

UNIVERZITA KOMENSKÉHO V BRATISLAVE

FAKULTA MATEMATIKY, FYZIKY  
A INFORMATIKY

FENOMÉN “MALÉHO SVETA”  
V SOCIÁLNYCH SIETĎACH

bakalárska práca

UNIVERZITA KOMENSKÉHO V BRATISLAVE  
FAKULTA MATEMATIKY, FYZIKY  
A INFORMATIKY

FENOMÉN “MALÉHO SVETA”  
V SOCIÁLNYCH SIEŤACH

bakalárska práca

Študijný program: Poistná matematika

Študijný odbor: 6211 Štatistika

Školiace pracovisko: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky

Školiteľ: RNDr. Beáta Stehlíková, PhD.



Univerzita Komenského v Bratislave  
Fakulta matematiky, fyziky a informatiky

## ZADANIE ZÁVEREČNEJ PRÁCE

**Meno a priezvisko študenta:** Kristína Fukasová  
**Študijný program:** poisťná matematika (Jednoodborové štúdium, bakalársky I. st., denná forma)  
**Študijný odbor:** štatistika  
**Typ záverečnej práce:** bakalárska  
**Jazyk záverečnej práce:** slovenský  
**Sekundárny jazyk:** anglický

**Názov:** Fenomén "malého sveta" v sociálnych sieťach  
*"Small world" phenomenon in social networks*

**Cieľ:** Pojem "malý svet" sa vzťahuje na známu situáciu, keď pri stretnutí s neznámym človekom zistíme, že máme spoločných známych.

Jednoduchý model náhodného grafu je tzv. Erdős-Rényiho graf, v grafoch tohto typu môžeme pozorovať vlastnosť "malého sveta", teda malé vzdialenosti medzi vrcholmi. Toto sa dá vidieť simulovaním takýchto grafov, ako aj matematickými odvodzeniami - v práci sa ukážu oba prístupy.

V čom je však rozdiel, je vytváranie zhlukov. V realite je častá situácia typu "ak sú A aj B moji kamaráti, tak aj A a B sú kamaráti". Erdős-Rényiho graf však túto vlastnosť nemá. Miera zhlukovania sa dá matematicky popísať a dajú sa pomocou tohto popisu grafy analyzovať, čo sa v práci tiež spraví.

Súčasťou práce bude aj zozbieranie potrebných dát a následná analýza (vzdialenosti vrcholov, zhlukovanie) niekoľkých reálnych sietí.

Nakoľko Erdős-Rényiho graf je dobrým modelom malých svetov, ale nevystihuje dobre zhlukovanie, existujú aj iné prístupy k vytváraniu náhodných grafov, ktoré by mohli slúžiť na modelovanie reálnych sietí. Práca sa ukončí uvedením niektorých z nich, ich vlastností, ukážkami vygenerovaných grafov, referenciami na literatúru, kde sa tieto prístupy použili na modelovanie reálnych situácií.

**Vedúci:** RNDr. Beáta Stehlíková, PhD.  
**Katedra:** FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky  
**Vedúci katedry:** prof. RNDr. Daniel Ševčovič, CSc.  
**Dátum zadania:** 28.10.2015

**Dátum schválenia:** 29.10.2015  
doc. RNDr. Katarína Janková, CSc.  
garant študijného programu

.....  
študent

.....  
vedúci práce

Úprimné poďakovanie patrí mojej školiteľke RnDr. Beáte Stehlíkovej, PhD. za ochotu, pomoc pri práci a za všetok čas a námahu, ktorú vynaložila, aby ma viedla pri písaní tejto bakalárskej práce.

# Abstrakt

V tejto bakalárskej práci je predstavený model náhodných sietí Erdősa a Rényiho, v ktorom pozorujeme fenomén “malého sveta”. Vlastnosti spomínaného modelu sú matematicky odvodené, a tiež ukázané na konkrétnych simulovaných náhodných sieťach. Práca obsahuje základné pojmy z teórie grafov ilustrované na príkladoch a najdôležitejšie vlastnosti reálnych grafov spolu s ilustračnými obrázkami. Tiež sa v nej nachádza analýza reálnych sociálnych sietí zobrazených z nazbieraných dát a následné porovnanie týchto reálnych sietí s modelom náhodných sietí Erdősa a Rényiho. Fenomén malého sveta je vlastnosť mnohých sietí, ktorá sa vzťahuje k dvom podobným, ale odlišným vlastnostiam: malé vzdialenosti medzi vrcholmi a veľká miera zhlukovania. Keďže sa ukáže, že model Erdősa a Rényiho nepopisuje dobre zhlukovanie, nakoniec sú uvedené aj iné prístupy k vytváraniu náhodných grafov.

Kľúčové slová: fenomén malého sveta, model Erdősa a Rényiho, náhodné siete, zhlukovanie.

# Abstract

In this bachelor thesis is introduced Erdős-Rényi model, which has “small world” properties. Properties of mentioned model are mathematically derived and shown at concrete simulation of random network. Thesis contains basics concepts of graph theory, illustrated in the examples and the most important properties of real graphs all together with illustrative images. It contains real social network analysis by using the collected data and subsequent comparison of real networks with random network model of Erdős and Rényi as well. Small world phenomenon is a feature of many networks, which relates to two similar but distinct characteristics: small distance between the vertices and the large degree of clustering. Since it turned out that the model of Erdős and Rényi does not describe clustering appropriately, eventually are mentioned also other approaches to the creation of random graphs.

Klíčové slová: small world phenomenon, Erdős-Rényih model, random network, clustering.

# Obsah

<b>1</b>	<b>Siete a ich charakteristiky</b>	<b>10</b>
1.1	Sociálne siete . . . . .	10
1.2	Základné pojmy . . . . .	11
1.3	Náhodné siete . . . . .	13
<b>2</b>	<b>Náhodné siete Erdősa a Rényiho</b>	<b>15</b>
2.1	Rozdelenie stupňa uzlov . . . . .	16
2.2	Vlastnosti a prahové funkcie pre siete Erdősa a Rényiho . . . . .	18
2.2.1	Existencia aspoň jednej hrany . . . . .	19
2.2.2	Ďalšie prahové funkcie . . . . .	21
2.2.3	Súvislosť náhodných sietí . . . . .	22
<b>3</b>	<b>Vlastnosti reálnych sociálnych sietí</b>	<b>29</b>
3.1	Malý svet . . . . .	29
3.1.1	Priemer náhodných sietí . . . . .	29
3.2	Zhlukovanie . . . . .	31
3.3	Homofília . . . . .	34
<b>4</b>	<b>Analýza reálnych sociálnych sietí</b>	<b>36</b>
4.1	Sieť poistných matematikov . . . . .	36
4.2	Sieť hudobníkov . . . . .	37
4.3	Sieť vedcov . . . . .	38
<b>5</b>	<b>Ďalšie modely pre sociálne siete</b>	<b>40</b>
5.1	Model malého sveta . . . . .	40
5.2	Mocninový model . . . . .	41





# Úvod

Siete sú všade okolo nás. Či už sú to trate električiek, potravinový reťazec alebo internet. Dokonca aj my sami sme ich súčasťou. Siete vzťahov medzi ľuďmi voláme sociálne siete. Tie ovplyvňujú celý chod spoločnosti. Odovzdávame si informácie, zdieľame názory, ovplyvňujeme navzájom svoje rozhodnutia atď.

Známy je Milgramov experiment, ktorý spočíval v tom, že náhodne vybraný človek mal doručiť balíček druhému náhodne vybranému človeku tak, že ho odovzdal svojmu známemu, ten zase svojmu a tak ďalej. Z experimentu vyplynulo niekoľko záverov. Jedným z nich bol, že z tých balíčkov, ktoré sa podarilo doručiť, bol priemerný počet takýchto poslaní zásielok šesť, čo je možno menšie číslo, než by sme čakali. Tento fenomén voláme malý svet. Úlohou tejto bakalárskej práce je matematicky popísať fenomén malého sveta a analyzovať ďalšie vzťahy v sociálnych sieťach, pomocou teórie grafov a pravdepodobnosti. Hlavným zdrojom, z ktorého sme pri tejto práci čerpali bol kurz “Social and Economic Networks: Models and Analysis” Matthewa Jacksona, profesora Stanfordskej univerzity na stránke coursera.org. V práci sme tiež použili software Gephi, ktorý využívame na simuláciu a zobrazovanie náhodných sietí.

Hlavným cieľom práce je oboznámiť čitateľa s problematikou, popísať siete pomocou náhodných grafov a zhodnotiť ich dobré vlastnosti a nedostatky. Vysvetliť, v čom tieto nedostatky spočívajú, prečo sa reálne siete odlišujú od náhodne vygenerovaných sietí a navrhnúť alternatívne modely, ktoré môžu byť použité na modelovanie reálnych sociálnych sietí. Zistené skutočnosti sa budeme snažiť zobrazovať na zozbieraných dátach.

Práca sa člení na päť kapitol. Prvá kapitola uvádza teoretické základy a definície z teórie grafov, ktoré budú využité v ďalších častiach práce. Napríklad definície priemeru siete, najkratšej cesty a tak ďalej. Čitateľ sa tiež dozvie, čo sú to náhodné siete a v čom spočíva ich náhodnosť.

Druhá kapitola sa venuje modelu Erdősa a Rényiho. Sú tu uvedené najdôležitejšie

vlastnosti a prahové funkcie modelu náhodných sietí Erdősa a Rényiho. Tieto vlastnosti sú ilustrované na príkladoch simulovaných sietí.

Tretia kapitola sa zaoberá vlastnosťami, ktoré sú špecifické pre reálne sociálne siete. Ukážeme si, ako závisí priemer siete od počtu vrcholov, zavedieme dve rozličné miery zhlukovania a vlastnosť homofílie, ktorú si ukážeme na príklade študentov amerických stredných škôl.

V nasledujúcej časti sú analyzované a zobrazené reálne sociálne siete.

Obsahom záverečnej, piatej časti sú ďalšie modely a ich vlastnosti, ktorými môžeme modelovať malý svet.

# Kapitola 1

## Siete a ich charakteristiky

### 1.1 Sociálne siete

Pod pojmom sieť rozumieme množinu objektov, ktoré sú medzi sebou nejakým spôsobom poprepájané. Môžeme hovoriť o priateľstvách medzi ľuďmi, počítačových sieťach, či neutrónoch v mozgu. Všetko toto sú siete, avšak sú veľmi odlišné. Na to, aby sme ich mohli viac charakterizovať, potrebujeme vedieť nielen čo to siete sú, ale aj aké rôzne druhy sietí vieme vyrobiť.

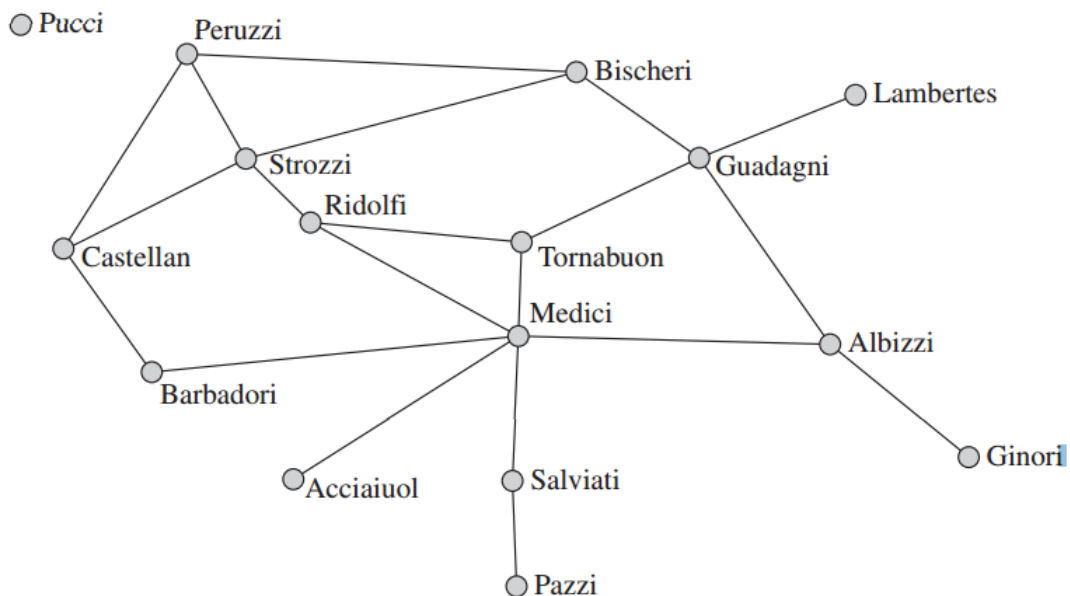
Veľa ekonomických, politických a sociálnych vzťahov je založených na štruktúre sociálnych sietí. Tieto siete zahŕňajú vzťahy medzi príbuznými a priateľmi, s ktorými pravidelne zdieľame informácie, láskavosti, názory. Sú to tiež vzťahy medzi svetovými spoločnosťami, ktoré sa navzájom ovplyvňujú pri rozhodovaní, kde a s kým budú obchodovať. Aj keď siete vykazujú určitú rôznorodosť, majú rovnakú základnú štruktúru vzťahov. Tie vieme matematicky analyzovať.

Začiatky teórie a sietí sa spájajú s rokom 1736, keď Leonhard Euler, jeden z najväčších matematikov vôbec, riešil problém v pruskom meste Königsberg, ako prejsť cez sedem mostov bez toho, aby prešiel po niektorom dva krát. Dokázal, že to nie je možné a tým položil základy teórii grafov.[6] Teraz jej rozsah siaha do mnohých odlišných vedných odborov ako sociológia, ekonomika, počítačové vedy, fyzika, matematika a iné.

Príkladom nám môže byť sociálna sieť (viď. Obr. 1.1), kde vrcholy sú rodiny vo Florencii v 15. storočí a neorientované hrany predstavujú manželstvá medzi nimi. V minulosti bola Florencia riadená oligarchiou elitných rodín.

Túto sieť analyzovali Padgett a Ansell (1993), ktorí vo svojej knihe vysvetľujú kľúčovú rolu rodiny Mediciovcov. Tá bola svojho času najmocnejšou rodinou či už politicky alebo obchodne. Manželské vzťahy boli kľúčom k prenosu informácií, obchodných ponúk či dosiahnutie politického rozhodnutia. Z tohto hľadiska sú Mediciovcí v oveľa lepšej pozícii než ostatné rodiny. Padgett a Ansell zdôrazňujú, že postavenie Mediciovcov v sieti nie je vecou náhody, ale starostlivého plánovania. Tu si môžeme uvedomiť, aké dôležité je modelovanie sietí.

