

**UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY**

**HĽADANIE GRAFLETOV
VO FUNKČNÝCH SIEŤACH MOZGU**

Diplomová práca

**UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY**

**HĽADANIE GRAFLETOV
VO FUNKČNÝCH SIEŤACH MOZGU**

Diplomová práca

| | |
|----------------------|----------------------------------|
| Študijný program: | Aplikovaná informatika |
| Študijný odbor: | 2511 Aplikovaná informatika |
| Školiace pracovisko: | Katedra aplikovanej informatiky |
| Školiteľ: | doc. RNDr. Mária Markošová, PhD. |

Bratislava, 2016

Bc. Lucia Budinská



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Lucia Budinská
Študijný program: aplikovaná informatika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.2.9. aplikovaná informatika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Hľadanie grafletov vo funkčných sieťach mozgu
Graphlets in the functional brain networks

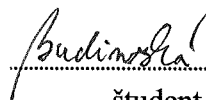
Cieľ: Cieľom práce je analyzovať motívy v reálnych komplexných sieťach, funkčných sieťach mozgu

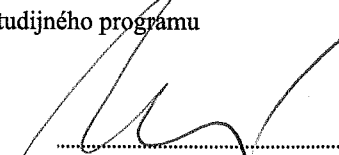
Anotácia: Študent zvládne dostupný softvér na hľadanie grafletov. Doprogramuje do neho potrebné funkcie a zanalyzuje súbor reálnych sietí nameraných pre mladých a starých zdravých participantov. Porovná štatistické vlastnosti a graflety vyskytujúce sa v daných sieťach s podobnými vlastnosťami v sieťach nameraných pre ľudí s Alzheimerovou chorobou.

Vedúci: doc. RNDr. Mária Markošová, PhD.
Katedra: FMFLKAI - Katedra aplikovanej informatiky
Vedúci katedry: doc. PhDr. Ján Rybár, PhD.
Dátum zadania: 10.10.2014

Dátum schválenia: 24.10.2014

prof. RNDr. Roman Ďurikovič, PhD.
garant študijného programu


.....
študent


.....
vedúci práce

Čestné vyhlásenie

Čestne vyhlasujem, že som diplomovú prácu „Hľadanie grafletov vo funkčných sieťach mozgu" vypracovala samostatne s použitím uvedených zdrojov.

V Bratislave dňa 4.5.2016

Bc. Lucia Budinská

Pod'akovanie

Rada by som sa pod'akovala svojej školiteľke, doc. RNDr. Márii Markošovej, PhD., za čas a úsilie, ktoré tejto práci venovala, za cenné rady, odporúčania a pomoc, ktoré mi poskytla, a v neposlednom rade za všetky vedomosti, ktoré som od nej získala.

Abstrakt

Diplomová práca sa zaoberá skúmaním funkčných sietí mozgu z hľadiska ich grafletovej štruktúry. V sieťach hľadáme graflety s tromi a štyrmi vrcholmi, a distribúcie jednotlivých orbít trojvrcholových grafletov a štvorkliky. Skúmame funkčné siete mozgu mladých zdravých participantov do 25 rokov, zdravých participantov nad 65 rokov a participantov nad 65 rokov s diagnostikovanou Alzheimerovou chorobou. Siete porovnávame v rámci skupín, medzi skupinami a so sieťami vytvorenými podľa Barabási-Albert a Watts-Strogatz modelov. Na porovnanie používame relatívnu vzdialenosť frekvencií grafletov a distribúciu stupňov grafletov. Prvá metóda ukazuje väčšie rozdiely medzi funkčnými sieťami mozgu mladých participantov a participantov s Alzheimerovou chorobou, a zároveň sme pomocou nej zistili podobnosť s WS modelom. Druhá metóda neodhalila väčšie rozdiely medzi jednotlivými skupinami participantov, ale ukázala väčšiu podobnosť funkčných sietí mozgu s BA modelom.

Kľúčové slová: funkčné siete mozgu, graflety, distribúcia stupňov grafletov, relatívna vzdialenosť frekvencií grafletov, Wattz-Strogatz model, Barabási-Albert model

Abstract

Research in our master thesis is aimed at graphlet structure of functional brain networks. We are searching for graphlets with three and four vertices, and orbits distribution of 3-vertex graphlet and 4-clique in networks. We examine functional brain networks of young healthy participants under 25, healthy participants who are over 65 years old, and participants with diagnosed Alzheimer's disease over 65. We compare networks in groups, in between groups and also with networks based on Barabasi-Albert model or Wattz-Strogatz model. Relative graphlet frequency distance and graphlet degree distribution are used for networks' comparisons. First method shows greater differences between functional brain networks of young participants and participants with Alzheimer's disease and also this method discovered significant resemblance with WS model. Second method didn't show any significant differences in between groups, but it showed some resemblance of functional brain networks with BA model.

