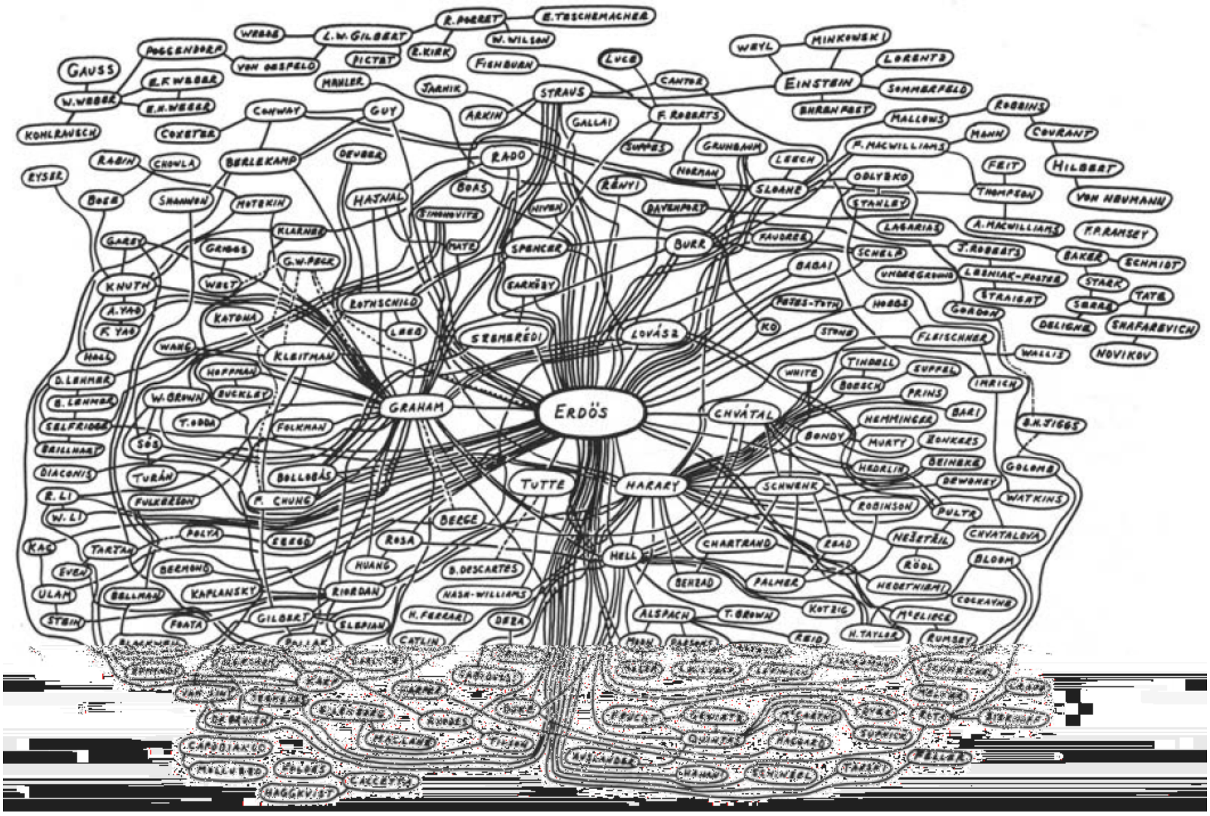
Středoškolská odborná činnost



Sociální sítě z pohledu teorie grafů

Matěj Žídek

Malenovice 2015

Středoškolská odborná činnost

Obor SOČ: 01. Matematika a statistika

Sociální sítě z pohledu teorie grafů

Autor: **Matěj Žídek**

Škola: Gymnázium, Frýdlant nad Ostravicí
nám. T. G. Masaryka 1260, příspěvková organizace
739 11 Frýdlant nad Ostravicí

Konzultant: **doc. Mgr. Petr Kovář, Ph.D.**
VŠB – Technická univerzita Ostrava

Malenovice 2015

Prohlášení
Prohlašuji, že jsem svou práci vypracoval samostatně pod vedením doc. Mgr. Petra Kováře, Ph.D. Použil jsem pouze podklady (literaturu, SW atd.) uvedené v přiloženém seznamu a postup při zpracování a dalším nakládání s prací je v souladu se zákonem č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) v platném znění.

V Malenovicích dne 31. března 2015

Poděkování
V první řadě děkuji panu doc. Mgr. Petru Kovářovi, Ph.D., za výborné vedení a pomoc při této práci. Dále taky mému bratrovi Augustinovi, nebýt něho, tak o SOČ nevím a tím bych přišel o tuto nádhernou matematiku. A děkuji i ostatním členům mé rodiny a Silvii, za psychickou podporu a trpělivost.

Anotace

Struktura sociálních sítí z pohledu teorie grafů dokáže odhalit mnoho. V této práci hledáme pomocí metrik ideální definici skupiny, podle které by se skupina dala nalézt pouze na základě dat z grafu sociální sítě, kde její členové jsou zastupovány vrcholy a souvislost mezi nimi odpovídá hranám. Protože propojení v sociálních sítích je symetrická vlastnost, zabýváme se pouze neorientovanými grafy. Takové a podobné metody používají dnešní sociální sítě, které mají množství reálných dat[1].

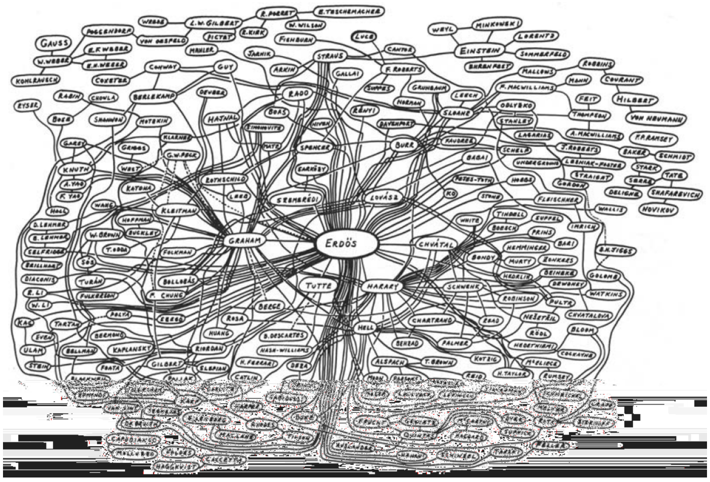
Klíčová slova: metrika; hranově disjunktní cesty; sociální sítě; teorie grafů; relace; Mengerova věta

Obsah

1. Úvod

V dnešní době výpočetní síla počítačů a sociální sítě typu Facebook, Google+, LinkedIn atd. umožňují aplikovat poměrně staré metody z teorie grafů nebývalými způsoby za účelem zkoumání statistických dat o lidech, jejich chování, interakcích a skupinách, které tvoří. Tato data se dají sehnat pomocí automatizovaného procházení těchto sítí. V této práci budeme na sociální sítě nahlížet jako na neorientovaný graf. Vrcholy v grafu odpovídají jejím členům a hrany odpovídají relacím mezi jednotlivými členy (např. přátelství, spoluautorství). Snažíme se získaná data analyzovat a této analýzy určit skupiny členů sociální sítě (výzkumná skupina, třída apod.).