Obr. 1.1: Manžestvá medzi rodinami vo Florencii, 15. storočie (Zdroj: Padgett a Ansell (1993))



## 1.2 Základné pojmy

**Definícia 1.1.** Definujme  $N = \{1, \dots, n\}$  ako množinu uzlov a maticu susednosti  $g \in \{0, 1\}^{n \times n}$ , kde  $g_{ij} = 1$ , ak medzi  $ij$  je väzba. Túto väzbu budeme volať hrana.

Sieť (graf)  $(N, g)$  je potom charakterizovaná množinou uzlov  $N$  a maticou  $g$ .

Zaveďme tiež označenie  $ij \in g$ , ktoré bude znamenať  $g_{ij} = 1$  a  $ij \notin g$ , ktoré bude znamenať  $g_{ij} = 0$ .

Sieť je orientovaná, ak existuje dvojica  $ij$  pre ktoré platí  $g_{ij} \neq g_{ji}$ .

Sieť je neorientovaná, ak pre všetky  $ij$  platí, že  $g_{ij} = g_{ji}$ .

Napr. pre orientovaný graf na Obr. 5.3 bude matica susednosti

$$g = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

a pre neorientovaný graf na Obr. 5.4 bude matica susednosti

$$g = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

**Definícia 1.2.** Ak  $g$  je matica susednosti a  $i$  je uzol, označme okolie uzla  $i \in g$ ,  $N_i(g) = \{j | ij \in g\}$ .

**Definícia 1.3.** Stupeň  $d_i = \#N_i(g)$  reprezentuje počet väzieb, ktoré uzol má.

**Definícia 1.4.** Prechádzkou z  $i_1 \rightarrow i_k$  nazveme postupnosť uzlov  $(i_1, i_2, \dots, i_k)$  a postupnosť väzieb  $(i_1i_2, i_2i_3, \dots, i_{k-1}i_k)$  takých, že  $i_{k-1}i_k \in g$ , pre všetky  $k$ .

Ak sú všetky  $i_k$  odlišné, túto prechádzku nazveme cestou. Jej dĺžka je  $k - 1$ .

**Definícia 1.5.** Sieť  $(N, g)$  je súvislá, ak existuje cesta medzi každými dvomi vrcholmi.

**Definícia 1.6.** Komponent je maximálny súvislý podgraf. To je,  $(N', g')$  je komponent  $(N, g)$  ak:

- $(N', g') \subset (N, g)$ ,
- $(N', g')$  je súvislý,
- $i \in N' \wedge ij \in g \Rightarrow j \in N' \wedge ij \in g'$ .

**Definícia 1.7.** Priemer siete  $(N, g)$  je maximálna najkratšia cesta. Vypočítame ho ako  $\bar{d}(g) = \max_{ij} d(i, j)$ , kde  $d(i, j)$  je najkratšia dĺžka cesty. Ak je sieť nesúvislá, tak priemer je nekonečno. Keďže veľa sociálnych sietí je nesúvislých, často sa priemer uvádza pre najväčší komponent.

Ukážme si tieto pojmy na príklade (Obr.1.1), vplyvu rodín vo Florencii. Táto sieť je neorientovaná, keďže manželstvo je obojstranný vzťah. Má dva komponenty. Jedným z nich je izolovaný uzol (Pucciovcí) a druhým komponentom sú všetky ostatné rody.

Keďže rod Pucciovcov je izolovaným uzlom, okolie bodu je prázdna množina, a teda aj stupeň uzla je rovný nule. Dominantnú pozíciu mala v tom čase rodina Mediciovcov. Prečo, to dobre vidíme z grafu. Má najväčšie okolie, to znamená najviac susedstiev, aj najväčší stupeň (šesť).

Ďalším pojmom ktorý sme už definovali, je prechádzka. Napríklad, prechádzkou z uzla Albizzi do Ridolfi môže byť postupnosť: (Albizzi, Medici, Tornabuon, Guadagni, Tornabuon, Ridolfi). Táto prechádzka nie je cestou, pretože sa tam dva krát vyskytuje rodina Tornabuoncov. Cestou, a zároveň aj najkratšou cestou (dĺžky dva) z Albizzi do Ridolfi, je postupnosť (Albizzi, Medici, Ridolfi).

Keďže je sieť nesúvislá, priemer siete je nekonečno. Ako sme už spomenuli, často sa zvykne uvádzať priemer najväčšieho komponentu. V tom prípade, je priemer siete päť. Napríklad z vrcholu Peruzzi do vrcholu Pazzi neexistuje kratšia cesta ako päť. Tiež z vrcholu Pazzi do vrcholu Lambertes.

### 1.3 Náhodné siete

Táto kapitola sa zaoberá náhodnými sieťami (grafmi). V čom spočíva ich náhodnosť? Pojmom náhodný graf označujeme graf, ktorého hrany, vrcholy alebo oboje sú generované náhodným procesom. Predstavme si takúto situáciu: Chceme modelovať sieť kamarátstiev žiakov 1.A triedy (predpokladajme, že tried prvého ročníka je viac). Vrcholy siete sú žiaci 1.A triedy. Hranou spojíme tých dvoch žiakov, ktorí sa skamarátili. Treba si uvedomiť, že to, ako budú žiaci rozmiestnení do tried, je náhodná udalosť. Taktiež, po vytvorení tried je náhodné aj vznikanie hrán (tzn. náhodné je aj to, kto s kým sa skamaráti).

V reálnych sieťach sa často stretávame s takzvaným rastom siete. Do siete vrcholy pribúdajú a pripájajú sa novými hranami k už existujúcim vrcholom. Napríklad: vytvorenie novej stránky na internete, či nadviazanie kontaktu medzi ľuďmi v sociálnych sieťach.

Skoro všetky modely, ktorými sa v bakalárskej práci zaoberáme, sú však takzvané

statické modely. Náhodný graf vzniká spôsobom, pri ktorom sa na začiatku pevne určí počet vrcholov, a potom sa medzi týmito vrcholmi náhodne vytvárajú hrany. Nakoľko Erdős-Rényiho graf je dobrým modelom malých svetov, ale nevystihuje dobre zhľukovanie, existujú aj iné prístupy k vytváraniu náhodných grafov, ktoré by mohli slúžiť na modelovanie reálnych sietí. Práca sa ukončí uvedením niektorých z nich,... Aj keď sú tieto siete statické, vo veľa prípadoch nás zaujíma, ako sa sieť zmení, ak  $n \rightarrow \infty$ . Množina takto vygenerovaných grafov spolu s pravdepodobnostnou funkciou vytvára diskretný pravdepodobnostný priestor  $(\Omega, S, P)$ , kde  $\Omega$  je množina všetkých sietí na týchto  $n$  vrcholoch.  $\sigma$ -algebra obsahuje všetky podmnožiny  $\Omega$ . Model spočíva v tom, že sa určí pravdepodobnostná miera  $P$ , teda pre každú sieť sa dá povedať, s akou pravdepodobnosťou vzniká.

# Kapitola 2

## Náhodné siete Erdősa a Rényiho

Paul Erdős (1913-1996) bol legendárny maďarský matematik, ktorý do svojej smrti napísal viac než 1500 matematických článkov. Niektoré z nich publikoval spoločne s ďalším maďarským matematikom, Alfrédom Rényim (1921-1970). Medzi nimi boli aj články zaoberajúce sa otázkou nášho vzájomne previazaného sveta, otázkou, ako vlastne vznikajú siete<sup>1</sup>. Ich výsledky položili základy teórie náhodných sietí.

Zjednodušenie všetkých reálnych sietí na grafy nás privedie k rôznym problémom. Napriek tomu, že ľudská spoločnosť, internet, bunka či mozog môžu byť dobre reprezentované pomocou grafu, rozhodne sa navzájom líšia tým, že vznikajú s veľmi odlišnými pravidlami. Tieto problémy vyriešili Erdős a Rényi tak, že zanedbali rozdiely vzniku siete a začali uzly spájať náhodne. Pravidlo znelo takto: Vyber dva uzly, hoď kockou, ak padne šestka spoj ich hranou, ak padne čokoľvek iné, nespájaj ich. [5]

**Definícia 2.1.** Majme graf, ktorý má  $n$  uzlov, pričom väzby medzi ľubovoľnými dvoma uzlami sú tvorené náhodne, nezávisle od seba s rovnakou pravdepodobnosťou  $p$ . Takýto graf nazveme graf Erdősa a Rényiho.

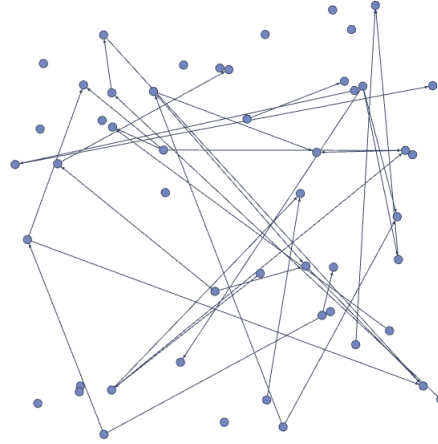
Na Obr. 2.1 môžeme vidieť náhodne vygenerovaný graf s  $n = 50$  uzlami. Pravdepodobnosť že medzi niektorými dvoma uzlami vznikne hrana je  $p = 0.03$ .

---

<sup>1</sup>On random graphs (1959), On the evolution of random graphs (1960), On the strength of connectedness of a random graph (1961), Asymmetric graphs (1963), On random matrices (1964), On a problem of graph theory (1966), On the existence of a factor of degree one of a connected random graph (1966), On random matrices II (1968) [18]



Obr. 2.1: Príklad náhodne vygenerovanej siete s  $n = 50$  uzlami a  $p = 0,03$  (Zdroj: vlastné spracovanie).



## 2.1 Rozdelenie stupňa uzlov

Dôležitá vlastnosť sociálnych sietí je rozdelenie stupňa uzlov. Napríklad, stupne uzlov v príklade rodín (Obr.1.1) sú: 0(Puzzi), 1(Pazzi), 1(Acciaiuol), 1(Ginori), 1(Lambertes), 2(Salviati), 2(Barbadori), 3(Albizzi), 3 (Bischeri), 3(Peruzzi), 3(Tornabuon), 3(Caltellan), 3(Ridolfi), 4(Strozzi), 4(Guadagni), 6(Medici). Z čoho môžeme vidieť, že Mediciovci majú viac než dva krát väčší stupeň než priemerný stupeň, a dva krát väčší stupeň než medián celej siete.

Rozdelenie stupňa uzlov siete  $(N, g)$ , označované ako  $P(d)$ , je pravdepodobnosť, že vrchol bude mať stupeň  $d$ .

Erdős–Rényiho siete s veľkým počtom vrcholov zvykneme nazývať aj Poissonove náhodné siete. Súvisí to s aproximáciou stupňa vrcholov, ktorú odvodíme v nasledujúcejvete. Majme daný Erdős–Rényiho graf a rozdelenie stupňa uzlov  $P(d)$ . Potom stupne vrcholov majú binomické rozdelenie,

$$P(d) = \binom{n-1}{d} p^d (1-p)^{n-d-1}.$$

**Veta 2.1.1.** Ak  $m \rightarrow \infty$  a  $p \rightarrow 0$  takým spôsobom, že  $mp = \lambda$ , kde  $\lambda$  je kladná konštanta a  $m = (n-1)$ , konverguje  $d$ -ty člen binomického rozdelenia ku  $d$ -temu členu Poissonovho rozdelenia. [16]

*Dôkaz.* Dôkaz vety na základe [17].

$$P(d) = \frac{(n-1)!}{d!(n-d-1)!} p^d (1-p)^{n-d-1}$$

Zaveďme substitúciu, aby sme mali štandardný tvar binomického rozdelenia:

$$m = (n-1).$$

$$P(d) = \frac{m!}{d!(m-d)!} p^d (1-p)^{m-d}$$

Z predpokladu  $mp = \lambda \Rightarrow p = \frac{\lambda}{m}$ .

$$P(d) = \frac{m!}{d!(m-d)!} \frac{\lambda^d}{m^d} \left(1 - \frac{\lambda}{m}\right)^{m-d}.$$

Predpoklad  $mp = \lambda$  pre  $m \rightarrow \infty$  môžeme zrejme nahradiť výrazom  $p = \lim_{m \rightarrow \infty} \frac{\lambda}{m}$ .

Pravdepodobnosť  $p \rightarrow 0$ . Z toho hneď plynie, že platí:

$$1 = \lim_{m \rightarrow \infty} \left(1 - \frac{\lambda}{m}\right)^{-d}. \quad (2.1)$$

Vypočítame teraz limitu ako:

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{m!}{(m-k)!m^k} &= \lim_{m \rightarrow \infty} \frac{m(m-1)\dots(m-k+1)(m-k)!}{(m-k)!m^k} \\ &= \lim_{m \rightarrow \infty} \frac{m}{m} \frac{m-1}{m} \dots \frac{m-k+1}{m} = 1 \end{aligned} \quad (2.2)$$

Využitím vzorca  $\lim_{x \rightarrow 0} (1+x)^{\frac{1}{x}} = e$  a zavedením substitúcie  $\frac{-\lambda}{m} = x$  dostávame

$$e^{-\lambda} = \lim_{m \rightarrow \infty} \left(1 - \frac{\lambda}{m}\right)^m. \quad (2.3)$$

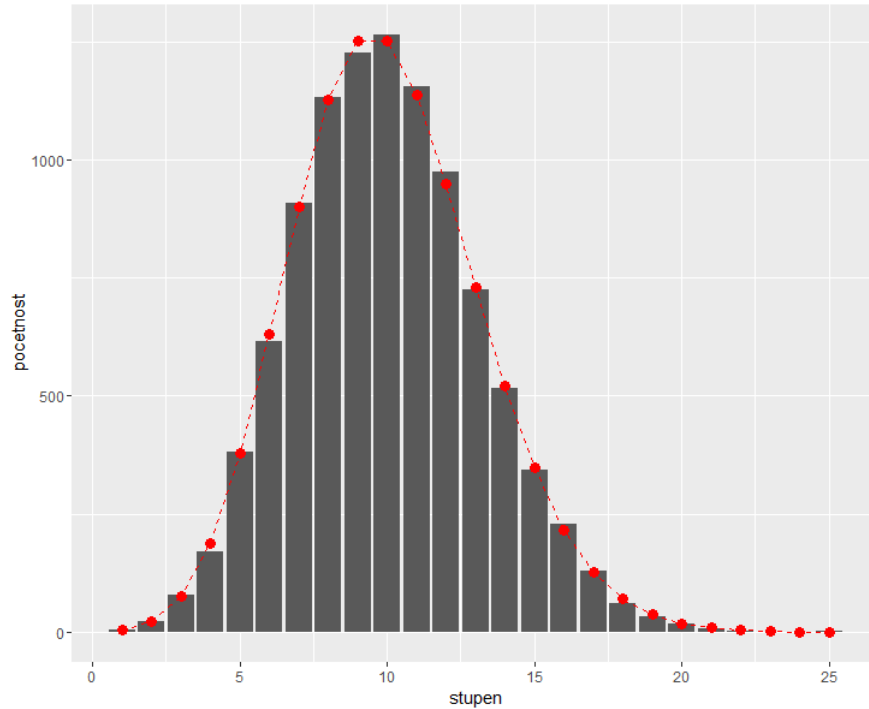
Využitím predošlých výpočtov (rovnice 2.1, 2.1, 2.1) a predpokladu že  $m \rightarrow \infty$  a  $p \rightarrow 0$  takým spôsobom, že  $mp = \lambda$  dostávame Poissonove rozdelenie,

$$\begin{aligned} P(d) &\rightarrow \frac{e^{-mp}(mp)^d}{d!} \\ P(d) &\rightarrow \frac{e^{-(n-1)p}((n-1)p)^d}{d!}. \end{aligned}$$

□

Z grafu (Obr. 2.2) vidíme, že pre veľké  $n$  a malé  $p$  je Poissonove rozdelenie dobrou aproximáciou binomického rozdelenia. Takéto siete, ktoré sú náhodne vygenerované s rovnakou pravdepodobnosťou sa stali akýmsi štandardom, s ktorými sa oplatí siete, ktoré nie sú vytvorené náhodným procesom, porovnávať.