Keywords: functional brain networks, graphlets, graphlet degree distribution, relative graphlet frequency count, Wattz-Strogatz model, Barabasi-Albert model

Obsah

| | |
|---|----|
| Úvod..... | 1 |
| 1. Teoretické východiská | 3 |
| 1.1. Základy teórie grafov | 3 |
| 1.2. Modely sietí | 9 |
| 1.2.1. Náhodné siete | 10 |
| 1.2.2. Siete malého sveta | 11 |
| 1.2.3. Siete s preferenčným pripájaním..... | 13 |
| 1.3. Graflety | 14 |
| 1.3.1. Počet grafletov | 15 |
| 1.3.2. Vektor distribúcie grafletov | 16 |
| 1.3.3. Distribúcia stupňov grafletov | 18 |
| 1.4. Funkčné siete mozgu | 20 |
| 2. Ciele práce a dostupné dáta..... | 23 |
| 2.1. Ciele práce | 23 |
| 2.2. Reálne dáta..... | 23 |
| 3. Implementácia výskumných metód..... | 26 |
| 3.1. Hľadanie grafletov | 26 |
| 3.1.1. Grafletové vzory..... | 27 |
| 3.1.2. Počet grafletov | 31 |
| 3.1.3. Distribúcia orbít grafletov | 33 |
| 3.2. Analýza sietí | 34 |
| 3.3. Umelo vytvorené dáta..... | 36 |
| 4. Výsledky | 38 |
| 4.1. Počty grafletov vo funkčných sieťach mozgu | 38 |
| 4.2. Relatívna vzdialenosť frekvencií grafletov..... | 40 |
| 4.3. Distribúcia stupňov grafletov..... | 45 |
| Záver | 48 |
| Literatúra a internetové zdroje | 50 |
| Prílohy..... | 52 |

Úvod

Ľudský mozog ešte stále nie je úplne prebádanou oblasťou. Ľudia sa mnoho rokov snažia prísť na to, ako funguje, ako sa správa a v modernej dobe sa snažia napodobniť jeho činnosť pomocou rôznych počítačových algoritmov. Matematici a informatici skúmajú mozog aj pomocou funkčných sietí mozgu. Zjednodušene by sa dalo povedať, že sa jedná o siete tvorené aktívnymi časťami mozgu, ktoré sa aktivizujú pri nejakej činnosti. Tieto siete sa získavajú pomocou funkčnej magnetickej rezonancie (fMRI).

Vedci sa snažia nájsť model, ktorý by presne popísal dynamiku rôznych funkčných sietí. Existujú štúdie, v ktorých sa autori zaoberali funkčnými sieťami mozgu vo vzťahu k ľudskej inteligencii [1], či skúmaniu schizofrénie [2]. Zaujímavý bol tiež výskum vývinu mozgu u detí a mladých dospelých [3], alebo naopak výskum starnutia [4].

My sme mali k dispozícii funkčné siete mozgu mladých ľudí do 25 rokov a ľudí nad 65 rokov a to jednak zdravých, alebo s diagnostikovanou Alzheimerovou chorobou. Pri tejto chorobe dochádza k postupnému zániku nervových buniek, čo vedie k degenerácii mozgu. Na splnení akejkoľvek úlohy, či sa jedná o reč, chôdzu, alebo len jednoduché poklepanie prstom po podložke, spolupracuje v mozgu viac centier. V každom z nich sa nachádzajú nervové bunky, ktoré danú činnosť riadia. Pacienti s Alzheimerovou chorobou majú problémy s pamäťou, chôdzou, orientáciou a podobne, čo priamo súvisí s nefunkčnosťou niektorých centier mozgu alebo ich nekoordináciou. Činnosť mozgu týchto ľudí sa preto odlišuje od činnosti mozgu zdravých ľudí. Tento fakt by mali reflektovať aj funkčné siete mozgu.

Existuje viacero prístupov k skúmaniu funkčných sietí, v našej práci sme sa rozhodli tieto siete skúmať z hľadiska ich grafletovej štruktúry. Graflety sú malé indukované navzájom neizomorfné spojené grafy veľkej siete [5]. Pre lepšiu ilustráciu si môžeme priblížiť graflety s tromi vrcholmi. Sú dva – jeden má vrcholy postupne pospájané tak, že tvoria cestu, druhý má prepojené vzájomne všetky vrcholy, ktoré tak tvoria trojuholník. Grafletová štruktúra je nový prístup k skúmaniu sietí a môže priniesť iný pohľad na funkčné siete mozgu.