Příkladem takové sociální sítě je i obrázek na titulní straně (viz obr. 1.1). Jedná se o graf, kde vrcholy tvoří autoři matematických článků, hrany mezi nimi značí spoluautorství na aspoň jednom článku. Uprostřed je maďarský matematik Paul Erdős, který napsal přes 1500 matematických článků, mnohé z nich byly z oblasti teorie grafů.



Obrázek 1.1: Erdősův graf

K určení skupin je nejdřív třeba nadefinovat pojem *skupina*. V této práci je ukázáno několik možností definic. Zkoumáme, jak je která definice skupiny praktická, jaké má výhody a nevýhody. Hlavně se věnujeme definici pomocí metriky #. Metriky jsou obecně výhodné pro definování skupin, protože když o členech skupiny nevíme nic osobně, pak potřebujeme něco, čímž jsme schopni uchopit intuitivně chápaný pojem „skupina“. A to právě metriky dělají: každým dvěma vrcholům přiřadí jejich vzdálenost – určité číslo a na základě těchto čísel jsme schopni vyvodit informace o skupinách. Zjednodušeně řečeno, vrcholy, které jsou blízko sobě pod danou metrikou, tvoří skupinu.

Metrika je přesně definovaný pojem: než můžeme nějakou funkci prohlásit za metriku, musí splňovat určité podmínky uvedené podrobně v kapitole 4, které je třeba ověřit. K tomu na začátku práce dokážeme větu 1. Týká se hranově disjunktních cest mezi různými dvojicemi ze třech vrcholů.

2. Přehled použitých pojmů z teorie grafů

- **Cesta**

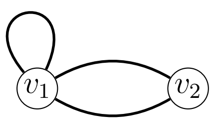
  Cesta z vrcholu a do vrcholu b je posloupnost hran a vrcholů, která tyto dva vrcholy spojuje. Žádný vrchol ani hrana se při tom nesmí opakovat – cesta nesmí obsahovat cykly.

- **Jednoduchý, souvislý, neorientovaný graf**

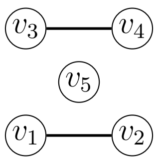
  Jednoduchý graf je takový, který neobsahuje smyčky (hrany, které začínají a končí ve stejném vrcholu), ani násobné hrany (více hran mezi dvěma vrcholy) a ani není orientovaný. Graf, který smyčky a násobné hrany obsahuje se nazývá pseudograf (viz obr. 2.1a), ale těmi se v této nezabýváme.

  Souvislý graf je takový, ve kterém existuje mezi každými dvěma vrcholy cesta. Příklad nesouvislého grafu je na obr. 2.1b, těmi se ale taky příliš nezabýváme.

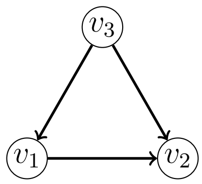
  Neorientovaný graf je takový, který nemá hrany orientované, tedy u hran není důležité pořadí jejích koncových vrcholů. Zde se zabýváme pouze nimi, jelikož vztahy v sociální síti jsou vzájemné. Na obr. 2.1c je příklad orientovaného grafu.



(a) Pseudograf



(b) Nesouvislý graf



(c) Orientovaný graf

- **Hranově disjunktní cesty**

  Dvě (u, v)-cesty jsou hranově disjunktní, jestliže nemají žádnou společnou hranu. [3] (str. 95)

- **Metrika** Metrika je funkce, která každé dvojici vrcholů přiřadí nezáporné číslo, která splňuje trojúhelníkovou nerovnost a navíc je symetrická. Podrobněji se jim včetně definice věnujeme v kapitole 4.

- **Automorfismus**

Grafy G a H se nazývají isomorfní, jestliže existuje bijekce $\Phi : V(G) \to V(H)$ taková, že každé dva vrcholy $u; v \in V(G)$ jsou sousední právě tehdy, když jsou sousední vrcholy $\Phi(u); \Phi(v) \in V(H)$. Zobrazení Φ se nazývá isomorfismus. Isomorfismus $f : V(G) \to V(G)$ se nazývá automorfismus grafu G. [3] (str. 82 a 86)

Jinými slovy, pokud lze v grafu přejmenovat libovolné dva vrcholy tak, že zůstane stejná matice sousednosti (matice, ve které je zaznamenané s jakými vrcholy sousedí každý vrchol). Takovými grafy jsou např. všechny kompletní grafy (viz obr. 5.5, na kterém je kompletní graf na pěti vrcholech).

3. Věta o hranově disjunktních cestách

Nejdřív dokážeme pro tuto práci velice důležitou větu, se kterou budeme v dalších kapitolách pracovat.

Hranově disjunktní cesty mezi vrcholy x a y jsou takové cesty, které nemají žádnou společnou hranu. Jejich maximální počet označme $p(x, y)$.

Věta 1. *Mějme jednoduchý souvislý graf $G = (V, E)$ a tři vrcholy $x, y, z \in V$. Pak největší počet hranově disjunktních cest mezi dvěma dvojicemi z těchto vrcholů je stejný a počet hranově disjunktních cest mezi třetí dvojicí z nich je větší nebo roven ostatním. Tedy BÚNO (bez újmy na obecnosti) platí $p(x, y) \geq p(x, z) = p(y, z)$.*

Důkaz. Označme k minimum z $p(x, y)$ a $p(y, z)$. Platí tedy:

$$p(x, y) \geq k$$
$$p(y, z) \geq k$$

Pokud odebereme méně než k libovolných hran z grafu G, tak i pak bude určitě aspoň jedna (x, y) i (y, z)-cesta. To vyplývá z Mengerovy věty níže, jejíž důkaz nalezneme zde: [2].