Obr. 2.2: Aproximácia binomického rozdelenia Poissonovým. Dáta(10000) vygenerované z binomického rozdelenia s pravdepodobnosťou  $p = 0,001$  tvoria histogram, červená krivka predstavuje Poissonove rozdelenie. (Zdroj: vlastné spracovanie)



## 2.2 Vlastnosti a prahové funkcie pre siete Erdősa a Rényiho

Uvažujme Erdős–Rényiho model pre veľké  $n$  na množine  $n$  vrcholov, kde index pravdepodobnosti vzniku hrany je funkciou  $n$ , označovaný  $p(n)$ . Je prirodzené, že niektoré vlastnosti siete závisia od počtu uzlov. Pravdepodobnosť, že sieť danú vlastnosť má, sa s rastúcim  $n$  môže prudko meniť medzi 0 a 1. Napríklad sa môžeme pýtať, či v sieti vzniknú nejaké izolované uzly. Pre niektoré náhodne generované siete je takmer isté že vznikne aspoň jeden izolovaný uzol (v prípade že sieť je veľká), o iných náhodne generovaných sieťach môže platiť, že takmer iste nevzniknú žiadne izolované body.

Takéto vlastnosti sú zvyčajne špecifikované ako množina sietí, pre každé  $n$ , pre ktoré je daná vlastnosť splnená. Označme,  $G(N)$  je množina všetkých neorientovaných sietí na množine  $N$  vrcholov. Vlastnosť je množina takých  $A(N)$ , pre každé  $N$ , že  $A(N) \subset G(N)$ . Napríklad vlastnosť, že sieť nemá žiadne izolované uzly zapíšeme ako  $A(N) = \{g | N_i(g) = \emptyset \text{ pre } \forall i \in N\}$ .

**Definícia 2.2.** Vlastnosť siete je monotónna, ak  $g \in A(N) \wedge g \subset g' \Rightarrow g' \in A(N)$ .

**Definícia 2.3.** Funkcia  $t(n)$  je prahová funkcia pre monotónnu vlastnosť  $A(n)$ , ak  $A(n)$  je funkcia  $t(n)$  taká, že

$$\lim_{n \rightarrow \infty} Pr [A(n)|p(n)] = \begin{cases} 0 & \text{pre } \frac{p(n)}{t(n)} \rightarrow 0, \\ 1 & \text{pre } \frac{p(n)}{t(n)} \rightarrow \infty. \end{cases}$$

Ak existuje prahová funkcia, hovorí sa, že v  $t(n)$  nastane fázový prechod.

## 2.2.1 Existencia aspoň jednej hrany

Veta, ktorá je obsahom tejto podkapitoly, hovorí o prahovej funkcii existencie aspoň jednej hrany. Dokážeme ju dvoma spôsobmi. Najprv priamo a potom iným spôsobom, v ktorom využijeme Markovu a Čebyševovu nerovnosť (2.2.2 a 2.2.3). Druhý spôsob uvádzame preto, že dôkaz má podobnú myšlienku, ako zložitejší dôkaz, ktorý uvádzame neskôr.

**Veta 2.2.1.** [12] *Nech  $X$  je náhodná veličina s konečnou strednou hodnotou  $\mu = E(X^r)$ . Potom pre ľubovoľné  $\epsilon > 0$  platí*

$$P(|X| \geq \epsilon) \leq \frac{E(|X|^r)}{\epsilon^r}.$$

*Dôkaz.* [12] Nech  $F(x)$  je distribučná funkcia náhodnej premennej  $X$ , potom  $E(|X|^r) = \int_{-\infty}^{\infty} |x|^r dF \geq \int_{|x| \geq \epsilon} |x|^r dF \geq \int_{|x| \geq \epsilon} \epsilon^r dF = \epsilon^r \lambda_F(|x| \geq \epsilon) = \epsilon^r Pr(|x| \geq \epsilon)$ .

Kde  $\lambda_F(|x| \geq \epsilon)$  je miera množiny  $\{|x| \geq \epsilon\}$ . □

Nasledujúce vety (2.2.2 a 2.2.3) sú dôsledkami vety 2.2.1.

**Veta 2.2.2** (Markovova nerovnosť). [12] *Nech  $X$  je náhodná veličina s konečnou strednou hodnotou  $E(X)$ . Potom pre ľubovoľné  $\epsilon > 0$  platí*

$$P(|X| \geq \epsilon) \leq \frac{E(|Y|)}{\epsilon}.$$

*Dôkaz.* Markovova nerovnosť je len špeciálny prípad vety 2.2.1 pre  $r = 1$ . □

**Veta 2.2.3** (Čebyševova nerovnosť). [12] *Nech  $X$  je náhodná veličina s konečnou strednou hodnotou  $E(X)$  a kladnou disperziou  $D(X)$ . Potom pre ľubovoľné  $\epsilon > 0$  platí*

$$P(|X - E(X)| \geq \epsilon) \leq \frac{D(X)}{\epsilon^2}.$$

*Dôkaz.* Čebyševova nerovnosť je špeciálny prípad vety 2.2.1 pre  $r = 2$ .

$$P(\{\omega \in \Omega : |X(\omega) - E(x)| \geq \epsilon\}) \leq \frac{E[(X - E(X))^2]}{\epsilon^2}.$$

□

**Veta 2.2.4.** [8] *Nech  $A(n)$  je vlastnosť definovaná ako tvrdenie, že v náhodnej sieti  $G_{n,p}$  vznikne aspoň jedna hrana, potom  $t(n) = \frac{1}{n^2}$  je prahová funkcia pre monotónnu vlastnosť  $A(n)$ .*

$$\lim_{n \rightarrow \infty} Pr [A(n)|p(n)] = \begin{cases} 0 & \text{pre } \frac{p(n)}{n^2} \rightarrow 0, \\ 1 & \text{pre } \frac{p(n)}{n^2} \rightarrow \infty. \end{cases}$$

*Dôkaz 1.* Majme náhodnú premennú  $X$ , ktorá reprezentuje počet hrán. Teda,  $X$  má binomické rozdelenie s parametrami  $X \sim bin(\binom{n}{2}, p)$ . Najprv dokážme tvrdenie, že pre  $p(n)n^2 \rightarrow \infty$  sa pravdepodobnosť, že vznikne aspoň jedna hrana bude blížiť k jednotke. Počítajme opačnú pravdepodobnosť. Dokážme, že pravdepodobnosť, že ne- vznikne ani jedna hrana pôjde k nule.

$$P(X = 0) = (1 - p(n))^{\binom{n}{2}} = e^{\ln(1-p) \frac{n(n-1)}{2}} = e^{\frac{n(n-1)}{2} \ln(1-p)} \leq e^{\frac{n(n-1)}{2} (-p)} = e^{-\frac{n(n-1)}{2} \frac{pn^2}{n^2}} \rightarrow 0,$$

kde sme využili, že vieme  $\ln(1+x) \leq x$  pre  $x > -1$ . Z toho vyplýva, že  $\ln(1-p) \leq -p$ .

Keďže sme počítali opačnú pravdepodobnosť, pravdepodobnosť, že vznikne aspoň jedna hrana ide v tomto prípade do jednotky, čo bolo treba dokázať.

Teraz ukážeme, že pre  $p(n)n^2 \rightarrow 0$  sa pravdepodobnosť, že vznikne aspoň jedna hrana bude blížiť k nule:

$$P(X = 0) = (1 - p(n))^{\binom{n}{2}} = e^{\ln(1-p) \frac{n(n-1)}{2}} = e^{-\frac{n(n-1)}{2} \ln(1-p)}, \text{ kde pre exponent platí, že:}$$

$$\lim_{pn^2 \rightarrow 0} \frac{-n(n-1)}{2} \ln(1-p) = \frac{-n(n-1)}{2} \frac{p \ln(1-p)}{p} = \frac{-n(n-1)}{2n^2} \frac{pn^2 \ln(1-p)}{p} \rightarrow 1, \text{ lebo}$$

$$\text{platí, že } \lim_{p \rightarrow 0} \frac{\ln(1-p)}{p} = \lim_{p \rightarrow 0} \frac{\frac{1}{1-p}}{1} = 1. \text{ Pravdepodobnosť } p \rightarrow 0 \text{ pretože } p = \frac{pn^2}{n^2} \rightarrow 0.$$

Počítali sme opačnú pravdepodobnosť, a teda pravdepodobnosť, že vznikne aspoň jedna hrana ide v tomto prípade do nuly, čo bolo treba dokázať.

*Dôkaz 2.* Dokážme prvý prípad, čiže pre  $\frac{p(n)}{n^2} \rightarrow 0$ . Nech  $X$  je náhodná premenná reprezentujúca počet hrán v grafe.  $X$  má binomické rozdelenie s parametrami  $X \sim bin(\binom{n}{2}, p)$  a teda jej stredná hodnota je  $E(X) = \binom{n}{2}p$  a disperzia  $D(X) = \binom{n}{2}p(1-p) =$

$= (1 - p)E(X)$ . Pravdepodobnosť, že vznikne najmenej jedna hrana môžeme ohraničiť strednou hodnotou (z Markovovej nerovnosti 2.2.2)

$$P(X \geq 1) \leq E(X) \leq \binom{n}{2} p(n) \leq \frac{n(n-1)}{2} p(n) \leq \frac{n^2}{2} p(n).$$

Ak  $\lim_{n \rightarrow \infty} \frac{p(n)}{\frac{1}{n^2}} \rightarrow 0$  tak  $n^2 p(n) \rightarrow 0$  Z toho vyplýva, že  $E(X) \rightarrow 0$  pre  $n \rightarrow \infty$ . V druhom prípade  $\left(\frac{p(n)}{\frac{1}{n^2}} \rightarrow \infty\right)$  využijeme Čebyševovu nerovnosť (2.2.3):

$$\begin{aligned} P(X \geq 1) &= 1 - P(X \leq 0) = 1 - P(-X \geq 0) = 1 - P(E(X) - X \geq E(X)) \geq \\ &\geq 1 - P(|E(X) - X| \geq E(X)) \geq 1 - \frac{D(X)}{[E(x)]^2} \geq 1 - \frac{(1-p)E(x)}{[E(x)]^2} \geq \\ &\geq 1 - \frac{(1-p)}{\frac{p(n)n^2}{2}} \geq 1 - \frac{2(1-p)}{p(n)n^2}. \end{aligned}$$

Ak  $\lim_{n \rightarrow \infty} \frac{p(n)}{\frac{1}{n^2}} \rightarrow \infty$  tak  $\frac{D(X)}{[E(x)]^2} \rightarrow 0$ . Z toho vyplýva, že  $P(X \geq 1) \rightarrow 1$  pre  $n \rightarrow \infty$ . □

To znamená, že ak  $t(n) = 1/n$  je prahová funkcia pre vznik aspoň jednej hrany, tak ak pravdepodobnosť  $p(n) \ll 1/n$  v limite bude pravdepodobné, že sieť nebude mať žiadne hrany. Naopak, ak  $p(n) \gg 1/n$ , pravdepodobnosť, že sieť má najmenej jednu hranu sa limitne bude blížiť k jednotke. □

## 2.2.2 Ďalšie prahové funkcie

**Veta 2.2.5.** [8]. *Nech  $A(n)$  je vlastnosť definovaná ako tvrdenie, že v náhodnej sieti  $G_{n,p}$  vznikne komponent väčší ako dva vrcholy, potom  $t(n) = n^{\frac{3}{2}}$  je prahová funkcia pre monotónnu vlastnosť  $A(n)$ .*

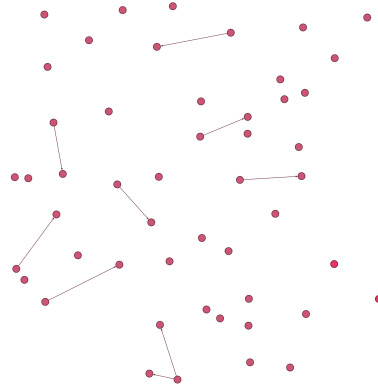
**Veta 2.2.6.** [8]. *Nech  $A(n)$  je vlastnosť definovaná ako tvrdenie, že v náhodnej sieti  $G_{n,p}$  začnú vznikáť cykly, potom  $t(n) = \frac{1}{n}$  je prahová funkcia pre monotónnu vlastnosť  $A(n)$ .*

**Veta 2.2.7.** *Nech  $A(n)$  je vlastnosť definovaná ako tvrdenie, že v náhodná sieť  $G_{n,p}$  je súvislá, potom  $t(n) = \frac{\ln(n)}{n}$  je prahová funkcia pre monotónnu vlastnosť  $A(n)$ .*

Vygenerujme náhodnú sieť s 50 vrcholmi a pravdepodobnosťou, že medzi každými dvoma vrcholmi vznikne hrana  $p = 0,01$ . Pre  $n = 50$  sa začnú objavovať spojenia