Celá práca je rozdelená do štyroch kapitol. Prvá kapitola práce sa venuje základom teórie grafov, na ktorú nadväzujeme pri základoch teórie sietí. Bližšie v nej tiež opíšeme funkčné siete mozgu a už spomínané metódy využívajúce grafletovú štruktúru. V nasledujúcej kapitole zadefinujeme ciele našej práce a popíšeme jednotlivé funkčné siete mozgu, ktoré

sme mali k dispozícii. Tretia kapitola sa bude bližšie venovať výskumným metódam a nástrojom použitým pri našom skúmaní. Vysvetlíme v nej, aké algoritmy sme používali a ako sme postupovali pri následnom testovaní. V poslednej kapitole sa tak budeme venovať výsledkom, ktoré vyplývajú z nášho skúmania a analýzy funkčných sietí mozgu.

1. Teoretické východiská

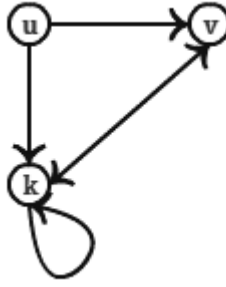
Prvá kapitola obsahuje informácie potrebné na pochopenie riešenej problematiky. Keďže matematickým modelom siete je graf, v časti 1.1. uvedieme niektoré pojmy z teórie grafov, na čo v časti 1.2. nadviažeme teóriou komplexných sietí. Časť 1.3. bude obsahovať popis metód využívajúcich graflety na opísanie štruktúry sietí, ktorým sa neskôr budeme venovať aj vo výskumnej časti našej práce. V poslednej časti tejto kapitoly popíšeme funkčné siete mozgu a stručne niektoré z ich vlastností.

1.1. Základy teórie grafov

Teória grafov je rozsiahla oblasť, ktorej je venované nespočetné množstvo publikácií. Označenie jednotlivých pojmov sa môže medzi učebnicami a knihami líšiť, označenia a definície použité v tejto podkapitole vychádzajú z učebníc [6] a [7], no pre zjednotenie a zamedzenie duplicit boli mierne modifikované. Definície uvádzame stručne a bez dôkazov a zameriavame sa hlavne na pojmy potrebné v ďalších častiach práce.

Graf $G = (V, E)$ je reprezentovaný množinou vrcholov V a množinou hrán E , pričom $E \subseteq [V]^2$; teda prvky množiny E sú dvojprvkové podmnožiny množiny V . Prvky množiny V sú **vrcholy** grafu G , prvky množiny E sú jeho **hranami**. Veľkosť množiny V , t. j. počet vrcholov, označujeme ako $|V|$. Veľkosť množiny E , počet hrán, označujeme $|E|$.

Hranu medzi vrcholmi u a v môžeme označiť ako (u, v) . Rozlišujeme dva typy hrán – orientované a neorientované. Pri reprezentácii **orientovaných** hrán záleží na usporiadaní hrán v dvojici. Hrana (u, v) je tak usporiadanou dvojicou a vedie z vrcholu u do vrcholu v . Orientované hrany (u, v) a (v, u) nie sú totožné. Keby boli dané hrany **neorientované**, daná dvojica by predstavovala totožnú hranu. Hrana začínajúca aj končiaca v tom istom vrchole, t. j. napríklad na obrázku 1 hrana z a do vrcholu k (k, k) , sa nazýva **slučka**. Obr. 1 zobrazuje slučku aj orientované hrany. Graf obsahujúci orientované hrany sa nazýva orientovaný, pokiaľ sa v ňom nachádzajú iba neorientované hrany, tento graf je neorientovaný.



Obr. 1 - Orientovaný graf

Graf označujeme ako **multigraf**, ak medzi dvoma vrcholmi môže existovať viac hrán. Pokiaľ medzi každou dvojicou vrcholov môže existovať maximálne jedna hrana a v grafe sa nenachádzajú slučky, označujeme tento graf ako **jednoduchý**. Pri analýze funkčných sietí mozgu sa budeme zaoberať jednoduchými neorientovanými grafmi, preto ďalšie definície uvádzame pre ne.

Vrchol je **incidentný** s hranou, ak daná hrana z neho vychádza (alebo do neho vchádza). Stupeň vrcholu $d(v)$ označuje počet hrán, s ktorými je daný vrchol incidentný. Pri orientovaných grafoch je stupeň vrcholu súčtom hrán, ktoré do vrcholu vchádzajú a ktoré z neho vychádzajú, čo môžeme označiť ako:

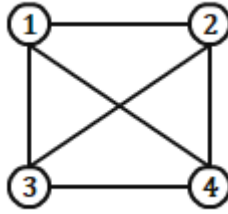
$$d(v) = d(v)^{in} + d(v)^{out}, \quad (1)$$

kde $d(v)^{in}$ nazývame vchádzajúcim stupňom vrcholu a $d(v)^{out}$ vychádzajúcim stupňom. Pre každý graf vieme určiť **minimálny** a **maximálny stupeň vrcholu**, čo je minimum respektíve maximum zo stupňov všetkých vrcholov grafu. **Priemerný stupeň grafu**, $d(G)$, vypočítame ako:

$$d(G) = \frac{1}{|V|} \sum_{v \in V} d(v) \quad (2)$$

Ak je stupeň vrcholu 0, tento vrchol nazývame **izolovaný**. Ak všetky vrcholy majú rovnaký stupeň $d(v)=k$, takýto graf nazývame **k-regulárny**. Pokiaľ je $k = |V| - 1$, tento graf nazývame **kompletný**. V kompletnom grafe existuje hrana medzi každou dvojicou vrcholov. Označujeme ho ako K^n , kde n je počet vrcholov. Na Obr. 2 vidíme K^4 . Počet hrán v kompletnom grafe vypočítame podľa vzťahu (3):

$$|E| = \binom{|V|}{2} = \frac{|V| \cdot (|V| - 1)}{2} \quad (3)$$



Obr. 2 - Kompletný graf stupňa 4

Susedstvom vrcholu v nazývame všetky vrcholy, ktoré sú spojené s vrcholom v hranou.

Klasterizačný koeficient vrcholu v , označovaný ako Cl_v , je pomer počtu hrán medzi susedstvom vrcholu a počtu hrán, ktoré by medzi jeho susedmi mohli existovať, keby susedstvo v tvorilo kompletý graf. Tento koeficient hovorí o prepojení okolia vrcholu v .

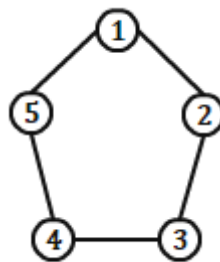
$$Cl_v = \frac{e_v}{\binom{d(v)}{2}} \quad (4)$$

V rovnici (4) e_v značí počet hrán, ktoré existujú medzi susedmi vrcholu v . Kombinačné číslo v menovateli vyjadruje počet všetkých možných hrán medzi týmito vrcholmi. Čím vyšší je klasterizačný koeficient vrcholu, tým viac sú jeho susedia prepojení.

Klasterizačný koeficient celého grafu G , Cl , je priemerom klasterizačných koeficientov všetkých jeho vrcholov. Ak n je počet vrcholov v grafe, môžeme ho vypočítať ako:

$$Cl = \frac{1}{n} \sum_{v=1}^n Cl_v \quad (5)$$

Graf môžeme označiť ako **kružnicu** dĺžky n , alebo n -uholník, C^n , ak obsahuje n vrcholov a n hrán. Jedna hrana je (v_n, v_1) a $(n-1)$ hrán je typu (v_i, v_{i+1}) , pre všetky $i < n$. Takýto graf je 2-regulárny. Na Obr. 3 je pre ilustráciu znázornený C^5 .



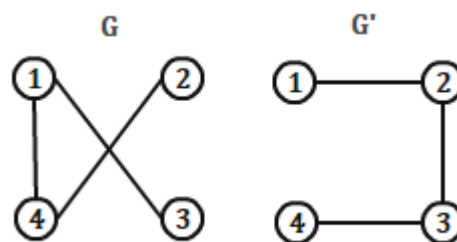
Obr. 3 - Kružnica dĺžky 5

Bipartitný graf je taký graf, ktorého vrcholy môžeme rozdeliť na dve disjunktné podmnožiny tak, že hrany môžu existovať iba medzi vrcholmi z rôznych podmnožín (v rámci vrcholov jednej podmnožiny neexistuje žiadna hrana).

Ak bipartitný graf obsahuje všetky hrany, ktoré môžu vzniknúť medzi vrcholmi oboch podmnožín, nazývame ho **kompletný bipartitný graf** a označujeme ho ako $K_{n,m}$, kde n a m predstavujú veľkosť jednotlivých podmnožín vrcholov.

Špeciálnym typom kompletných bipartitných grafov sú takzvané **hviezdy**. Jedná sa o grafy typu $K_{1,n}$. Z toho vyplýva, že jedna podmnožina obsahuje iba jeden vrchol – stredový – ktorý je ďalej spojený so všetkými ostatnými vrcholmi, ktorých je n . Ak sú usporiadané do kruhu okolo stredového, pripomínajú hviezdu. Pre hviezdy používame označenie S^n , kde n predstavuje počet vrcholov hviezdy usporiadaných okolo stredového.

Dva grafy sú **komplementárne**, ak obsahujú rovnaké vrcholy a zjednotením ich hrán vznikne kompletný graf. Ak grafy G a G' sú komplementárne, graf G je komplementom grafu G' a naopak. Obr. 4 znázorňuje príklad komplementárnych grafov.