Věta 2. ***Mengerova věta***
Graf G je hranově k-souvislý právě tehdy, když mezi libovolnými dvěma vrcholy grafu G existuje alespoň k hranově disjunktních cest. (Převzato z [3] (str. 96).)

Tudíž určitě existuje i aspoň jedna (x, z)-cesta, která vznikne spojením (x, y) a (y, z)-cest. Pokud by přímo tímto spojením nevznikla cesta, tedy aspoň jeden vrchol by byl použit dvakrát, vezmeme úsek $x - u$, kde u je první (nejblíže vrcholu x) společný vrchol (x, y)-cesty s (y, z)-cestou. Tento úsek spojíme s úsekem $u - z$ z původní (y, z)-cesty a vznikne (x, z)-cesta.

Když tedy platí, že po odebrání méně než k libovolných hran z grafu G, existuje aspoň jedna (x, z)-cesta, pak z Mengerovy věty plyne, že v původním grafu G existuje aspoň k hranově disjunktních (x, z)-cest. Platí tedy i:

$$p(x, z) \geq k \tag{3.1}$$

Analogicky když označíme l minimum z $p(x, y)$ a $p(x, z)$, pak platí:

$$\begin{aligned} p(x, y) &\geq l \\ p(x, z) &\geq l \\ p(y, z) &\geq l \end{aligned} \tag{3.2}$$

Jedna z hodnot $p(x, y)$, $p(y, z)$ a $p(x, z)$ je určitě větší nebo rovna zbylým, BÚNO můžeme předpokládat, že je to $p(x, y)$, jelikož pokud by to tak neplatilo, jen si přejmenujeme vrcholy tak, aby to platilo. Z toho vyplývá

$$k = p(y, z) \tag{3.3}$$

$$l = p(x, z) \tag{3.4}$$

jelikož jsou to minima ze dvou hodnot, kde druhá - $p(x, y)$ - je jim větší nebo rovna. Když dosadíme (3.3) do (3.1) a (3.4) do (3.2), vyjde

$$\begin{aligned} p(x, z) &\geq p(y, z) \\ p(y, z) &\geq p(x, z) \end{aligned}$$

Z toho plyne (spolu s podmínkou, že $p(x, y)$ je největší):

$$p(x, y) \geq p(y, z) = p(x, z)$$

□

3.1 Jiná cesta k důkazu

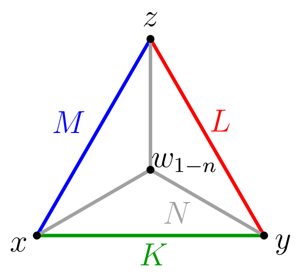
Větu 1 jsem se snažil dokázat dlouho jiným způsobem. Nemyslím si, že by byl špatný, jen jsem vždy narazil na to, že je třeba ošetřit příliš mnoho možností. Ale velice se mi takový důkaz líbil, i když nebyl kompletní. Doufám, že se mi to v blízké budoucnosti podaří. Zde předkládám jeho náčrt.

Začneme tím, že najdeme mezi každou dvojicí z vrcholů x, y a z takové hranově disjunktních cesty [1], že jejich počet je největší možný. To je určitě možné, byť to v hustějších grafech může být složité. Pak ukážeme, že tento původní výběr cest lze přeuspořádat tak, že tyto cesty budeme moci všechny zařadit do čtyřech skupin popsaných níže tak, aniž by se jejich počet změnil, tedy počet zůstane stále největší možný.[2]. Potom už jen spočítáme pomocí těch čtyř skupin, kolik je $p(x, y)$, $p(x, z)$ a $p(y, z)$ a z toho vyplyne, že dvě z těchto hodnot se rovnají a ta třetí je buď stejná nebo větší.

Skupiny, do kterých chceme všechny cesty zařadit jsou následovné:

[1] Hranově disjunktních pouze v rámci každé dvojice zvlášť, tedy (x, y)-cesta může mít společné hrany s (x, z)-cestou apod.

[2] Na tomto jsem se vždy zastavil, i když jsem to zkoušel mnohými způsoby



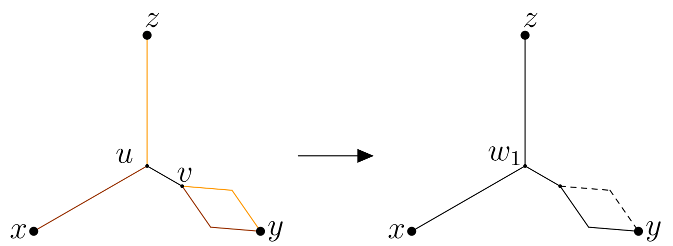
Obrázek 3.1: Rozdělení cest do skupin

Všechny čtyři skupiny jsou hranově disjunktní (vrcholy ale společné mít můžou). Skupina K (skupiny M a L obdobně) obsahuje hranově disjunktní (x, y)-cesty, které neprochází bodem z a zároveň nemají žádnou společnou hranu s (x, z)-, resp. (y, z)-cestami, které neprochází bodem y, resp. x.

Ve skupině N jsou určitě trojice cest. Každá trojice je složená z vzájemně hranově disjunktních cest mezi $x - w_a, y - w_a$ a $z - w_a$. Vrchol w je s indexem a, protože tento vrchol může být pro každou trojici jiný. Do počtů $p(x, y)$, $p(x, z)$ a $p(y, z)$ se každá tato trojice započítá právě jednou, protože z každé trojice sedící v jedné lavici vznikne jedna (x, y)-cesta (analogicky i (x, z)- a (y, z)-cesty) spojením úseků $x - w_a$ a $w_a - y$.