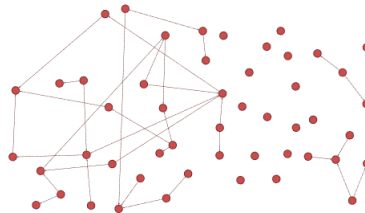
medzi vrcholmi s pravdepodobnosťou  $p = \frac{1}{n^2} = 0,0004$ . Obr. 2.3 ilustruje tvrdenie, že s malým  $p$  vzniká veľa izolovaných bodov a komponenty sú najviac z dvoch uzlov. Hranica, kedy vzniká komponent obsahujúci viac ako dva uzly je  $p = n^{-\frac{3}{2}} = 0,003$ . Ak

Obr. 2.3: Poissonova náhodná sieť  $n = 50, p = 0,01$ . (Zdroj: vlastné spracovanie)



zväčšíme pravdepodobnosť na hodnotu  $p = 0,03$  ( Obr. 2.4), vzniká viac spojení medzi uzlami, počet izolovaných bodov je menší a vznikajú cykly (je postupnosť hrán, ktoré vedú z jedného vrcholu a vrátia sa opäť do toho istého vrcholu). Hranica, kedy začínajú cykly vznikáť je  $p = \frac{1}{n} = 0,02$ . Na Obr. 2.5 je vygenerovaná sieť s  $p = 0,05$ . Všimnime

Obr. 2.4: Poissonova náhodná sieť  $n = 50, p = 0,03$ . (Zdroj: vlastné spracovanie)



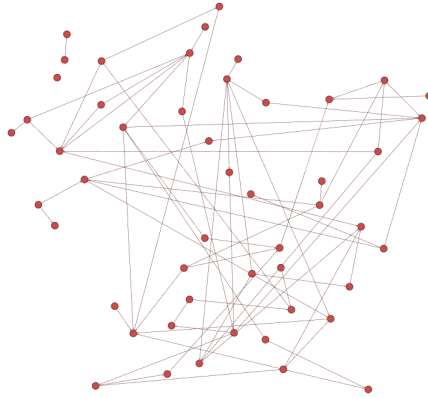
si, že počet komponentov sa znižuje, vzniká jeden veľký komponent. Graf generovaný s pravdepodobnosťou  $p = 0,1$  (Obr. 2.6) nemá žiadne izolované body. Hranica, kedy začína byť graf súvislý (vznik jedného komponentu) je  $p = \frac{\ln(n)}{n} = 0,08$ .

### 2.2.3 Súvislosť náhodných sietí

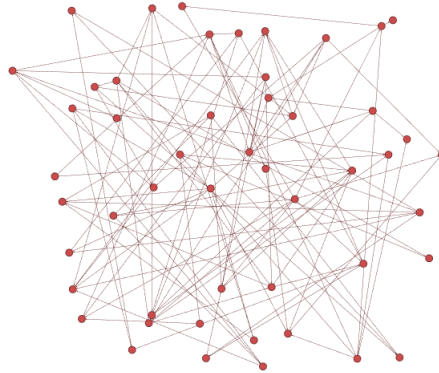
**Veta 2.2.8.** [15] Nech  $p(n) = \frac{c \ln(n)}{n}$ . Ak  $c \geq 3$  a  $n \geq 100$ , tak  $P_{n,p}(G \text{ je súvislý}) > 1 - \frac{1}{n}$ .

Predstavme si štát, ktorý má 2000 strategicky významných miest, pričom každé dve z nich sú spojené samostatnou železničnou traťou.

Obr. 2.5: Poissonova náhodná sieť  $n = 50, p = 0,05$ . (Zdroj: vlastné spracovanie)



Obr. 2.6: Poissonova náhodná sieť  $n = 50, p = 0,1$ . (Zdroj: vlastné spracovanie)



Štát má “potenciálneho protivníka”. V prípade konfliktu protivník zničí určitú železničnú trať spájajúcu dve miesta s pravdepodobnosťou  $q$ , nezávisle od ostatných. Pýtame sa, aké veľké môže byť  $q$ , ak si štát chce zachovať “infraštruktúru”, s dostatočnou veľkou pravdepodobnosťou? Infraštruktúra je práve súvislosť. Z jedného miesta sa vieme dostať do druhého.

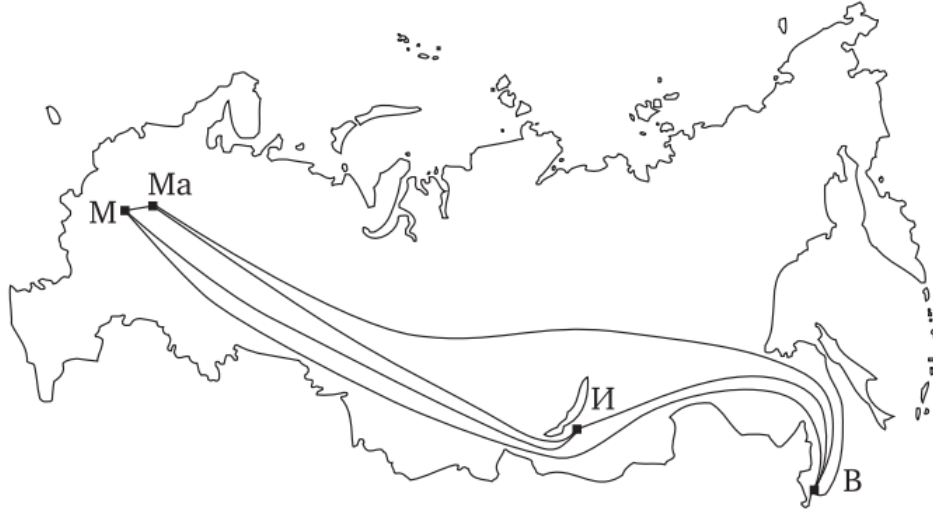
Podľa vety 2.2.8 ak  $p = \frac{3 \ln 2000}{2000} \doteq 0.01$  a teda  $q \doteq 0.99$ , tak pravdepodobnosť, že graf je súvislý, je viac ako  $1 - \frac{1}{2000} > 0.999$ . Teda protivníkovi “dovolíme” zničiť každú trať s pravdepodobnosťou 0.99, ale súvislosť sa zachová s pravdepodobnosťou viac ako 0,999. Čo je prekvapujúce.

*Dôkaz.* [15] Definujme náhodnú premennú  $X_n = X_n(G) = \begin{cases} 0 & \text{ak je } G \text{ súvislý,} \\ k & \text{ak má } k \text{ komponentov, pre } k > 1. \end{cases}$

Dokážme, že pravdepodobnosť  $P_{n,p}(X_n = 0) \rightarrow 1$  pre  $n \rightarrow \infty$ , čo je ekvivalentné tomu, že  $P_{n,p}(X_n \geq 1) \rightarrow 0$ . Podľa Markovovej nerovnosti (2.2.2)  $P_{n,p}(X_n \geq 1) \leq E(X_n)$ ,



Obr. 2.7: Infraštruktúra (súcisosť) železničnej siete štátu (Zdroj: [15])



takže stačí dokázať, že  $E(X_n) \rightarrow 0$ .

Potom  $X_n = X_{n,1} + \dots + X_{n,n-1}$ , kde  $X_{n,k}$  je počet komponentov tvorených práve z  $k$  vrcholov, ktoré má graf s  $n$  vrcholmi. Počet  $k$ -prvkových množín vrcholov je  $\binom{n}{k}$ , očísľujme ich  $K_1, \dots, K_{\binom{n}{k}}$ .

Ďalej,  $X_{n,k} = X_{n,k,1} + \dots + X_{n,k,\binom{n}{k}}$ , kde  $X_{n,k,i} = \begin{cases} 1 & \text{ak } K_i \text{ je komponent,} \\ 0 & \text{inak.} \end{cases}$

Strednú hodnotu  $X_n$  vieme napísať ako

$$E(X_n) = \sum_{k=1}^{n-1} \sum_{i=1}^{\binom{n}{k}} E(X_{n,k,i}),$$

kde  $E(X_{n,k,i}) = P_{n,p}(K_i \text{ je komponent gragu } G)$

$$\begin{aligned} &\leq P_{n,p}(\text{neexistujú hrany spájajúce vrcholy z } K_i \text{ s vrcholmi z } V \setminus K_i) \\ &= q^{k(n-k)}. \end{aligned}$$

Nerovnosť vyplýva z toho, že  $K_i$  sú  $k$ -prvkové množiny. To, že sú komponentom je silnejšia podmienka, teda pravepodobnosť je menšia.

Vráťme sa k strednej hodnote, a dosadíme:  $E(X_n) \leq \sum_{k=1}^{n-1} \sum_{i=1}^{\binom{n}{k}} q^{k(n-k)} = \sum_{k=1}^{n-1} \binom{n}{k} q^{k(n-k)}$ .

Táto suma je symetrická v tom zmysle, že členy pre indexy  $i, n - i$  sú rovnaké. Často budeme využívať, že  $\ln(1 + x) \leq x$  pre  $x > -1$ .

Z toho vyplýva, že  $\ln(1 - p) \leq -p$  a zároveň, pre  $c \geq 3$  platí, že  $p = \frac{c \ln n}{n}$ . (2.4)

Využitím rovnice (2.4) dostávame pre  $k = 1$ :

$$nq^{n-1} = n(1 - p)^{n-1} = ne^{(n-1)\ln(1-p)} \leq ne^{-p(n-1)} \leq e^{-\frac{3 \ln n(n-1)}{n}}.$$

Pre dosť veľké  $n$  ( $n \geq 10$ ) je  $\frac{n-1}{n} \geq 0.9$ , vtedy  $nq^{n-1} \leq e^{-2.7 \ln n} = \frac{1}{n^{2.7}}$ . Zoberme teraz podiel dvoch po sebe idúcich členov sumy:

$$\frac{\binom{n}{k+1} q^{(k+1)(n-k-1)}}{\binom{n}{k} q^{k(n-k)}} = \frac{n-k}{k+1} q^{-k+n-k-1}$$

Kvôli uvedenej symetrii členov nám stačí zaoberať sa členmi pre  $k \leq \frac{n}{2}$ . Rozlíšme dve možnosti:

1. Ak  $k \leq \frac{n}{8}$ , tak

$$\frac{n-k}{k+1} q^{-k+n-k-1} \leq (n-1) q^{\frac{3n}{4}-1} \leq (n-1) e^{\frac{-9 \ln n}{4} + p} \leq n e^{\frac{-9 \ln n}{4} + p} \leq n e^{\frac{-9 \ln n}{4} + 1}.$$

Prvá nerovnosť je jednoduchá:  $\frac{n-k}{k+1} q^{-k+n-k-1} \leq \frac{n-1}{k+1} q^{-k+n-k-1} \leq (n-1) q^{-k+n-k-1} \leq (n-1) q^{\frac{3n}{4}-1}$ , pretože výraz  $q^{-k+n-k-1}$  je najväčší pre  $k = \frac{n}{8}$ .

Druhú nerovnosť dostávame využitím rovnice(2.4):

$$(n-1) q^{\frac{3n}{4}-1} \leq (n-1) \left(1 - \frac{3 \ln n}{n}\right)^{\frac{3n}{4}-1} = (n-1) e^{\left(\frac{3n}{4}-1\right) \ln \left(1 - \frac{3 \ln n}{n}\right)} \leq (n-1) e^{\left(\frac{3n}{4}-1\right) \left(-\frac{3 \ln n}{n}\right)}.$$

Pre dosť veľké  $n$  máme:

$$n e^{\frac{-9 \ln n}{4} + 1} \leq \frac{n e^{\frac{\ln n}{4} - 1}}{e^{\frac{9 \ln n}{4} - 1}} = n e^{-2 \ln n} = \frac{1}{n}.$$

Podiel dvoch po sebe idúcich členov je  $\frac{1}{n} < 1$ , a teda,  $k$ -ty sčítanec je menší ako prvý. Vyriešme ešte, čo znamená “pre dosť veľké  $n$ ”. Pretože aby platila nerovnosť  $n e^{\frac{-9 \ln n}{4} + 1} \leq \frac{n e^{\frac{\ln n}{4} - 1}}{e^{\frac{9 \ln n}{4} - 1}}$ , musí byť výraz, ktorým prenasobujeme väčší ako jedna. To však je iba pre  $n \geq 55$ .

$$e^{\frac{\ln n}{4} - 1} \geq e^0$$

$$\frac{\ln n}{4} - 1 \geq 0$$

$$n \geq e^4$$

2. Ak  $k > \frac{n}{8}$ , tak

(a)  $\binom{n}{k} < 2^n$ , pretože množinu všetkých kombinácií  $k$ -tej triedy z  $n$  prvkov definujeme ako podmnožinu potenčnej množiny.

(b)  $q^{k(n-k)}$  keďže  $q \leq 1$ ,  $\frac{n}{8} < k < \frac{n}{2}$ ,  $(n-k) > \frac{n}{2}$  je  $q^{k(n-k)} \leq q^{\frac{n}{8} \frac{n}{2}} = q^{\frac{n^2}{16}} = e^{\frac{n^2}{16} \ln(1-p)} \leq e^{-\frac{pn^2}{16}} \leq e^{-\frac{3(\ln n)n}{16}}$ , kde posledné dve nerovnosti sme dostali opäť využitím 2.4.