Obr. 4 - Komplementárne grafy

Grafy $G = (V, E)$ a $G' = (V', E')$ môžeme označiť ako **izomorfné**, ak obsahujú rovnaký počet vrcholov a existuje zobrazenie také, že:

$$\varphi: V \rightarrow V'; (u, v) \in E \Leftrightarrow (\varphi(u), \varphi(v)) \in E' \text{ pre } \forall u, v \in V \quad (6)$$

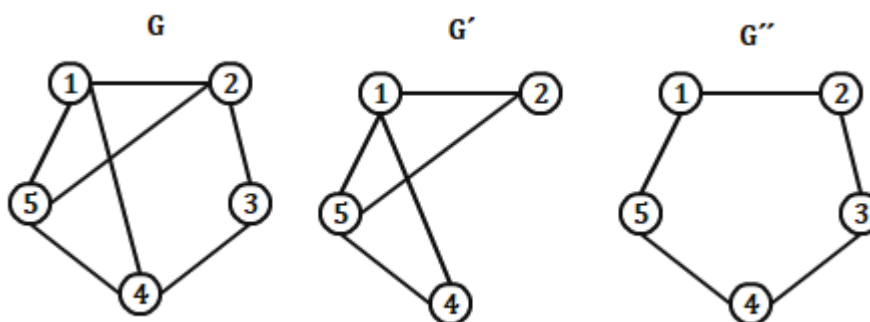
Teda pôvodných vrcholy z grafu $G = (V, E)$ vieme zobrazit' na nové vrcholy v grafe $G' = (V', E')$ tak, že ak v pôvodnom grafe existovala hrana medzi dvoma vrcholmi, bude existovať aj medzi ich zobrazeniami v novom grafe. Ak sa graf zobrazí sám na seba, takže $G = G'$ (tj. $V = V'$ a $E = E'$), označujeme to ako **automorfizmus**. Graf označujeme ako **samokomplementárny**, ak je izomorfný so svojim komplementom.

Majme dva grafy, graf $G=(V, E)$ a graf $G'=(V', E')$. Nech graf G' je **podgrafom** grafu G , potom musí platiť:

$$(V' \subseteq V) \wedge (E' \subseteq E) \quad (7)$$

Graf G môžeme označiť aj ako **nadgraf** grafu G' .

Ak podgraf G' obsahuje všetky hrany $(u, v) \in E$ pre $\forall u, v \in V'$, hovoríme o **indukovanom** podgrafe. Indukovaný podgraf tak zachováva všetky hrany ktoré existovali v pôvodnom grafe medzi podmnožinou vrcholov tvoriacich daný podgraf. **Pokrývajúci** podgraf je graf, v ktorom $V = V'$, t. j. graf, ktorý obsahuje všetky vrcholy pôvodného grafu s tým, že nemusí zachovať všetky hrany. Graf, ktorý je zároveň pokrývajúcim aj indukovaným grafom, je graf totožný s pôvodným grafom.



Obr. 5 - Indukovaný a pokrývajúci podgraf

Na Obr. 5 je graf G' indukovaným podgrafom grafu G , keďže zachováva všetky hrany medzi vrcholmi 1, 2, 4 a 5. Graf G'' je pokrývajúci podgraf grafu G .

Sledom dĺžky k rozumieme takú postupnosť vrcholov a hrán $v_0 e_1 v_1 e_2 \dots e_k v_k$, kde:

$$e_1 = (v_0, v_1), e_2 = (v_1, v_2) \dots e_k = (v_{k-1}, v_k) \quad (8)$$

V slede sa môžu opakovať aj hrany aj vrcholy. Ak sa v slede neopakujú hrany, nazýva sa **ťah**. Ak sa v ťahu neopakujú vrcholy, označujeme ho ako **cesta**.

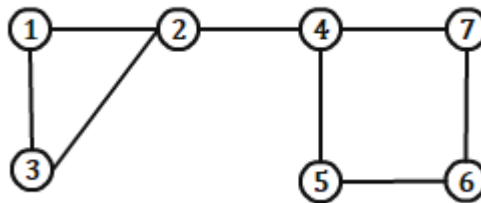
Cestou v grafe označujeme postupnosť vrcholov postupne pospájaných hranami, kde vrchol v_0 je začiatok cesty a vrchol v_k jeho koniec. Počet hrán určuje dĺžku cesty. Ak je cesta dlhšia ako 2 a začína a končí v rovnakom vrchole, nazývame ju **cyklus**.

Vzdialenosť vrcholov v grafe G , označovaná ako $dist_G(u, v)$ je dĺžka najkratšej cesty medzi vrcholmi u a v . Ak takáto cesta neexistuje, potom $dist_G(u, v) = \infty$. vzdialenosť vrcholu od seba samého je 0.

Separácia vrcholu l je priemerom vzdialenosti všetkých dvojíc vrcholov v grafe. Ak má graf n vrcholov, môžeme tento vzťah zapísať nasledovne:

$$l = \frac{\sum_{u,v \in V} \text{dist}_G(u,v)}{\binom{n}{2}} \quad (9)$$

Ak existuje cesta medzi každou dvojicou vrcholov, graf nazývame **spojitý** alebo súvislý. V nespojitom grafe existujú aspoň 2 **komponenty**, čo sú podgrafy, ktoré nie sú navzájom prepojené. Ak existuje v grafe hrana, odobratím ktorej sa súvislý graf rozpadne na dva komponenty, túto hranu označujeme ako **most**. Vrcholy tejto hrany nazývame **artikulácie**.



