a_{ij}^{(k)})$ prvky k-té mocniny matice A_G. Pak $a_{ij}^{(k)}$ je počet sledů délky k mezi vrcholy v_i a v_j.

Corollary

Jsou-li $G = (V, E)$ a A_G jako v předchozí větě, pak lze všechny vrcholy G spojit cestou právě, když má matice $(A + \mathbb{I}_n)^{n-1}$ samé nenulové členy (zde $\mathbb{I}_n$ označuje jednotkovou matici s n řádky a sloupci).

Jaký je vliv má permutace našeho uspořádání uzlů V na matici souslednosti grafu?

Permutace uzlů grafu G má za následek jednu a tutéž permutaci řádků a i sloupců matice A_G. Každou takovou permutaci můžeme zadat právě jednou tzv. permutační maticí, tj. maticí z nul a jedniček, která má v každém řádku a každém sloupci právě jednu jedničku a jinak nuly. Je-li P taková parmutační matice, pak nová matice souslednosti izomorfního grafu G' bude

$$A_{G'} = P \cdot A_G \cdot P^T,$$

kde P^T značí transponovanou matici a tečkou označujeme násobení matic.

Každou permutaci umíme napsat jako složení transpozic a proto příslušnou permutační matici dostaneme jako součin příslušných matic pro transpozice.

V případě permutačních matic je matice trnasponovaná zároveň maticí inverzní. Tyto úvahy lze dále rozvíjet a přemýšlet o souvislostech matic souslednosti a matic lineárních zobrazení mezi vektorovými prostory. Nebudeme zde zacházet do podrobností.