Z (a) a (b) máme:  $\binom{n}{k} q^{k(n-k)} \leq 2^n n^{-\frac{3n}{16}}$ . Pre dosť veľké  $n$  platí  $2^n n^{-\frac{3n}{16}} \leq \frac{1}{n^{2.7}}$ , čo je naše ohraničenie pre prvý člen. A tak je prvý sčítanec v sume najväčší. Teda:

$$E(X_n) \leq \sum_{k=1}^{n-1} \frac{1}{n^{2.7}} < \frac{n}{n^{2.7}} = \frac{1}{n^{1.7}} \xrightarrow{n \rightarrow \infty} 0.$$

Nerovnosť  $2^n n^{-\frac{3n}{16}} \leq \frac{1}{n^{2.7}}$  platí pre  $n \geq 86$ , čo vidno na (Obr. 5.5), kde sme zobrazili funkciu  $f(n) = \frac{1}{n^{2.7}} - 2^n n^{-\frac{3n}{16}}$  pre  $n \in (0, 250)$ . Čo však potrebujeme dokázať je, že  $f(n) \geq 0$  pre všetky  $n \geq 86$ . Nerovnosť upravíme ekvivalentnými úpravami:

$$\begin{aligned} 2^n &\leq n^{\frac{3n}{16}} n^{-2.7} \\ n \ln 2 &\leq \frac{3n}{16} \ln n - 2.7 \ln n \\ 0 &\leq \frac{3}{16} \ln n - 2.7 \frac{\ln n}{n} - \ln 2 \end{aligned}$$

Označme pravú stranu rovnice  $g(n) = \frac{3}{16} \ln n - 2.7 \frac{\ln n}{n} - \ln 2$ . Platí,

$$\begin{aligned} g'(n) &= \frac{3}{16} \frac{1}{n} - 2.7 \left( \frac{\frac{1}{n} n - \ln n}{n^2} \right) \\ g'(n) &= \frac{3}{16} \frac{1}{n} - 2.7 \left( \frac{1 - \ln n}{n^2} \right) \\ g'(n) &= \frac{1}{n^2} \left( \frac{3}{16} n - 2.7 + 2.7 \ln n \right) \end{aligned}$$

Označme výraz  $\left( \frac{3}{16} n - 2.7 + 2.7 \ln n \right) = h(n)$

$$h'(n) = \frac{3}{16} + \frac{2.7}{n} > 0, \text{ čiže } h(n) \text{ je rastúca}$$

$h(2) \doteq -0,4535$ ;  $h(3) \doteq 0,8288 > 0$ , teda  $h(n) > 0$  pre  $n \geq 3$ . Pre  $n \geq 3$  je teda aj  $g'(n) > 0$ , čiže  $g(n)$  je rastúca. Keďže  $g(85) \doteq -0,00123696 < 0$ ;  $g(86) \doteq 0,0021972 > 0$ , tak  $g(n) \geq 0$  pre  $n \geq 86$ . Čo znamená, že pre  $n \geq 86$  platí nerovnosť.  $\square$

**Veta 2.2.9.** [15] Nech  $p(n) = \frac{c \ln(n)}{n}$ , potom:

$$\lim_{n \rightarrow \infty} Pr[\text{graf je súvislý}] = \begin{cases} \rightarrow 0 & \text{pre } c < 1, \\ \rightarrow 1 & \text{pre } c > 1. \end{cases}$$

*Dôkaz.* [15] Dokážeme tvrdenie pre  $c < 1$ . Dôkaz tvrdenia pre  $c > 1$  je podobný ako v predchádzajúcej vete, ale treba robiť presnejšie odhady. Dôkaz sa dá nájsť napríklad v [15]. Nech náhodná premenná  $\xi$  reprezentuje počet izolovaných vrcholov. Dokážme, že sieť nie je súvislá (pre  $c < 1$ ). Pravdepodobnosť, že sieť je súvislá potom bude  $P(\xi = 0) = 1 - P(\xi \geq 1)$ .

Využitím vety (2.2.2)  $P(\xi \geq 1) = E(\xi)$ . Z toho plynie, že nám stačí počítať strednú hodnotu:

$$E(\xi) = E(\xi_1 + \dots + \xi_n) = nE(\xi_i), \text{ kde } \xi_i(G) \text{ je indikátor } \xi_i(G) = \begin{cases} 1 & \text{Ak je vrchol izolovaný,} \\ 0 & \text{inak.} \end{cases}$$

Vrchol  $i$  je izolovaný, ak s ním nie je žiadny iný bod spojený hranou. (To znamená, neexistuje  $n - 1$  hrán). Pravdepodobnosť, že vznikne medzi nimi hrana je  $p$ . Strednú hodnotu  $E(\xi_i)$  potom vypočítame ako  $E(\xi_i) = (1 - p)^{n-1}$  a teda,  $E(\xi) = n(1 - p)^{n-1} = ne^{(n-1)\ln(1-p)}$ . Vieme, že  $\ln(1 + x) \leq x$  pre  $x > -1$ . Z toho vyplýva, že  $\ln(1 - p) \leq -p$ .

$$E(\xi) = ne^{(n-1)\ln(1-p)} \leq ne^{-(n-1)p} \leq ne^{-(n-1)c\frac{\ln n}{n}} = ne^{[\ln n]^{-c\frac{n-1}{n}}} = n^{1-c\frac{n-1}{n}}.$$

Pre  $c < 1$  je  $1 - c > 0$ . Pre všetky  $n > 0$ , platí:  $\frac{(n+1)-1}{n+1} > \frac{n-1}{n}$ , preto pre  $n > n_0$ , pre nejaké  $n_0$  ( $n_0 = 2$ ) platí:

$$\begin{aligned} \frac{n-1}{n} &> \frac{1}{2} \\ (1-c)\frac{n-1}{n} &> \frac{1-c}{2} \\ n^{(1-c)\frac{n-1}{2}} &> n^{\frac{1-c}{2}}. \end{aligned}$$

Keďže  $\lim_{n \rightarrow \infty} n^{\frac{1-c}{2}} \rightarrow \infty$ , aj  $E(\xi) \rightarrow \infty$ . Z (2.2.3) platí, že  $P(\xi \geq 1) \geq 1 - \frac{D(\xi)}{E^2(\xi)}$ . Počítajme disperziu:  $D(\xi) = E(\xi^2) - E^2(\xi)$  kde,

$$E(\xi^2) = E(\xi_1^2 + \dots + \xi_n^2 + \sum_{i \neq j} \xi_i \xi_j) = E(\xi_1 + \dots + \xi_n) + \sum_{i \neq j} E(\xi_i \xi_j).$$

Platí:  $P(\xi_i \xi_j = 1) = (1 - p)^{2n-3} \Rightarrow E(\xi_i \xi_j) = (1 - p)^{2n-3}$ , pretože  $\xi_i \xi_j$  má hodnotu 1 iba ak sú oba vrcholy izolované. V opačnom prípade je to 0. Aby boli oba vrcholy izolované, musíme vylúčiť existenciu  $2(n - 1) - 1 = 2n - 3$  hrán. (Pre každý z nich

$n - 1$ , avšak hranu, ktorá spája ich dva navzájom sme započítali dva krát, preto ju musíme odčítať).

$$E(\xi^2) = n(1-p)^{n-1} + \sum_{i \neq j} (1-p)^{2n-3} = n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3}$$

$$\begin{aligned} \frac{D(\xi)}{E^2(\xi)} &= \frac{E(\xi^2) - E^2(\xi)}{E^2(\xi)} = \frac{E(\xi^2)}{E^2(\xi)} - 1 \\ &= \frac{n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3}}{(n(1-p)^{n-1})^2} - 1 = \frac{1}{n(1-p)^{n-1}} + \frac{n-1}{n}(1-p)^{-1} - 1 \\ &= \frac{1}{E(\xi)} - \frac{1}{n(1-p)} + \frac{p}{1-p} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

□

**Veta 2.2.10.** [15] *Nech  $p(n) = \frac{c \ln(n)}{n}$ .*

1. *Ak  $c > 1$ , tak náhodný graf je skoro iste súvislý pre  $n \rightarrow \infty$ .*
2. *Ak  $c < 1$ , tak náhodný graf je skoro iste nesúvislý pre  $n \rightarrow \infty$ .*

# Kapitola 3

## Vlastnosti reálnych sociálnych sietí

### 3.1 Malý svet

Pojem “malý svet” sa vzťahuje na známu situáciu, keď pri stretnutí s neznámym človekom zistíme, že máme spoločných známych. Stanley Milgram, renomovaný profesor sociológie, opísal fenomén malého sveta známym experimentom, pri ktorom sa mala zásielka doručiť “náhodne zvolenému” obyvateľovi Bostonu, štartujúc od “náhodne zvoleného” obyvateľa zo vzdialeného mesta, pričom každý ju mohol poslať iba niekomu, koho osobne pozná. Skúmal sa počet takýchto odovzdávaní zásielky, kým sa dostala do cieľa. Z 296 zásielok sa do cieľa dostalo 29 % a takto nameraná “vzdialenosť” medzi dvoma náhodne vybranými obyvateľmi bola v priemere 5.7 medzičlánkov (“six degrees of separation”). [9]

Takýto druh výsledku dosiahlo mnoho ďalších podobných štúdií, ktoré tento fenomén analyzovali na oveľa väčších dátach. Napr. sociálna sieť Facebook skúmala túto skutočnosť na aktívnych užívateľoch v máji 2011. V tom čase mal Facebook približne 721 milión užívateľov. Priemerná vzdialenosť medzi náhodne vybranými dvoma užívateľmi tejto sociálnej siete bola 4.74, čo znamená, v priemere 3.74 medzičlánkov (“degrees of separation”). [2]

#### 3.1.1 Priemer náhodných sietí

Malý svet poukazuje na pomerne malý priemer vo veľkých sieťach. Preto sa táto podkapitola zaoberá vzťahom medzi počtom vrcholov v grafe a priemerom grafu. Pri tvorbe sme sa opierali o zdroje [3],[11].

**Veta 3.1.1.** *Nech je daný graf  $G$  s  $n$  vrcholmi, priemerom  $D$  a najväčším stupňom vrcholov  $\Delta$  ( $\Delta > 2$ ). Potom priemer grafu spĺňa:*

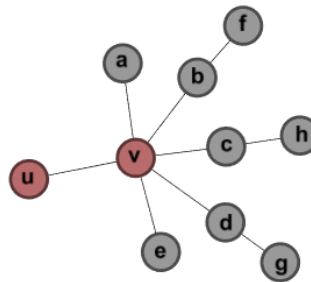
$$n \leq 1 + \Delta \frac{(\Delta - 1)^D - 1}{\Delta - 2}.$$

*Ekvivalentne*

$$D \geq \frac{\log(n [1 - \frac{2}{\Delta}] + \frac{2}{\Delta})}{\log(\Delta - 1)}.$$

Pre lepšie pochopenie, si predstavme takýto jednoduchý graf (Obr. 3.1):

Obr. 3.1: Sieť s 10 vrcholmi, priemerom 3 a najväčším stupňom 6.



Označme  $d_i(u)$  ako počet vrcholov vzdialených o  $i$  od  $u$ . Uvedomme si, že  $d_1(u) \leq \Delta$  pretože vrchol  $u$  môže mať najviac  $\Delta$  susedov. Vrchol  $v$  je vzdialený od  $u$  na vzdialenosť  $i = 1$ . Vidíme, že môže mať najviac  $\Delta - 1$  susedov vzdialených od  $u$  na vzdialenosť  $i + 1$ , keďže maximálny stupeň vrchola je  $\Delta$  a jedným zo susedných vrcholov je práve  $u$  (v tomto prípade sú to vrcholy  $a, b, c, d, e$ ).

Pre počet vrcholov vzdialených od  $u$  na vzdialenosť  $i = 2$  musí platiť  $d_2 \leq (\Delta - 1)d_1$ . Zoberme teraz vrchol, ktorý je od  $u$  vzdialený na vzdialenosť dva (v našom prípade napríklad vrchol  $b$ ). Existuje teda cesta dĺžky dva medzi vrcholmi  $u$  a  $b$ , v našom prípade  $u, v, b$ . Vrchol  $b$  môže mať maximálne  $\Delta$  susedov, jedným z nich je však ten, ktorý je na tejto najkratšej ceste, a teda jeho vzdialenosť od  $u$  je menšia ako vzdialenosť  $b$  od  $u$ . Počet susedov  $b$ , ktoré sú od  $u$  vzdialené na vzdialenosť  $i + 1$ , je teda maximálne  $\Delta - 1$ .

Pre počet vrcholov vzdialených od  $u$  na vzdialenosť  $i = 3$  musí platiť  $d_3 \leq (\Delta - 1)d_2$ . Zoberme teraz vrchol, ktorý je od  $u$  vzdialený na vzdialenosť tri (v našom prípade napríklad vrchol  $h$ ). Existuje teda cesta dĺžky tri medzi vrcholmi  $u$  a  $h$  (cesta  $u, v, c, h$ ). Vrchol  $h$  môže mať maximálne  $\Delta$  susedov, jedným z nich je však ten, ktorý je na tejto najkratšej ceste, a teda jeho vzdialenosť od  $u$  je dva (nie tri). Počet susedov  $h$ , ktoré sú od  $u$  vzdialené na vzdialenosť  $i + 1$ , je teda maximálne  $\Delta - 1$ . Preto platí  $d_3 \leq (\Delta - 1)d_2$ .

*Dôkaz.* Označme  $d_i(u) = |\Gamma_i(u)|$ , kde  $\Gamma_i := \{v : d_G(u, v) = i\}$  je množina vrcholov, ktoré sú vzdialené na vzdialenosť  $i$  od vrcholu  $u$ , pre všetky  $0 \leq i \leq D$ . Predpokladajme, že  $\Delta > 2$ . Ľahko vidieť, že ak maximálny stupeň vrchola je najviac  $\Delta$ , pre nejaký vrchol  $u$ ,  $d_1(u) \leq \Delta$  a pre všetky  $i \geq 2$  je  $d_i(u) \leq \Delta(\Delta - 1)^{i-1}$ . Ak priemer nie je väčší ako  $D$ , tak

$$n = 1 + d_1(u) + d_2(u) + \dots + d_D(u) = 1 + \Delta\{1 + (\Delta - 1) + \dots + (\Delta - 1)^{D-1}\}.$$

Súčtom geometrickej postupnosti dostávame

$$n \leq 1 + \Delta \frac{(\Delta - 1)^D - 1}{\Delta - 2}.$$

Z toho úpravami dostaneme vzťah pre priemer:

$$\begin{aligned} n(\Delta - 2) &\leq (\Delta - 2) + \Delta(\Delta - 1)^D - \Delta \\ n \left(1 - \frac{2}{\Delta}\right) - \left(1 - \frac{2}{\Delta}\right) + 1 &\leq (\Delta - 1)^D \\ \log \left[ n \left(1 - \frac{2}{\Delta}\right) + \frac{2}{\Delta} \right] &\leq D \log [\Delta - 1] \\ D &\geq \frac{\log \left( n \left[1 - \frac{2}{\Delta}\right] + \frac{2}{\Delta} \right)}{\log(\Delta - 1)}. \end{aligned}$$

□

Pre ilustráciu vety 3.1.1 sme na Obr.3.2 (obrázok zľava) zobrazili vzťah medzi priemerom a počtom vrcholov siete, pri pevnom najväčšom stupni vrcholov. Ten sme pevne určili  $\Delta = 30$ . Keďže priemer siete môže byť iba celé číslo, červená čiara na grafe je horná celá časť priemeru siete pri danom počte vrcholov. Obrázok teda hovorí o tom, ako rýchlo rastie priemer siete s rastúcim počtom vrcholov pre  $n \in (1, 1000)$ .