Obr. 6 - Súvislý graf

Na Obr. 6 je mostom hrana (2, 4) a vrcholy 2 a 4 sú artikuláciami.

Eulerovský ťah, je taký ťah, ktorý prejde každou hranou grafu práve raz. **Eulerovská kružnica** je Eulerovský ťah, ktorý začína a aj končí v rovnakom vrchole.

Známa Eulerova veta z roku 1736, ktorá bola prvou vetou z teórie grafov, znie: „*Graf má Eulerovskú kružnicu práve vtedy, keď majú všetky jeho vrcholy párny stupeň.*“ Z toho vyplýva, že graf má Eulerovský ťah práve vtedy, keď má dva vrcholy nepárneho stupňa.

Hamiltonovská cesta, je naopak taká cesta, ktorý prejde každým vrcholom grafu práve raz. Analogicky, **Hamiltonovská kružnica**, je Hamiltonovská cesta, ktorá začína a aj končí v rovnakom vrchole.

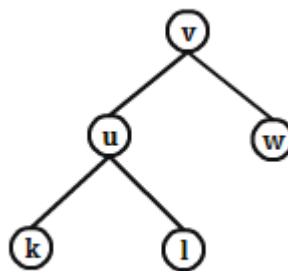
Graf, ktorý neobsahuje cykly, nazývame **acyklický**. Pokiaľ je acyklický graf spojitý, nazývame ho **strom**. V strome všetky vrcholy so stupňom 1 nazývame **listy** a každý strom obsahuje aspoň dva listy. Ak spojitý graf obsahuje n vrcholov, potom musí obsahovať presne $n-1$ hrán, aby bol stromom. Pre stromy tiež platí, že medzi každou dvojicou vrcholov musí existovať práve jedna cesta.

Graf obsahujúci viac stromov (acyklický, nespojitý), nazývame **les**. Každý jeho komponent je potom strom.

Stromy, ktoré obsahujú jeden špeciálne vyznačený vrchol, nazývame **koreňové**. V závislosti od koreňa môžeme zvoliť orientáciu. Potom **predchodca** (otec, rodič) vrcholu u je vrchol v , ak existuje orientovaná hrana (v,u) . Vrchol u nazývame **nasledovníkom** (synom) vrcholu v . Predok vrcholu u , je akýkoľvek vrchol, nachádzajúci sa na ceste z koreňa do vrcholu u . Potomok vrcholu u je, analogicky, taký vrchol, pre ktorý je vrchol u predkom.

Okrem koreňa poznáme v koreňovom strome ešte dva typy vrcholov – **listy**, čo sú vrcholy, ktoré nemajú žiadnych synov, a **vnútorné vrcholy**, čo sú všetky ostatné vrcholy s aspoň jedným synom.

Hĺbka stromu je dĺžka najdlhšej cesty z koreňa do listu meraná počtom hrán.



Obr. 7 – Strom

Na Obr. 7 vidíme strom hĺbky 2 s koreňom v , vnútorným vrcholom u a listami w , k a l .

Každý spojitý graf obsahuje indukovaný pokrývajúci podgraf, ktorý je koreňovým stromom. Ten nazývame **kostra** grafu.

1.2. Modely sietí

Pojmom **komplexná sieť**, označujeme taký graf, ktorý má netriviálne topologické vlastnosti a veľký počet vrcholov. Komplexné siete sa majú ešte ďalšie typické vlastnosti vhodné pre modelovanie reálnych situácií. [8]

Jednou z nich je **distribúcia stupňa vrcholu** $p(k)$. Je to funkcia, ktorá reprezentuje pravdepodobnosť, že náhodne vybraný vrchol v grafe má práve stupeň rovný k .

$$p(k) = \frac{\sum_{v \in V} \delta(k - d(v))}{N} \quad (10)$$

Ďalšou, pomerne ľahko definovateľnou a merateľnou, štatistickou vlastnosťou je tzv. **klasterizačná distribúcia**, ktorá popisuje distribúciu priemerných klasterizačných koeficientov pre vrcholy stupňa k .

$$cl(k) = \frac{\sum_{v \in V} \delta(k - d(v)) \cdot Cl_v}{k} \quad (11)$$

Existuje viacero modelov sietí, medzi najznámejšie patria náhodné siete, siete malého sveta, bezškálové siete, či siete s pravidelnou štruktúrou.

1.2.1. Náhodné siete

Pojem **náhodné siete** zdefinovali P. Erdős a A. Rényi v 1959 [9], neskôr sa viacero matematikov venovalo ich vlastnostiam. Najznámejšia je asi práca B. Bollobasa [10].