Do těchto skupin nelze (alespoň ne vždy) zařadit hranově disjunktní (x, y)-, (x, z)- a $p(y, z)$-cesty z původního výběru (který byl udělaný tak, že jejich počty jsou největší možné). Proto je třeba původní výběr cest upravit tak, aby šly zařadit do našich skupin aniž by se změnil jejich počet. Ukázat, jak tyto cesty přeuspořádat, se mi ale nepodařilo. Nicméně ukážeme si aspoň jeden lehký případ na obr. 3.2 pro lepší představu:

Máme hnědou (x, y)-cestu a žlutou (y, z)- cestu, které mají společný černý úsek (u, v). Úsek (v, y) má každá cesta jiný, tak si vybereme libovolný z nich (např. ten žlutý). Ten odhodíme a nahradíme ho tímto úsekem od té druhé cesty (tedy ten hnědý). Vrchol u je vrcholem w_1. Všechny z počtů $p(x, y)$, $p(x, z)$ a $p(y, z)$ zůstaly zachované (v tomto případě jsou všechny stále jedna) a podařilo se zařadit tyto cesty do skupiny N (tvoří jednu trojici $x - w_1, y - w_1$ a $z - w_1$).



Obrázek 3.2: Příklad úpravy do skupin z původního výběru cest

Když máme tedy cesty přeuspořádané, tak je už lehké spočítat, kolik jich je. Vysvětlíme, jak se počítá $p(x, y)$, zbylé dva počty se počítají analogicky.

BÚNO můžeme předpokládat, že platí $K \geq L \geq M$. Cesty započítané do $p(x, y)$ můžou být následovných typů:

1. Každá (x, y)-cesta tohoto typu je tvořena jedním členem ze skupiny K a jsou tak použiti všichni členové skupiny K. Těchto cest je tolik, co v je členů v K. Počet členů označíme $|K|$.
2. Každá (x, y)-cesta tohoto typu je tvořena jedním členem ze skupiny N. Analogicky je těchto cest $|N|$.
3. Každá (x, y)-cesta tohoto typu je tvořena jedním členem ze skupiny M a jedním členem ze skupiny$L$. Zde se ale nepoužijí všichni členové skupiny L, jelikož se spojuje vždy jeden a jeden člen z každé skupiny a skupina L má členů více. Tedy tak vznikne minimum z $|L|$ a $|M|$, tedy $|M|$.

Platí tedy:

$$
\begin{aligned}
p(x, y) &= |K| + |N| + |M| \\
p(x, z) &= |M| + |N| + |L| \\
p(y, z) &= |L| + |N| + |N|
\end{aligned}
$$

Z toho je jasně vidět, že $p(x, y) \geq p(x, z) = p(y, z)$.

4. Metriky

Než budeme schopni skupinu v sociální síti nějak nadefinovat, potřebujeme způsob, jakým získávat nějaká konkrétní data o vrcholech. K tomu jsou dobré metriky, protože jejich výstupem je číslo.

Definice metriky[3]: Metrika ρ na množině A je takové zobrazení $\rho : A \times A \to \mathbb{R}$, že $\forall x, y, z \in A$ platí:

1. $\rho(x, y) \geq 0$, přičemž $\rho(x, y) = 0$ jen pro $x = y$,
2. $\rho(x, y) = \rho(y, x)$,
3. $\rho(x, y) + \rho(y, z) \geq \rho(x, z)$.

4.1 Metrika

4.1.1 Definice a důkaz metriky

Mějme zobrazení # na množině vrcholů jednoduchého souvislého grafu $G = (V, E)$, které je definováno následovně:

$$\forall x, y \in V \; : \; \#(x, y) = \begin{cases} \frac{1}{p(x,y)} & x \neq y, \\ 0 & x = y, \end{cases}$$

kde $p(x, y)$ je počet hranově disjunktních cest mezi x a y. Ukážeme, že dané zobrazení je metrikou.

1. $\#(x, y) = 0$ jen pro $x = y$, což je splněno z definice zobrazení # a $\#(x, y)$ pro $x \neq y$ musí být kladné, protože $p(x, y)$ musí být na souvislém grafu aspoň 1. Pro nesouvislý graf není metrika # definována (můžeme ji definovat pro jeho souvislé komponenty).
2. Metriku hledáme na neorientovaném grafu, kde pro každou cestu z x do y najdeme i cestu z y do x a platí tedy $p(x, y) = p(y, x)$.
3. Třetí vlastnost dokážeme pomocí následujícího lemmatu:
   Lemma 1. *V souvislém neorientovaném grafu $G = (V, E)$ platí $\forall x, y, z \in V$, že dvě z hodnot $\#(x, y)$, $\#(x, z)$ a $\#(y, z)$ jsou shodné a třetí je menší nebo jim rovna.*

Důkaz. Využijeme věty 1. Hodnota zobrazení $\#(x, y)$ je obrácená hodnota počtu $p(x, y)$, takže větu 1 pouze obrátíme hodnotu (spolu s otočením nerovnosti). BÚNO pro $p(x, y) \geq p(x, z) = p(y, z)$ platí $\#(x, y) \leq \#(x, z) = \#(y, z)$. □

Z toho plyne, že pro zobrazení # vždy platí trojúhelníková nerovnost. To proto, že buď můžou nastat dvě možnosti ($A = \#(x, y);\ B = \#(x, z) = \#(y, z)$): $A + A \geq B$ - to platí, jelikož je $A \leq B$; nebo $A + B \geq A$, to platí triviálně, když A i B jsou kladné.

Zobrazení # tedy je metrikou.

4.1.2 Relace ekvivalence $R_\#$

Ještě než se pustíme na hledání skupin v sociálních sítích, zjistíme si o této metrice zajímavou vlastnost, ze které vyplynou později vlastnosti skupin nadefinovaných pomocí této metriky.

Definice relace ekvivalence: Je-li relace R na množině A:

- reflexivní $\forall x \in A \ :\ (x, x) \in R$
- symetrická $\forall x, y \in A \ :\ (x, y) \in R \ \Leftrightarrow\ (y, x) \in R$
- tranzitivní $\forall x, y, z \in A \ :\ (x, y) \in R \ \wedge\ (y, z) \in R \ \Leftrightarrow\ (x, z) \in R$

pak je R relací ekvivalence a má smysl definovat rozklad množiny A.

Lemma 2. *Relace $R_\#$ na množině vrcholů V definovaná*

$$\forall x, y \in V \ :\ (x, y) \in R_\# \ \Leftrightarrow\ \#(x, y) \leq k; \ \textit{pro nějaké } k \geq 0$$

je relací ekvivalence.