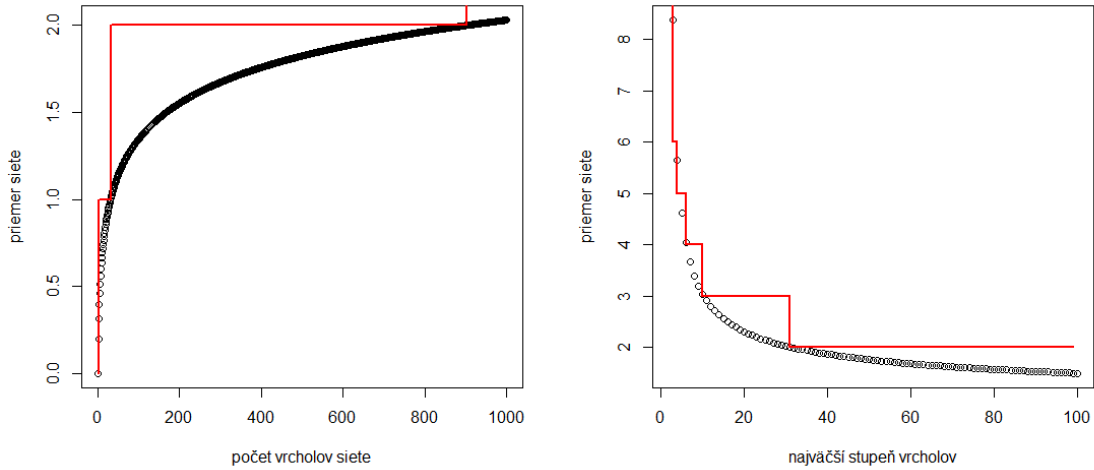
Na Obr.3.2 (obrázok sprava) sme znázornili vzťah medzi priemerom siete a najväčším stupňom vrcholov, pri pevnom počte vrcholov siete  $n = 1000$ . Obrázok teda hovorí, ako rýchlo rastie priemer siete s rastúcim najväčším stupňom vrcholov v sieti, ktorá má tisíc vrcholov.

## 3.2 Zhlukovanie

Je veľmi prirodzené predpokladať, že ak poznám Eriku a Nikolu, pravdepodobnosť toho, že sa aj ony dve poznajú navzájom je oveľa väčšia, než pravdepodobnosť v náhodných grafoch Erdősa a Rényiho. To nie je až také prekvapivé. Ľudia majú už od



Obr. 3.2: Vývoj priemeru siete s rastúcim počtom vrcholov (obrázok zľava) a s rastúcim najväčším stupňom vrcholov (obrázok sprava). (Zdroj:vlastné spracovanie)



pradávná tendenciu tvoriť spoločenstvá a zhluky. Avšak, zistilo sa, že zhlukovanie nie je len vlastnosť sociálnych sietí, ale napríklad aj sietí www-stránok a ďalších. Ukazuje sa, že je veľmi dôležité vedieť toto zhlukovanie hlbšie analyzovať. Preto, aby sme mohli zhlukovanie matematicky popísať, zaviedol sa pojem klasterizačný koeficient, alebo koeficient zhlukovania. Ten vlastne popisuje, aká hustá je sieť. Tiež sme si v tejto podkapitole definovali dve rozdielne miery zhlukovania: Priemerný klasterizačný a celkový klasterizačný koeficient.

**Definícia 3.1.** [8] Zaveďme označenie  $Cl_i(g)$ , čo je klasterizačný koeficient (koeficient zhlukovania) uzla  $i$ ,

$$Cl_i(g) = \frac{\#\{kj \in g | k, j \in N_i(g)\}}{\#\{kj | k, j \in N_i(g)\}}.$$

Potom priemerný klasterizačný koeficient je

$$Cl^{avg}(g) = \frac{\sum_{i=1}^n Cl_i(g)}{n}.$$

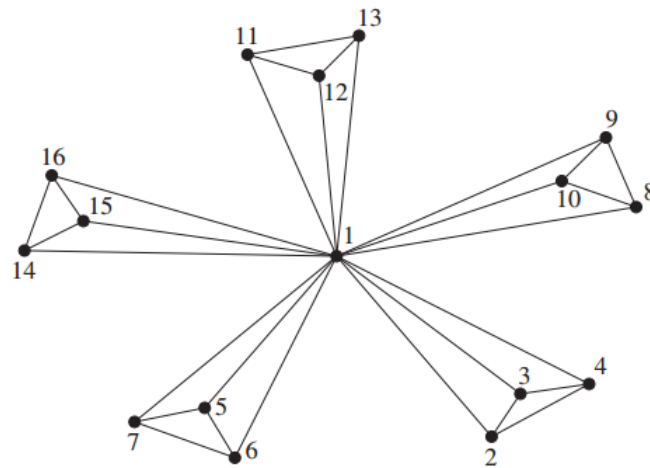
Celkový klasterizačný koeficient vypočítame ako

$$Cl(g) = \frac{\sum_{i=1}^n \#\{kj \in g | k, j \in N_i(g)\}}{\sum_{i=1}^n \#\{kj | k, j \in N_i(g)\}}.$$

Na Obr. 1.1 sú napríklad takéto klasterizačné koeficienty:  $Cl_{MEDICI}(g) = \frac{1}{6}$ , pretože okolie Mediciovcov sú rody Ridolfi, Tornabuon, Albizzi, Acciaiuol, Salviati, Barbadori a z nich sú navzájom spojené rodiny Ridolfiovcov a Tornabuoncov.  $Cl_{BISTERI}(g) = \frac{1}{3}$ ,

lebo okolie rodu Bisteri sú uzly Peruzzi, Strozzi a Guadagni, z nich sú prepojené väzbou rodiny Peruzzi a Strozzi. Klasterizačný koeficient uzla Guadagni je nula. Guadagni majú síce stupeň 4, ale tieto uzly nie sú prepojené väzbou, čiže  $Cl_{GUADAGNI}(g) = \frac{0}{4} = 0$ . Na Obr. 3.3 sme ilustrovali rozdielnosť mier klasterizácie. Poukazuje na to, ako veľmi sa priemerný a klasterizačný koeficient môže často líšiť. Na Obr. 3.3 môžeme vidieť sieť, ktorej priemerný klasterizačný koeficient je  $C^{AVG}(g) = 0,946$ , celkový klasterizačný koeficient je  $Cl(g) = 0,4$ . S rastúcim počtom takýchto zhlukov by sa priemerný koeficient blížil k jednotke, avšak celkový koeficient k nule. Metóda priemerného klasterizačného koeficientu dáva totiž väčšiu váhu vrcholom s malým stupňom uzlov.

Obr. 3.3: Rozdielne miery klasterizácie(Zdroj: [8])



Ako sme už spomenuli, ak poznám Eriku a Nikolu, pravdepodobnosť toho, že aj ony dve sa poznajú navzájom je oveľa väčšia, než pravdepodobnosť v náhodných grafoch Erdősa a Rényiho. Tam platí, že pravdepodobnosť, že sa aj ony dve poznajú, je  $p$  (keďže v tomto modeli vznikajú väzby nezávisle od seba). Táto pravdepodobnosť je pre všetky vrcholy rovnaká, preto celkový klasterizačný koeficient je tiež  $Cl(g) = p$ .

Teda, pre reálnu sieť s  $n$  vrcholmi, vieme vypočítať jej priemerný stupeň uzlov  $d$ . Ak predpokladáme, že náš graf vznikol takým istým spôsobom ako náhodné grafy, môžeme tiež predpokladať, že celkový klasterizačný koeficient bude  $Cl(g) = \frac{d}{n} = p$ . Dokázalo sa, že rozdiely medzi týmito reálnymi klasterizačnými koeficientami a koeficientami náhodných sietí sú v mnohých prípadoch veľmi veľké (viď. tabuľka 3.1). Z toho plynie, že zhlukovanie nie je len vlastnosť sociálnych, ale aj iných komplexných sietí (neutróny v mozgu, www-stránky...). [13], [8].

Tabuľka 3.1: Porovnanie klasterizačných koeficientov pre reálne a náhodné siete

Sieť	$n$	$d$	$Cl_{REAL}$	$Cl_{RAND}$	Referencia
Spolupráca matematikov	253339	3,9	0,15	0,00001	Newman(2001)
Spolupráca biológov	1520251	15,5	0,081	0,000010	Newman(2001)
Neutróny	282	14,0	0,28	0,049	Watts, Strogatz (1998)
WWW	153127	35,2	0,11	0,00023	Adamic(1999)

(Zdroj [13], vlastné spracovanie)

### 3.3 Homofília

V sieťach, kde vrcholy sú ľudia, často poprepájanie uzlov závisí od veľkého počtu iných charakteristík týchto uzlov. Je to preto, že ľudia majú svoju identitu, ktorá zahŕňa veci ako: vek, pohlavie, etnický pôvod, povolanie, úroveň vzdelania a iné. Inak povedané, ľudia majú tendenciu spájať sa podľa rovnakých názorov, záľub a tak ďalej.

Majme množinu vrcholov  $N$ , rozdelenú do skupín podľa nejakých vlastností na  $N_1, \dots, N_i, \dots, N_m$ . Všetky vrcholy v jednej skupine majú rovnaké vlastnosti a uzly z rôznych skupín majú rozdielne vlastnosti. Nech  $n$  a  $n_1, \dots, n_i$  je príslušná početnosť. Nech  $p(g)$  je počet väzieb v sieti  $g$  v porovnaní s tým, koľko by ich mohlo byť. Teda:

$$p(g) = \frac{\sum_{j \in N} d_j(g)}{n(n-1)},$$

kde  $d_j(g)$  je stupeň uzla  $j$ . Teraz urobme rovnaký výpočet, ale pozrime sa, aký je počet spojení medzi vrcholmi rovnakého typu v sieti  $g$ , v porovnaní s tým, aký by mohol byť. Túto veličinu označme  $p_s(g)$ . Je to teda:

$$p_s(g) = \frac{\sum_{i=1, \dots, m} \sum_{i, k \in N_i} g_{ik}}{\sum_{i=1, \dots, m} n_i(n_i - 1)}.$$

Homofíliu potom vyrátame

$$h(g) = \frac{p_s(g)}{p(g)}.$$

Na ilustráciu si zoberme príklad<sup>1</sup> o priateľstve medzi študentmi stredných škôl v súbore dát “Adolescent Health”, ktoré sú založené na rozhovore s viac ako deväťdesiat-tisíc študentmi, ktorí boli vybratí ako reprezentatívna zložka amerických stredných škôl

<sup>1</sup>Tento príklad pochádza z článku Golub and Jackson (2008)

vrátane miest i vidieka, z rôznych geografických regiónov a iných rôznych etnických a sociálno-ekonomických skupín študentov.

Tabuľka 3.2: Príklad homofílie študentov v amerike

	Rasa	Pohlavie	Ročník
Priemerná homofília škôl	1,4	1,2	4
Minimálna homofília škôl	0,9	1	1,5
Maximálna homofília škôl	2,7	1,5	5,6
Štandardná odchýlka homofília škôl	0,43	0,8	0,9

*(Zdroj [8], vlastné spracovanie)*

Z tabuľky 3.2 napríklad vidíme, že so štyrikrát väčšou pravdepodobnosťou sa zvyknú študenti kamarátiť so študentmi z ich ročníka, než s ostatnými študentmi, alebo s 1,4krát väčšou pravdepodobnosťou vzniknú priateľstvá v rámci jednej rasy, než medzi rasami a tak ďalej. V tejto časti sme čerpali z [8]. Viac o tejto vlastnosti sociálnych sietí sa môžete dozvedieť práve tam.

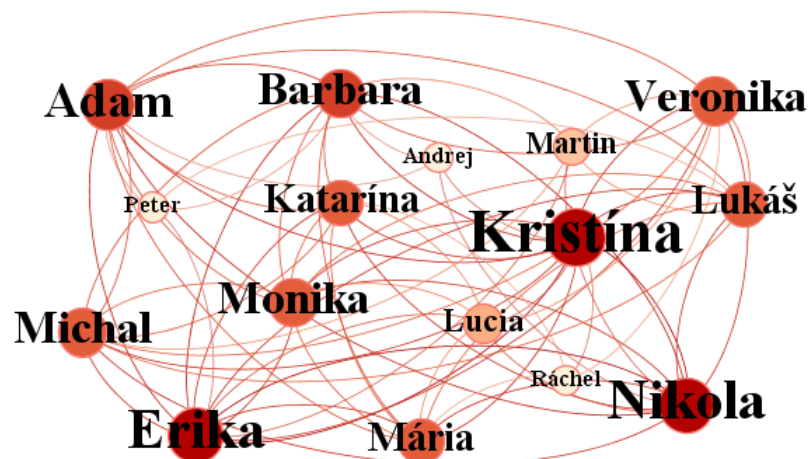
# Kapitola 4

## Analýza reálnych sociálnych sietí

### 4.1 Sieť poistných matematikov

Majme sieť tvorenú  $n = 16$  vrcholmi (študenti z krúžku poistná matematika). Hrana predstavuje priateľstvo na sociálnej sieti Facebook. Na tejto veľmi rozšírenej sociálnej sieti je možné spriatelíť sa s iným užívateľom tak, že ho o priateľstvo požiadate. Ak túto žiadosť potvrdí, stávate sa obaja navzájom priateľmi. Teda, sieť poistných matematikov na Facebooku je neorientovanou sieťou.

Obr. 4.1: Sociála sieť študentov krúžku poistná matematika (Zdroj: vlastné spracovanie)



V triede poistných matematikov nie je ani jeden človek, ktorý by na spomínanej sociálnej sieti nemal v priateľoch svojho spolužiaka (je súvislá). Tvorí ju iba jeden komponent.