Erdős a Rényi zdefinovali dva modely (neskôr označované ako ER modely). Jeden z nich je označovaný ako $G_{N,m}$ model. Je to graf s N vrcholmi a m hranami, ktoré náhodne vyberieme zo všetkých možných hrán, ktorých počet je $\binom{N}{2}$. Druhý model $G_{N,\pi}$ tiež obsahuje N vrcholov, ale hrany medzi nimi vznikajú s pravdepodobnosťou π . Počet hrán v tomto grafe teda bude:

$$|E_{N,\pi}| = \pi \binom{N}{2} \quad (12)$$

Zaujímavým zistením autorov je, že pre konkrétnu pravdepodobnosť buď takmer všetky grafy majú nejakú vlastnosť Q , alebo naopak takmer žiadne.

Pre nás sú však zaujímavejšie iné vlastnosti [10], [8]. **Distribúciu stupňov vrcholov** pre pravdepodobnostný model $G_{N,\pi}$ predstavuje vzťah:

$$p(k) = \binom{N-1}{k} \pi^k (1-\pi)^{N-1-k} \quad (13)$$

Táto distribúcia je binomická, pričom ale pre veľké počty vrcholov N sa táto distribúcia začne podobať na Poissonovo rozdelenie:

$$p(k) \simeq \frac{(N\pi)^k e^{-N\pi}}{k!} \quad (14)$$

Pre veľké náhodné grafy platí, že s veľkou pravdepodobnosťou nebude počet vrcholov vzdialených o vzdialenosť x od vrcholu v oveľa menší ako x -tá mocnina priemerného stupňa

vrcholu v grafe. Ak označíme priemerný stupeň vrcholu \bar{k} , potom **separáciu vrcholov** náhodného grafu l_{random} vypočítame ako:

$$l_{random} = \frac{\log(N)}{\log(\bar{k})} \quad (15)$$

Pri výpočte **klasterizačného koeficientu** sledujeme, ako je prepojené susedstvo vrcholu. V náhodných grafoch je pravdepodobnosť, že dva vrcholy v susedstve vrcholu sú prepojené, rovnaká ako pravdepodobnosť prepojenia dvoch náhodne vybraných vrcholov.

$$cl_{random} = \frac{\bar{k}}{N} \quad (16)$$

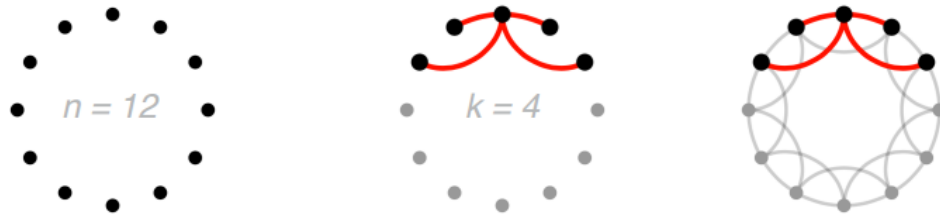
Keďže klasterizačný koeficient náhodných grafov priamo závisí od počtu vrcholov (N), tento model nie je dobre použiteľný na všetky reálne siete. V nich totiž ich štatistické vlastnosti nemusia závisieť od ich veľkosti.

1.2.2. Siete malého sveta

Pojem **siete malého sveta** vychádza zo známeho fenoménu malého sveta, bližšie opísaného napríklad v [11]. Stručne vysvetlené, ak to, že sa dvaja ľudia poznajú, nazveme ich prepojením, tak ak sú dvaja ľudia prepojení cez spoločného známeho, ich prepojenie je potom dĺžky 2. Pre dvoch náhodne vybraných ľudí tak existuje postupnosť prepojení cez spoločných známych a ich známych atď. Viacero výskumov potvrdzuje, že táto postupnosť je v priemere neočakávane krátka a pohybuje sa okolo čísla 5.

Tento fenomén inšpiroval sociológa Duncana J. Watts a matematika Stevena Strogatza, k vytvoreniu modelov sietí malého sveta [12]. Tieto modely sú často označované ako Watts-Strogatz modely. Popíšeme si jeden z ich (ďalej označovaný ako WS model).

Majme n vrcholov usporiadaných do tvaru kružnice. Zvoľme si stupeň k , a vytvoríme k -regulárny graf tak, že každý vrchol bude prepojený hranou s k najbližšími susedmi. Pre $k = 2$ to bude kružnica veľkosti n , t. j. C^n . Pre $k = 4$ a $n = 12$ je situácia zobrazená na Obr. 8.