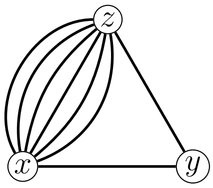
Důkaz. Reflexivita a symetričnost relace $R_\#$ platí z definice. Tranzitivita platí, protože pokud platí $(x, y) \in R_\# \ \wedge\ (y, z) \in R_\#$, tak by podle definice této relace mělo zároveň platit $\#(x, y) \leq k \wedge \#(y, z) \leq k$. A protože dva z $\#(x, y),\ \#(x, z),\ \#(y, z)$ musí být stejné, takže víme, že aspoň jedna z nich je menší než k a třetí menší (plyne z lemmatu 1), pak musí platit $\#(x, z) \leq k$. Z toho plyne, že $\#(x, z) \in R_\#$. □

4.2 Co není metrikou

Při hledání různých metrik jsme prošli několik slepých uliček. Vyjmenujeme si některé a popíšeme důvody, proč tyto návrhy funkcí $V(G) \times V(G) \to \mathbb{R}_0^+$ nemůžou být metrikou.

- $\mathbf{p(x, y)}$

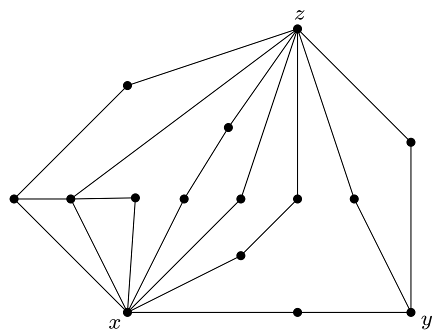
  Jak již bylo zmíněno dříve, $p(x, y)$ je počet hranově disjunktních cest mezi x a y. Toto nemůže být metrikou, protože lze libovolně zvětšovat počet cest mezi x a z, aniž by to nějak ovlivnilo počet cest mezi x a y, takže obecně nebude platit trojúhelníková nerovnost (viz obr. 4.1, na kterém $p(x, y) = 2$, $p(x, z) = 7$ a $p(y, z) = 2$. Neplatí $7 \not\leq 2 + 2$). U metriky # toto nevadí, protože zvětšením $p(x, z)$ se $\#(x, z)$ zmenší.



Obrázek 4.1: Mezi x a z vede hodně cest a ovlivňuje to pouze $p(x, z)$

- **Počet sledů $s_k(x, y)$ mezi vrcholy x a y délky nejvýše k**

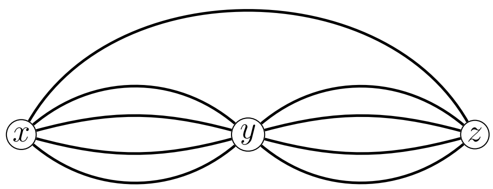
  Pokud by byly vrcholy x, y, z od sebe vzdálené právě k, tak můžeme při dodržení této vlastnosti přidat sledy mezi libovolné dva z nich podle libosti, takže by neplatila trojúhelníková vzdálenost (viz obr. 4.2, $s_k(x, y)$ je počet sledů mezi x a y délky nejvýše k).



Obrázek 4.2: $s_3(x, y) = 1$; $s_3(x, z) = 7$; $s_3(y, z) = 2$, trojúhelníková nerovnost neplatí: $7 \not\leq 2 + 3$

- $\#_{\mathbf{k}}(\mathbf{x}, \mathbf{y}) = \frac{1}{p_k(x,y)}$, kde p_k je počet hranově disjunktních cest délky nejvýše k.

  Opět lze nalézt protipříklad, viz obrázek 4.3, na něm je $\#_2(x, y) = \frac{1}{4}$, $\#_2(y, z) = \frac{1}{4}$ a $\#_2(x, z) = 1$. Neplatí trojúhelníková nerovnost: $1 \not\leq \frac{1}{4} + \frac{1}{4}$.



Obrázek 4.3: Hranově disjunktní cesty délky nejvýše k

Ve všech těchto případech bylo problémem, že u každého šlo jeden parametr ovlivňovat mezi vrcholy x a y tak, že to neovlivňovalo stejný parametr u x a z. Proto tyto případy nemůžou být metrikou.

5. Zavedení skupin

V této kapitole ukážeme několik možností, jak nadefinovat skupinu v sociální síti. Hlavně rozebereme detailně možnost pomocí metriky #. Popíšeme vlastnosti daných skupin, zhodnotíme jejich praktičnost.

Díky těmto zavedením může být počítač schopný nalézt skupiny i na větších grafech, kde člověk už není dostatečně rychlý. Bohužel nemám k dispozici reálná data z nějaké sociální sítě a ani neumím na to dostatečně programovat, abych takový program na vyhledávání skupin naprogramoval.

5.1 Na základě metriky

Definice skupiny: X je skupina v souvislém neorientovaném grafu $G = (V, E)$, pokud $\forall x, y \in X$ platí:

1. $\#(x, y) \leq k$,
2. další vrchol už není možné do X přidat, aniž by se neporušila vlastnost 1.

Na přesné určení vhodné konstanty k by bylo třeba reálná data, aby se zjistilo, jak to nejvíce odpovídá realitě, navíc může být různé v jiných sociálních sítích. Nám však stačí pracovat obecně s k.

5.1.1 Vlastnosti skupin v konkrétních grafech

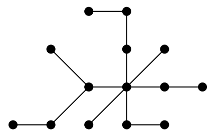
V závislosti na parametru k prozkoumáme počet skupin v některých konkrétních grafech a vyvodíme z toho obecné závěry. Definice pojmů jsou převzaty z [3].

- **Cesta/strom $\mathbf{T_n}$**

  Cesta je sled, ve kterém se neopakují vrcholy (viz obr. 5.1).

Obrázek 5.1: Příklad cesty

Strom je souvislý acyklický graf (tj. graf, který neobsahuje cyklus), viz obr. 5.2.



Obrázek 5.2: Příklad stromu

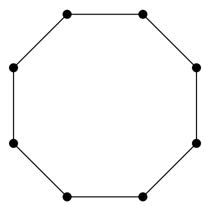
Pro cestu i strom platí, že mezi každými dvěma různými vrcholy je právě jedna cesta. Z toho plyne:

$$\forall x, y \in V(T) \ : \ \#(x,y) = \begin{cases} 1 & x \neq y \\ 0 & x = y \end{cases}$$

k	Počet skupin	Poznámka
$k \geq 1$	1	Všechny vrcholy jsou v jedné skupině.
$k < 1$	n	Každý vrchol je sám ve svojí skupině.