V Tabuľke 5.1 môžeme vidieť rozdelenie stupňa vrcholov tejto siete. Najväčší stupeň

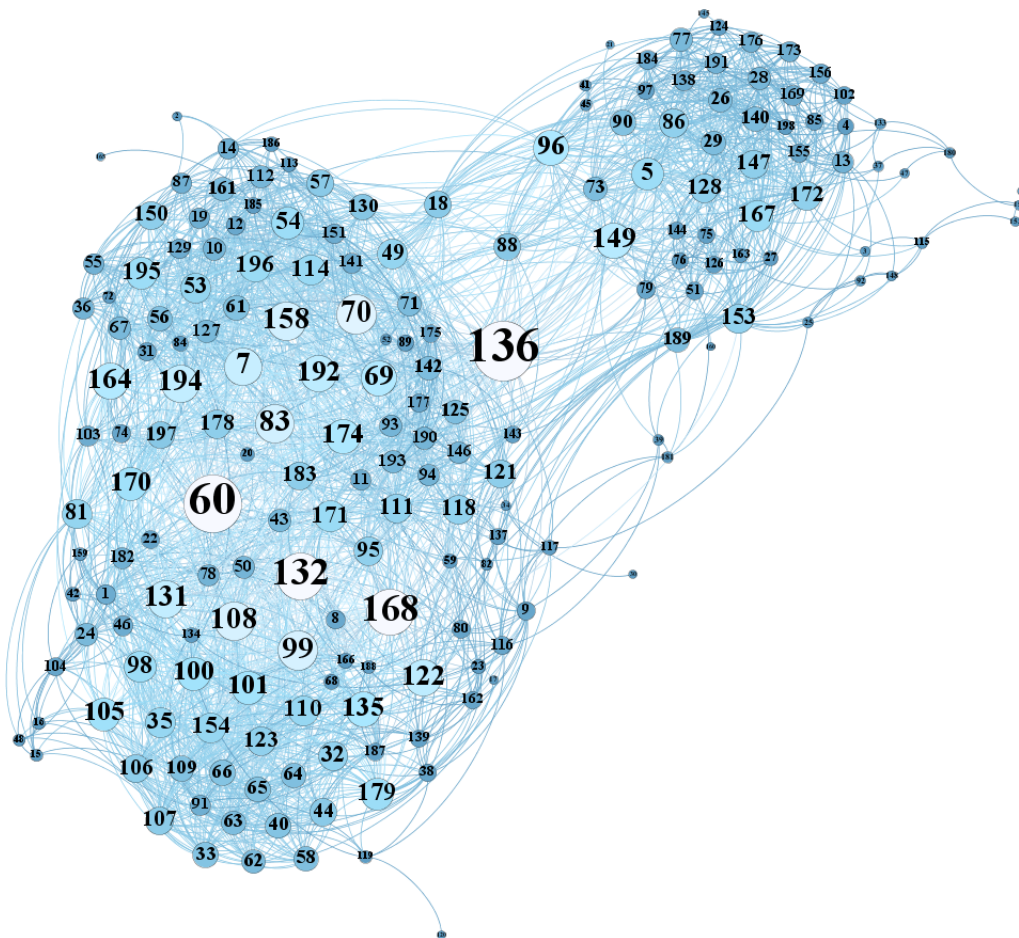
vrcholov je  $D = 13$ . Na základe vety 3.1.1 odhadujeme priemer siete  $D \geq 2$ . Z grafu vieme, že realne je táto vzdialenosť  $D = 3$  (vzdialenosť vrcholov “Peter” a “Ráchel” ). V tabuľke 5.1 môžeme tiež vidieť klasterizačné koeficienty všetkých študentov. Priemerný klasterizačný koeficient pre celú sieť je  $Cl^{AVG} = 0,756$  a celkový klasterizačný koeficient je  $Cl = 0,713$ . Ak predpokladáme, že náš graf vznikol takým istým spôsobom, ako náhodné grafy, celkový klasterizačný koeficient pre náhodnú sieť odhadneme takto:  $Cl_{RAND} = \frac{d}{n} = 0,544$ . Vidíme, že celkový klasterizačný koeficient reálnej siete je väčší a teda, táto sieť je pomerne dobre poprepájaná. To tiež vidno aj z toho, že priemerná dĺžka cesty je iba 1,392.

## 4.2 Sieť hudobníkov

Graf, ktorý spomínáme v tejto podkapitole, je graf sociálnej siete jazzových hudobníkov. Dáta spracovali Pablo Gleiser a Leon Danon v roku 2003. Údaje získali z digitálnej databázy “Red Hot Jazz Archív”. Pozreli sa na 198 jazzových kapiel, ktoré existovali v rokoch 1912-1940. Skupiny spojíme hranou, ak existuje aspoň jeden hráč, ktorý hral v oboch skupinách [1]. Dáta uvádzame v zdrojoch použitej literatúry [7].

Sieť je neorientovaná, súvislá. Pablo Gleiser a Leon Danon si na tejto sieti všimli, že zhlukovanie ovplyvňuje aj napr. rasové príslušnosť, či geografické rozpoloženie hudobných skupín. Môžeme to vidieť už z Obr. 4.2. Sieť je rozdelená na dva veľké zhluky. Toto rozdelenie do dvoch veľkých komunít môže byť interpretované ako prejav rasovej segregácie. Túto vlastnosť, že ľudia majú tendenciu spájať sa podľa toho, či majú nejakú spoločnú vlastnosť nazývame “homofília”. V tejto sieti sú znaky malého sveta, pretože priemerná vzdialenosť medzi dvoma náhodne vybranými vrcholmi je malá (priemerná dĺžka cesty je 2,23) a klasterizačný koeficient je veľký ( $Cl_{AVG} = 0,63$ ). Celkový klasterizačný koeficient, ktorý sme vypočítali z dát je  $Cl = 0,52$ . Teoretický klasterizačný koeficient, čiže taký, ktorý má náhodná sieť, ktorej pravdepodobnosť vzniku hrany medzi dvoma vrcholmi odhadneme ako podiel priemerného stupňa vrcholov a celkového počtu vrcholov je  $Cl_{RAND} = 0,1398$ .

Obr. 4.2: Sieť jazzových hudobníkov (Zdroj: vlastné spracovanie)



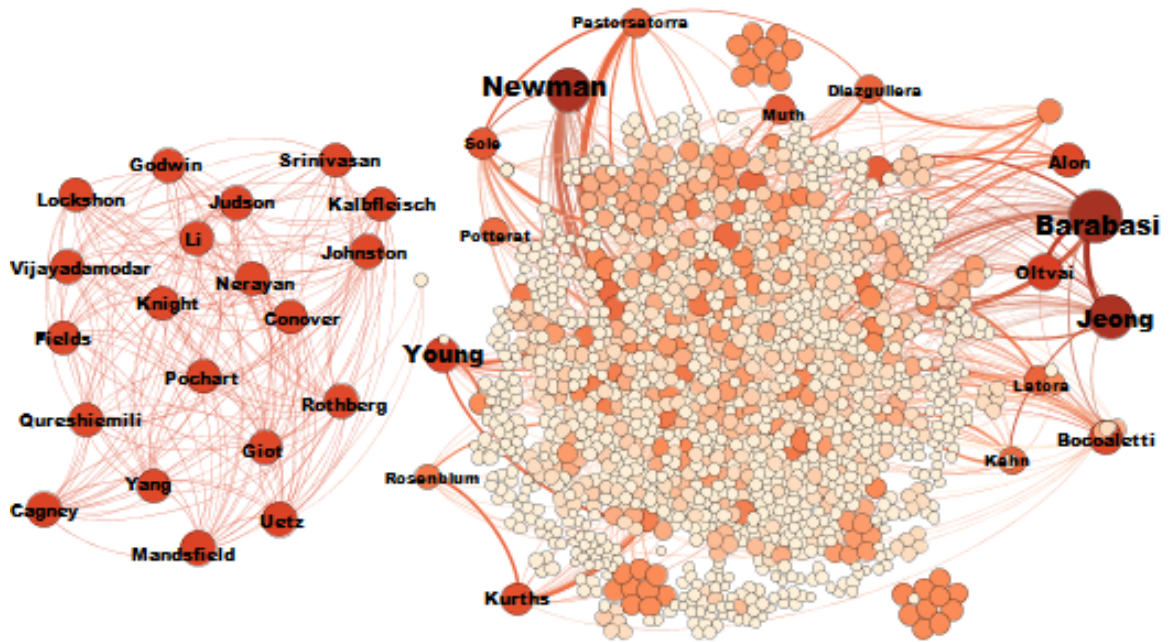
### 4.3 Sieť vedcov

V tejto podkapitole sme zobrazili a analyzovali sieť spolupráce vedcov, ktorí pracujú v oblasti teórie sietí a experimentov. Dáta zhromaždil M. Newman v máji roku 2006. Uvádzame ich v zdrojoch použitej literatúry [14].

Zobrazené sú len mená vedcov, ktorí majú najväčší stupeň, čiže sa najviac spolupodieľali na článkoch zaoberajúcimi sa analýzou sietí. Sieť má 1589 vrcholov. Nie je súvislá (tvorí ju až 396 komponentov). Keďže nie je súvislá, priemer siete by mal byť nula. Avšak, ako sme už spomínali, často sa zvykne uvádzať priemer pre najväčší komponent siete. Ten je 17.

Klasterizačný koeficient pre náhodnú sieť je  $Cl_{RAND} = 0,00217$ . Reálna sieť má celkový klasterizačný koeficient  $Cl = 0,693$  a teda, miera zhlukovania je veľká. Priemerný

Obr. 4.3: Sieť vedcov (Zdroj: vlastné spracovanie)



klasterizačný koeficient je  $Cl^{AVG} = 0,878$ . Vidíme, že celkový klasterizačný koeficient reálnej siete je rádovo väčší než klasterizačný koeficient náhodnej siete. Dôvod sme vysvetľovali v predošlej kapitole. Náhodné grafy nereprezentujú dobre zhhlukovanie.

Priemerný stupeň uzlov je 3,451. Zaujímavé je, že pri takomto malom priemernom stupni vrcholov, je priemerná dĺžka cesty 5,823. Sieť má znaky malého sveta, pretože priemerná dĺžka cesty je malá a klasterizácia veľká.



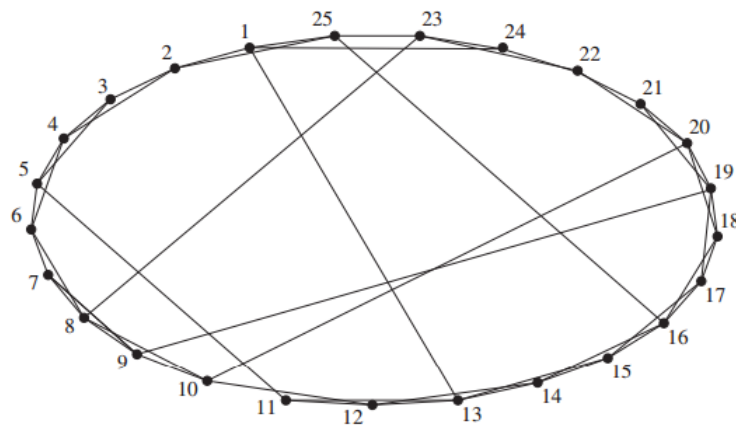
# Kapitola 5

## Ďalšie modely pre sociálne siete

### 5.1 Model malého sveta

Z uvedených dôvodov je zrejmé, že model Erdősa a Rényiho nevystihuje dobre zhlukovanie. Riešenie tohto problému prišlo v roku 1998, kedy Duncan Watts a Steven Strogatz publikovali nový model sietí. Vrcholy sú usporiadané do kružnice tak, že každý uzol je spojený so susedom, ktorý je od neho vzdialený  $k$  vrcholov (napríklad môže byť spojený so svojimi najbližšími dvoma susedmi a susedmi, ktoré su “ob jeden”). Tento model lepšie opisoval zhlukovanie, avšak stratil rysy malého sveta. Ak by sme napríklad chceli prejsť z jednej strany kružnice na druhú, museli by sme ísť cez veľa bodov a teda priemer siete by bol priveľký. [8]

Obr. 5.1: Wattsonov a Strogatzov model (Zdroj: [8])



Preto tento model Watts a Strogatz vylepšili. Náhodne vybrali dva uzly a zrušili väzbu medzi nimi (napríklad vrchol  $i$  a  $j$ ). Potom tento uzol náhodne spojili s iným

uzlom, ktorý nie je jeho susedným uzlom (čiže vrchol  $i$  by sme spojili s nejakým vrcholom  $k$ ). [8] Tieto náhodné väzby fungovali ako mosty a tým sa priemer siete rapídne zmenšil. Samozrejme, že ďalším takýmto odpájaním a pripájaním väzieb by zhlukovanie nakoniec zmizlo. Zistilo sa však, že stačí pomerne málo takýchto väzieb na to, aby sa priemer siete dostatočne zmenšil.

Model Wattsona a Strogatza je zaujímavý z hľadiska zhlukovania. Po vhodnom vzniknutí náhodných hrán zachováva aj malý priemer siete, avšak má aj nedostatky. Tými je najmä príliš pravidelná distribúciu stupňa uzlov, ktorá nie je špecifická pre sociálne siete. [8]