- **Cyklus $\mathbf{C_n}$**

  Cyklus je graf s vrcholovou množinou $V = \{x_1, x_2, \ldots, x_n\}$, pro $n \geq 3$, a množinou hran $E = \{x_1x_2, x_2x_3, \ldots, x_{n-1}x_n, x_nx_1\}$ (viz obr. 5.3)



Obrázek 5.3: Cyklus C_8

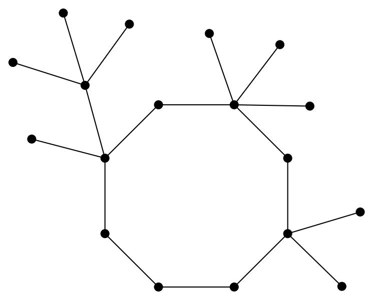
Pro cyklus platí, že mezi každými dvěma různými vrcholy v grafu jsou právě dvě cesty. Takže:

$$\forall x, y \in V(T) \ : \ \#(x,y) = \begin{cases} \frac{1}{2} & x \neq y \\ 0 & x = y \end{cases}$$

k	Počet skupin	Poznámka
$k \geq \frac{1}{2}$	1	Všechny vrcholy jsou v jedné skupině.
$k < \frac{1}{2}$	n	Každý vrchol je sám ve svojí skupině.

- **Unicyklický graf**

  Unicyklický graf je graf, který obsahuje právě jeden cyklus (viz obr. 5.4).



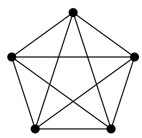
Obrázek 5.4: Unicyklický graf

$$\forall x, y \in V(G) \; : \; \#(x,y) = \begin{cases} 1 & x \neq y;\ \text{aspoň jeden z } x, y \text{ neleží na cyklu} \\ \frac{1}{2} & x \neq y;\ x, y \text{ leží na cyklu} \\ 0 & x = y \end{cases}$$

k	**Počet skupin**	**Poznámka**
$k \geq 1$	1	Vždy platí, že $\#(x,y) \leq 1$, protože $p(x,y)$ je vždy přirozené. Z toho plyne, že všechny vrcholy budou v jediné skupině.
$1 > k \geq \frac{1}{2}$	$1 + m$	Proměnná m je počet vrcholů neležících na cyklu. $\forall x, y$ ležících na cyklu je $p(x,y) = 2$, tedy $\#(x,y) = \frac{1}{2}$, takže jsou v jedné skupině. Mezi zbylými dvojicemi vrcholů je vždy právě jedna hranově disjunktní cesta, tedy jejich $\#(x,y) = 1 \geq k$, takže každý vrchol je sám ve svojí skupině.
$\frac{1}{2} > k$	n	Proměnná n je počet vrcholů. Mezi žádnou dvojicí není $\#(x,y) \leq \frac{1}{2}$, tudíž je každý vrchol sám ve své skupině.

- **Kompletní graf $\mathbf{K_n}$**

  Kompletní graf K_n je graf na n vrcholech obsahující všech $\binom{n}{2}$ hran (viz obr. 5.5).



Obrázek 5.5: Kompletní graf K_5

Libovolné dva různé vrcholy v kompletním grafu mají mezi sebou tolik hranově disjunktních cest, jako je stupeň každého z nich, tedy $n - 1$.

$$\forall x, y \in V(K_n) \; : \; \#(x,y) = \begin{cases} \frac{1}{n-1} & x \neq y \\ 0 & x = y \end{cases}$$

k	Počet skupin	Poznámka
$k \geq \frac{1}{n-1}$	1	Všechny vrcholy jsou v jedné skupině.
$k < \frac{1}{2}$	n	Každý vrchol sám ve svojí skupině.

- **Kompletní graf bez jedné hrany**

  V kompletním grafu bez jedné hrany mezi vrcholy, které:

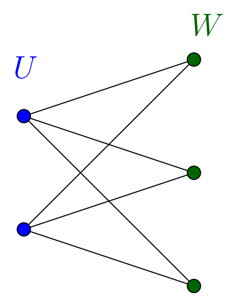
  - mají oba stupeň $n-1$ (tj. vedou od nich hrany do všech ostatních vrcholů), je hranově disjunktních cest $n-1$.
  - mají aspoň jeden z nich stupeň $n-2$ (tj. chybějící hrana do kompletnosti grafu vedla z tohoto vrcholu), je hranově disjunktních cest $n-2$.

$$\forall x, y \in V(G) \;:\; \#(x,y) = \begin{cases} \frac{1}{n-2} & x \neq y;\text{ aspoň jeden z}x,\ y \text{ je stupně } n-2 \\ \frac{1}{n-1} & x \neq y;\ x, y \text{ jsou stupně } n-1 \\ 0 & x = y \end{cases}$$

k	Počet skupin	Poznámka
$k \geq \frac{1}{n-2}$	1	Všechny vrcholy jsou v jedné skupině.
$\frac{1}{n-2} > k \geq \frac{1}{n-1}$	3	Dvě skupiny, každou s jedním členem, tvoří oba vrcholy se stupněm $n-2$ a třetí skupinu tvoří všechny ostatní vrcholy.
$\frac{1}{n-1} > k$	n	Každý vrchol je sám ve své skupině.

- **Kompletní bipartitní graf $\mathbf{K_{m,n}}$**

  Kompletní bipartitní graf $K_{m,n}$, kde $m = |U|$ a $n = |W|$, je graf, jehož vrcholová množina je sjednocením dvou neprázdných disjunktních množin U, W, množina hran je $E = uw \;:\; u \in U \wedge w \in W$ (viz obr. 5.6).



Obrázek 5.6: Kompletní bipartitní graf $K_{2,3}$

Pro BÚNO $m \leq n$ platí: $\forall x, y \in K_m, n \;:\; \#(x,y) = \begin{cases} \frac{1}{m} & x, y \in W \\ \frac{1}{n} & x, y \in U \\ \frac{1}{m} & x \in U,\ y \in W \\ 0 & x = y \end{cases}$

k	Počet skupin	Poznámka
$k \geq \frac{1}{m}$	1	Všechny vrcholy jsou v jedné skupině.
$\frac{1}{m} > k \geq \frac{1}{n}$	n	Každou skupinu tvoří celý podgraf U s jedním vrcholem z podgrafu W.
$\frac{1}{n} > k$	$m + n$	Každý vrchol je sám ve své skupině.

- **Souvislý vrcholově tranzitivní graf**

  Graf G se nazývá vrcholově tranzitivní, jestliže $\forall u, v \in V(G)$ existuje v grafu G takový automorfismus φ, že $\varphi(u) = v$.

  Podle věty z [4] je hranová souvislost rovná stupni d. Protože hranová souvislost udává nejmenší $p(x, y),\ x \neq y$ a současně $(x, y) \geq d$, tak

  $$\#(x, y) = \begin{cases} \frac{1}{d} & x \neq y \\ 0 & x = y \end{cases}$$

  Z toho plyne, že ve vrcholově tranzitivních grafech jsou vrcholy buď všechny v jedné skupině, nebo každý sám ve svojí.

5.1.2 Závěr ze skupin určených metrikou

Můžeme si všimnout, že počet skupin určených pomocí metriky # nikdy nepřekročil počet vrcholů v grafu. Platí to proto, že relace $R_{\#}$ je relací ekvivalence (lemma 2), tzn. je i tranzitivní, takže pokud je x ve skupině s y a y ve skupině s z, pak z tranzitivity musí být x ve skupině s z.

Navíc tranzitivita způsobuje, že žádný vrchol nemůže být současně ve dvou různých skupinách, což v reálné sociální síti není vždy pravda. Pokud např. máme jako sociální síť žáky jedné školy, hrany mezi nimi odpovídají třeba přátelstvím na facebooku a chceme dostat skupiny odpovídající třídám, pak tato definice skupiny vhodná je. Pokud ale uvažujeme sociální síť vědců, hrany mezi nimi jsou spoluautorství na článcích, pak to příliš není vhodné, protože jeden vědec může být i paralelně ve více výzkumných skupinách.

5.2 Metriky vedoucí k většímu počtu skupin než je vrcholů

Z metriky # vychází najevo, že aby byl počet skupin větší než je vrcholů, metrika nesmí být relací ekvivalence. V souvislém neorientovaném grafu má smysl se snažit nalézt takovou, která poruší tranzitivitu.

- **Vzdálenost dvou vrcholů v souvislém grafu určená délkou nejkratší cesty**

  V souvislém neorientovaném grafu splňuje vlastnosti metriky triviálně.

Existují samozřejmě i další, jen se v této práci zabýváme podrobně hlavně metrikou #.

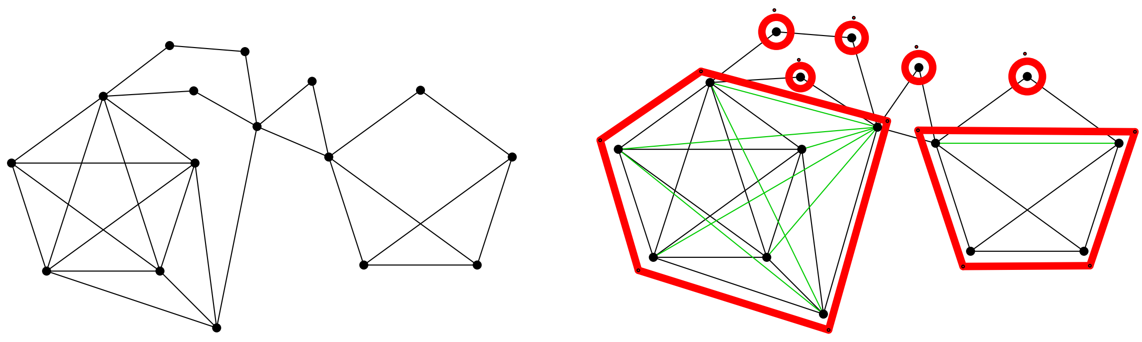
6. Uzávěr skupiny

Ideální by bylo doplnit hrany do grafu tak, že každá skupina bude kompletní podgraf, protože ne vždy jsou všichni, ač jsou spolu ve skupině, propojeni hranou. Způsob tohoto doplňování (nazvěme ho uzávěrem skupiny) ale není jednoduchý a bude třeba ověřit, zda pořadí přidávání hran neovlivní konečný výstup.

6.1 Uzávěr $U_\#$ pomocí metriky $\#$

Definice: Pokud v grafu není hrana xy a platí $\#(x,y) \leq k$, pak hranu x,y přidáme.

Na obrázku 6.1 vidíme příklad uzávěru $U_\#$ na grafu G pro $k = \frac{1}{3}$. Tedy doplníme hranu mezi takové vrcholy, mezi kterými zatím žádná nebyla a zároveň je mezi nimi aspoň tři hranově disjunktní cesty. Červeně jsou vyznačeny skupiny.



(a) Graf G před uzávěrem $U_\#$

(b) Graf G po uzávěru $U_\#$

Obrázek 6.1: Uzávěr $U_\#$ grafu G pro $k = \frac{1}{3}$

Chceme vědět, zda nám vždy vyjde po doplnění všech možných hran stejný graf, i když je budeme doplňovat v jiném pořadí. Dokážeme to pomocí silnějšího tvrzení, ze kterého plyne, že doplnění jedné hrany neovlivní doplňování ostatních. K důkazu budeme potřebovat pojmenovat několik věcí, proto je na konci jejich přehled.

Lemma 3. *Doplnění hrany yz do grafu G, který dosud takovou hranu neobsahoval, podle uzávěru $U_\#$ (tj. musí platit $\#(y,z) \leq k$) nezmění hodnotu $\#(x,y)$, $\forall x,y,z$.*

Důkaz. Dokážeme to sporem. Mějme graf G, ve kterém zatím nejsou hrany xy ani yz. Předpokládejme tedy, že i když v počátečním grafu G platí $\#(x,y) > k$ (tedy nelze přidat hranu xy) a zároveň platí $\#(y, z) \leq k$ (tedy lze doplnit hranu yz), tak po doplnění hrany yz platí $\#(x, y) \leq k$ (tedy mělo by jít doplnit hranu xy). Označme G' graf po doplnění hrany yz do grafu G.

Protože se lépe pracuje s $p(x, y)$ než s jeho převrácenou hodnotou $\#(x, y)$, zavedeme konstantu l tak, že platí:

$$l = \frac{1}{k}$$

Podmínka $p(x, y) \geq l$ je tak ekvivalentní s podmínkou $\#(x, y) \leq k$ na přidání hrany v definici uzávěru $U_\#$.