## 5.2 Mocninový model

V tejto časti sa opierame hlavne o [5] a [4], kde môže čitateľ nájsť viac informácií o tomto modeli. Skúmaním reálnych sietí sa ukázalo, že niektoré uzly siete majú až niekoľkonásobne väčší stupeň vrcholov než iné. [5] Taketo vrcholy označujeme centrá. Vznik centier je následkom iného uvažovania vzniku sietí. Na rozdiel od predchádzajúcich modelov, v tomto modeli sú vrcholy pridávané do siete postupne. Ďalším pravidlom pre pripájanie uzlov do siete je preferenčné pripájanie. Čo znamená, že uzly sa do siete nepripájajú rovnomerne náhodne, ale podľa nejakého kritéria. Nový uzol sa s väčšou pravdepodobnosťou pripojí k uzlu s väčším stupňom. Tak nám vznikne jedno veľké centrum, niekoľko menších, ešte menších a tak ďalej. Siete s takýmto charakterom voláme bezškálové siete. Bezškálová topológia je prirodzeným dôsledkom toho, že sieť svojou podstatou stále expanduje. [5] Tieto typy sietí popisujeme pomocou mocninového rozdelenia

$$P(d) = cd^{-\gamma},$$

kde  $P(d)$  je pravdepodobnosť, že vrchol má stupeň  $d$ ,  $c$  je normalizačná konštanta a  $\gamma$  je parameter. To môžeme prepísať ako:

$$\log(P(d)) = \log(c) - \gamma \log(d).$$

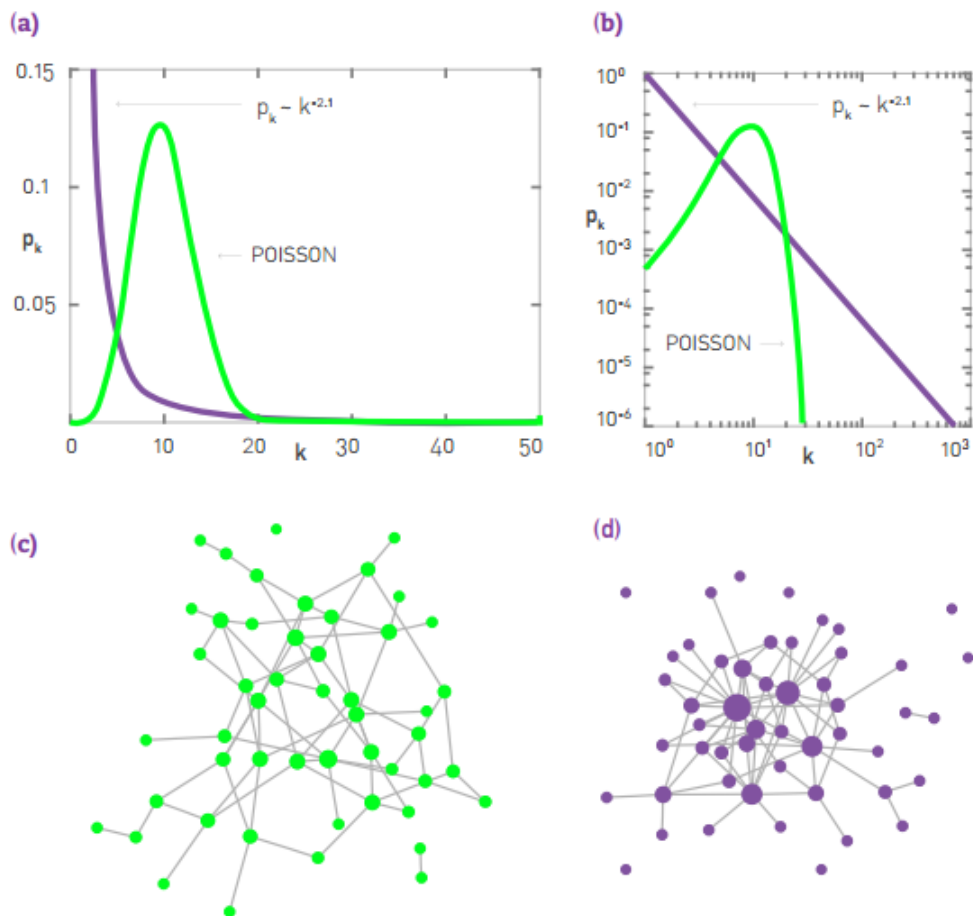
Siete riadiace sa mocninovým zákonom majú niektoré spoločné vlastnosti. Napríklad, že majú “ťažké chvosty” vzhľadom na Poissonovo rozdelenie. To znamená, že frekvencia veľmi vysokých a nízkych stupňov je väčšia, než v prípade, kde väzby vznikali nezávisle s rovnakou pravdepodobnosťou. [8], [5]

Kvôli centráam majú pomerne malý polomer a sú odolné voči náhodným výpadkom. Čiže, ak náhodne odoberieme niektorý uzol, sieť sa nerozpadne na komponenty. Avšak, keby sme odstraňovali uzly cielene, stačilo by odstrániť malý počet uzlov na to, aby sa sieť úplne rozpadla. V tomto prípade by sme sa zamerali práve na centrá. [5]

Na Obr.5.2 môžeme vidieť porovnanie Poissonovho a mocninového rozdelenia. Jednotlivé grafy predstavujú:

- (a) Porovnanie Poissonovej funkcie a Moninovej funkcie s parametrom  $\gamma = 2, 1$ , obidve funkcie majú priemerný stupeň jedenásť.
- (b) To isté čo v a), ale je zobrazený log-log graf.
- (c) Náhodne generovaná sieť s  $n=50$  vrcholmi a priemerným stupňom tri.
- (d) Bezškálová sieť s parametrom  $\gamma = 2, 1$ , a priemerným stupňom uzla tri.

Obr. 5.2: Poissonovo rozdelenie vs. Mocninové rozdelenie, (Zdroj: [4])



# Záver

Témou práce bol fenomén malého sveta v sociálnych sieťach. Ten sme modelovali pomocou modelu Erdősa a Rényiho, ktorí položili základy teórie náhodných grafov. Napriek tomu, že ľudská spoločnosť, internet, bunka či mozog môžu byť dobre reprezentované pomocou grafu, rozhodne sa navzájom líšia tým, že vznikajú s veľmi odlišnými pravidlami. Tieto problémy vyriešili Erdős a Rényi tak, že zanedbali rozdiely vzniku siete a začali uzly spájať náhodne.

V prvej kapitole sme sa oboznámili so základnými pojmami, ktoré sme v práci využívali, sociálnymi sieťami a náhodnými grafmi. Ukázalo sa, že takéto grafy sa stali akýmsi štandardom, s ktorými sa oplatí reálne siete porovnávať za účelom zistiť, či boli hrany v sieti vytvorené náhodne, alebo vznikajú s nejakým pravidlom.

Modelom Erdősa a Rényiho a jeho vlastnosťami sme sa zaoberali v druhej kapitole. Tieto vlastnosti sme ilustrovali na príkladoch simulovaných sietí.

Tretia kapitola sa zaoberala vlastnosťami reálnych sociálnych sietí. Ukázali sme ako priemer siete závisí od počtu vrcholov siete, definovali sme si dve rozličné miery zhlukovania a vlastnosť homofílie, ktorú sme si ukázali na príklade študentov amerických stredných škôl. Je prirodzené predpokladať, že ak mám dvoch kamarátov, pravdepodobnosť toho, že sa aj oni dvaja poznajú je väčšia, než v náhodných grafoch Erdősa a Rényiho, kde takúto skutočnosť neberú do úvahy. Hrany tam vznikajú náhodne s rovnakou pravdepodobnosťou a nezávisle od seba. Inak povedané, graf Erdősa a Rényiho nepopisuje dobre zhlukovanie.

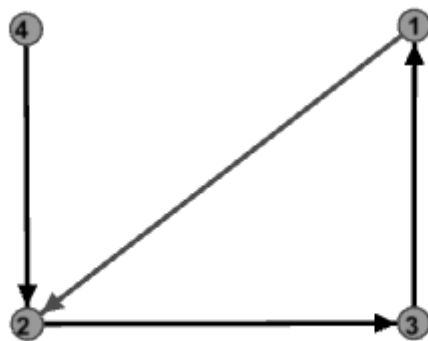
V nasledujúcej, štvrtej časti sú analyzované a zobrazené reálne sociálne siete. Zobrazili sme sieť spolupráce vedcov, sieť jazzových hudobníkov a sieť poisťných matematikov.

Obsahom záverečnej, piatej časti sú ďalšie modely a ich vlastnosti, ktorými môžeme modelovať malý svet. Uviedli sme model malého sveta Watta a Strogatza, ktorý veľmi

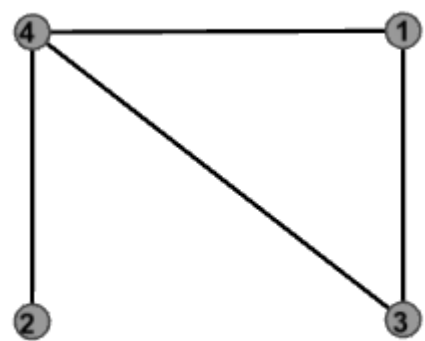
dobře popisuje malý svět, avšak distribúcia stupňa vrcholov je oproti reálnym sieťam príliš pravidelná. Tiež sme uviedli mocninový model, v ktorom sieť vzniká úplne iným spôsobom než v predošlých modeloch. Uvažuje bežškálovosť rastúcich sietí.

# Prílohy

## Príloha A



Obr. 5.3: Príklad orientovanej siete

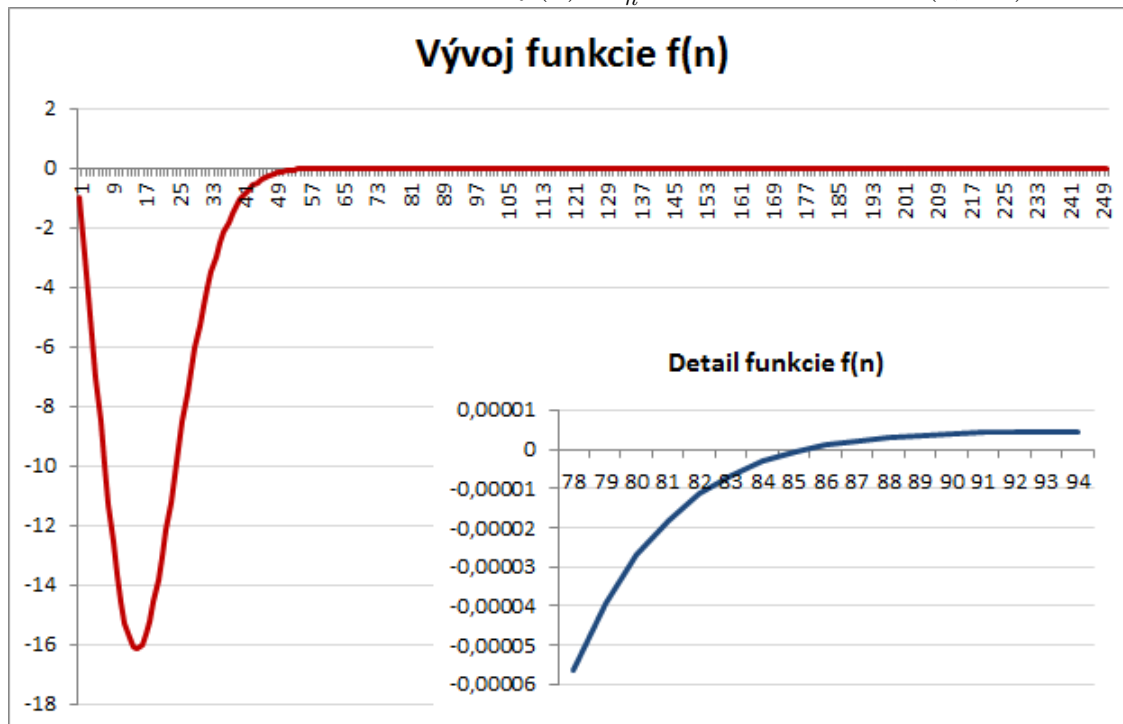


Obr. 5.4: Príklad neorientovanej siete

## Príloha B

Zobrazenie funkcie, ktorú využívame v dôkaze vety 2.2.9.

Obr. 5.5: Zobrazenie funkcie  $f(n) = \frac{1}{n^{2.7}} - 2^n n^{-\frac{3n}{16}}$  pre  $n \in (0, 250)$



## Príloha C

Tabuľka 5.1: Tabuľka charakteristík vrcholov (študentov poistnej matematiky)

---

<b>Meno</b>	<b>Stupeň</b>	<b>Klasterizačný koeficient</b>
Kristína	13	0.653
Adam	12	0.681
Monika	10	0.844
Katarína	10	0.755
Nikola	13	0.679
Erika	13	0.679
Lukáš	10	0.688
Michal	10	0.822
Veronika	10	0.644
Barbara	12	0.621
Martin	6	0.933
Ráchel	4	0.666
Peter	4	0.833
Andrej	4	1.0
Lucia	7	0.857
Mária	10	0.733
Priemerný klasterizačný koeficient	0,756	
Celkový klasterizačný koeficient	0,544	

---



# Literatúra

- [1] AMAIRANI, G., ABBEY, B., 2014. Network of Shared Jazz Musicians. [cit. 2016-04-20]. Dostupné na adrese:  
[http://humnet.scripts.mit.edu/wordpress2/wp-content/uploads/2011/09/1041\\_JazzBandsII.pdf](http://humnet.scripts.mit.edu/wordpress2/wp-content/uploads/2011/09/1041_JazzBandsII.pdf)
- [2] BACKSTROM, L., BOLDI, P., ROSA, M., UQANDER, J., VIGMA, S. 2011. Four degrees of separation. arxiv:1111.4570v1,2011
- [3] BALLOBÁS, B. 1985. Random Graphs. London: Academic Press.
- [4] BARABÁSI, A. L. 2014. Network science. [cit. 2016-05-12]. Dostupné na adrese:  
<http://barabasi.com/networksciencebook/>
- [5] BARABÁSI, A. L. 2005. V pavučině sítí. Vyd. 1. V Praze: Paseka. ISBN 80-718-5751-3.
- [6] DUNCAN, J. W. 2003. Six degrees: The Science of a Connected Age. W. W. Norton and Company. ISBN 0-393-04142-5.
- [7] GLEISER, P., DANON, L., 2003. Adv. Complex Syst.6, 565. Dostupné na adrese:  
<https://github.com/gephi/gephi/wiki/Datasets>
- [8] JACKSON, M. O. 2010. Social and Economic Networks. Princeton University Press. ISBN 978-0691148205.
- [9] LEVENE, M. 1957. An introduction to search engines and web navigation , second edition. ISBN 978-0-470-52684-2.
- [10] MARKECHOVÁ, D., STEHLÍKOVÁ, B., TIRPÁKOVÁ, A., 2011. Štatistické metódy a ich aplikácie, Nitra. Dostupné na adrese:  
[http://www.km.fpv.ukf.sk/upload\\_publicacie/20120125\\_143707\\_\\_1.pdf](http://www.km.fpv.ukf.sk/upload_publicacie/20120125_143707__1.pdf)

- [11] MILLER, M., ŠIRÁŇ J. "Moore graphs and beyond: a survey of the degree/diameter problem." *The Electronic Journal of Combinatorics* [electronic only] DS14 (2005): 61 p., electronic only-61 p., electronic only. <http://eudml.org/doc/125462>
- [12] NÁTHER, A. Prednáška z teórie pravdepodobnosti.
- [13] NEWMAN, M. E. J. *Random Graphs as Models of Networks*. Santa Fe Institute. 1399 Hyde Park Road. Santa Fe. NM 87501, U.S.A.
- [14] NEWMAN, M. E. J. 1985. *Phys. Rev. E* 74, 036104x. [cit. 2016-04-20]. Dostupné na adrese:  
<https://github.com/gephi/gephi/wiki/Datasets>
- [15] Райгородский, А. М. 2011. Модели случайных графов. Издательство МЦНМО. Москва . Dostupné na adrese:  
<http://www.mccme.ru/free-books/dubna/raigor-4.pdf>
- [16] RÉNYI, A. 1972. *Teorie pravděpodobnosti*. Academia, Praha.
- [17] RICE, J. A. 1995. *Mathematical Statistics and Data Analysis*, second edition. Belmont, CA: Duxbury Press.
- [18] Články Erdősa a Rényiho. [cit. 2016-04-20]. Dostupné na adrese:  
[https://renyi.hu/~p\\_erdos/Erdos.html](https://renyi.hu/~p_erdos/Erdos.html)