V grafu G musí být $p(x, y) < l$ a $p(x, z) \geq l$. Aby v grafu G' šlo uzávěrem doplnit hranu xy, musí platit $p(x, y) \geq l$ a jelikož přidáním jedné hrany xz určitě nemůže vzrůst $p(x, y)$ více než o jedna, tak v grafu G' musí platit $p(x, y) = l$. Z toho plyne, že v grafu G platí:

$$p(x, y) = l - 1$$

Označíme m počet hranově disjunktních (y, z)-cest takových, že neobsahují vrchol x a dále n počet hranově disjunktních (x, z)-cest takových, že neobsahují vrchol y. Když počítáme hranově disjunktní (x, y)-cesty takové, že se skládají pouze z předchozích dvou typů cest, tak z Mengerovy věty 2 vyplývá, že jich je aspoň minimum z hodnot m a n.

Když má po přidání hrany xz do grafu G vzniknout nová (x, y)-cesty, musí existovat taková (y, z)-cesta, která neobsahuje x a zároveň nebyla započítána jako část nějaké (x, y)-cesty v $p(x, y)$. Tato (y, z)-cesta by se pak napojila na novou hranu xz a vznikla by tím nová (x, y)-cesta, a tedy $p(x, y)$ by dosáhlo l. Ale aby taková (y, z)-cesta existovala, muselo by v grafu G platit $m > n$. To proto, že v opačném případě (tj $n \geq m$) by podle Mengerovy věty platilo, že by hranově disjunktních (x, z)-cest, bylo alespoň m a nezůstala by žádná požadovaná (y, z)-cesta.

Z toho plyne, že alespoň n je počet hranově disjunktních (x, y)-cest takových, které vznikly spojením hranově disjunktních cest (x, z), které neobsahují x, s hranově disjunktními (y, z)-cestami, které neobsahují x. Zbylých hranově disjunktních (x, y)-cest musí být $l - 1 - n$, aby jich dohromady bylo $l - 1$.

To ale znamená, že počet hranově disjunktních (x, z)-cest je n (cesty přímo z neobsahující y) plus minimum z $l - 1 - n$ a m (cesty, které vzniknou spojením (x, y) a (y, z)-cest, které neobsahují z a ty druhé x) podle Mengerovy věty. O m sice mnoho nevíme, nám stačí ohraničení $l - 1 - n$, protože tohle spolu s n dá dohromady $l - 1$ a ne l, jak byl původní předpoklad pro počet hranově disjunktních (x, z)-cest. To znamená, že přidáním hrany xz určitě nezměním $\#(x, y)$ za daných podmínek. $\square$

Uzávěr $U_\#$ je tedy jednoznačný, resp. ať přidáváme hrany v jakémkoliv pořadí, určitě to nezmění výsledek. To kdyby neplatilo, tak by to nebyl dobrý uzávěr.

Přehled proměnných a konstant v této kapitole:

- **k** - aby do grafu byla uzávěrem $U_\#$ doplněna hrana xy, musí platit podmínka $\#(x, y) \geq k$. Je to shodná konstanta i pro tvoření skupin na základě metriky $\#$, dá se nastavit podle toho, jak chceme, aby skupiny vypadaly

- $\mathbf{l} = \frac{1}{k}$, zavedené pro větší přehlednost
- **m** - počet hranově disjunktních (y, z)-cest v grafu G takových, že neobsahují vrchol x
- **n** - počet hranově disjunktních (x, z)-cest takových, že neobsahují vrchol y

7. Závěr

Asi největším úspěchem této práce je dokázání věty 1, která říká, že když vezmeme tři vrcholy v grafu, tak počet hranově disjunktních cest mezi dvěma dvojicemi z nich je stejný a mezi poslední dvojicí je větší. Jinak metrika # asi není nejlepší kritérium, pomocí kterého lze definovat skupiny. To proto, že počet hranově disjunktních cest mezi dvěma vrcholy není vždy tou rozhodující vlastností, která by indikovala, že zda spolu tyto vrcholy ve skupině jsou či nejsou. Navíc kvůli tranzitivitě relace $R_\#$ nemůže být jeden prvek v dvou různých skupinách, i když toto zrovna ne vždy vadí.

Na této práci jsem pracoval asi rok a získal jsem mnoho zkušeností. Do té doby jsem se s teorií grafů setkal maximálně v podobě kreslení obrázků na jeden tah, nebo hledání nejkratší cesty v grafu s ohodnocenými hranami. Rozšířila mi obzor o další krásný obor matematiky. Formulování matematických textů už dělám dlouho při řešení korespondenčních matematických seminářů, nicméně to nikdy nebyla ani zdaleka takto dlouhá práce, tedy mi její psaní dalo dost. Navíc jsem zdokonalil v práci s LaTeXem.

Rád bych se v budoucnu pokusil získat reálná data sociálních sítí a naprogramovat program, který by automaticky hledal skupiny na základě podmínek, které bych mu zadal. Dále v kapitole 3.1 je rozpracovaný jiný důkaz věty 1, který dokazuje silnější věc, tedy to, že původní výběr hranově disjunktních cest lze přeorganizovat do daných skupin. Ten se mi snad podaří dotáhnout do konce.

Literatura

[1] KASÍK, Pavel. Facebook ví, kdy se rozejdete. Ale přežije svou masovou popularitu?. Technet-cz [online]. 2014 [cit. 2015-03-31]. Dostupné z: http://technet.idnes.cz/budoucnost-a-rizika-facebooku-dkz-/sw_internet.aspx?c=A140106_135831_sw_internet_pka

[2] LORING, Terry. A Proof of Menger's Theorem. In: UNM Departmen of Mathematics and Stitistics [online]. The University of New Mexico, 2005 [cit. 2015-03-31]. Dostupné z: http://www.math.unm.edu/~loring/links/graph_s05/Menger.pdf

[3] KOVÁŘ, Petr. Teorie grafů. Ostrava, Plzeň: VŠB-TU Ostrava, Západočeská univerzita Plzeň, 2012

[4] GODSIL, Chris a ROYLE, Gordon. Algebraic graph theory. New York: Springer, c2001, xix, 439 s. Graduate texts in mathematics. ISBN 0-387-95241-1.