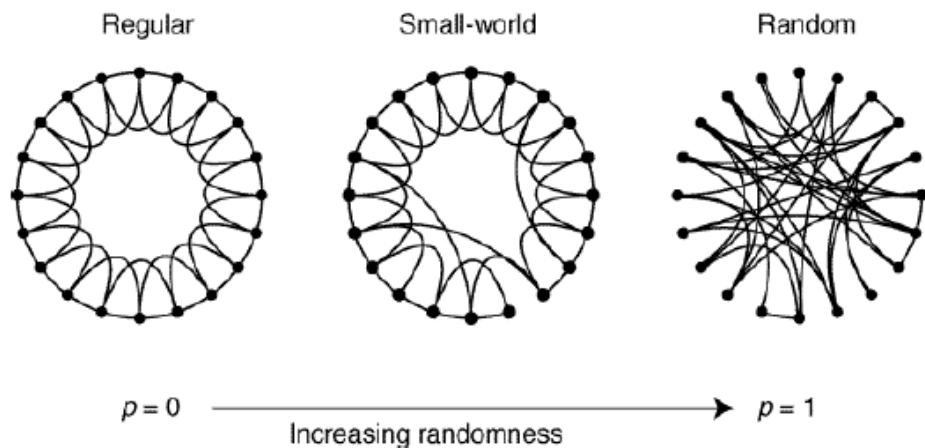
Obr. 8 - Vytvorenie počiatočného grafu (prevzatý z [13])

Na takto vytvorený graf aplikujeme nasledujúci algoritmus:

- 1) Vyberieme počiatočný vrchol v a jeho najbližšieho suseda v smere pohybu hodinových ručičiek. S pravdepodobnosťou p jeden koniec hrany, ktorá ich spája, prepojíme k náhodne vybranému vrcholu z grafu tak, aby sme nevytvorili duplicitné hrany. Inak (t. j. s pravdepodobnosťou $(1-p)$) túto hranu ponecháme.
- 2) Tento proces opakujeme v smere pohybu hodinových ručičiek, kým sa nedostaneme späť k počiatočnému vrcholu.
- 3) Pre počiatočný vrchol v vyberieme jeho druhého najbližšieho suseda v smere pohybu hodinových ručičiek a opakujeme postup z bodov 1) a 2).
- 4) Tento postup opakujeme, kým nie je prehodnotená každá hrana grafu. Keďže pôvodný graf má $\frac{nk}{2}$ hrán, počet cyklov algoritmu (jeden cyklus = prejde všetkých vrcholov) je $\frac{k}{2}$.

Pravdepodobnosť p sa nazýva **pravdepodobnosť prelinkovania hrany**. Pre pravdepodobnosť $p = 0$ sa pôvodný graf nezmení, pre pravdepodobnosť $p = 1$ vznikne náhodne poprepájaný graf. Sieť malého sveta tak interpoluje medzi obidvoma týmito hraničnými hodnotami. Najlepšie to ilustruje Obr. 9.

V takto vytvorených sieťach malého sveta sú všetky vrcholy vzájomne dobre prepojené. Klasterizačný koeficient siete je tak pomerne veľký. So zvyšujúcou sa pravdepodobnosťou p sa skracaje vzdialenosť medzi vrcholmi, keďže v sieti vznikajú tzv. krátke spojenia, čo sú spojenia dvoch nesusediacich, niekedy od seba veľmi vzdialených, vrcholov. Separácia vrcholov l je tak v sieťach malého sveta pomerne nízka.



Obr. 9 - Zmeny v algoritme v závislosti od p (prevzaté z [12])

1.2.3. Sieťe s preferenčným pripájaním

Ďalší zo známych modelov sietí je model A. L. Barabásiho a R. Albert, označovaný ako BA model. V článku [14] zistili, že distribúcia stupňov vrcholov v ich modeli je daná ako:

$$p(k) \propto k^{-\gamma}, \quad (17)$$

kde γ je škálovací exponent. Sieťe s takouto distribúciou nazvali **bezškálovými**, pretože sa v nich nenachádza žiaden stupeň vrcholu, ktoré by bol typický pre celú sieť.

BA model je založený na nasledujúcom algoritme: Začíname s n_0 vrcholmi. Každú časovú jednotku pridávame do siete nový vrchol s m ($m < n_0$) hranami, ktoré ho spoja s už existujúcimi vrcholmi v sieti. Pravdepodobnosť, že nový vrchol bude pripojený k vrcholu v , závisí od stupňa vrcholu v .

$$\Pi(d(v)) = \frac{d(v)}{\sum_{u \in V} d(u)} \quad (18)$$

Po t krokoch tak vzniká náhodná sieť s ($t < t_0$) vrcholmi a mt hranami. Táto sieť je v **škálovo invariantnom** stave. Pravdepodobnosť, že vrchol má práve k hrán je mocninová funkcia s exponentom $\gamma = 3.0$.

Rýchlosť s akou vrchol získava nové hrany v čase t môžeme vyjadriť vzt'ahom:

$$\frac{\partial d(v)}{\partial t} = \frac{d(v)}{2t} \quad (19)$$

Ak pre každý vrchol v označíme t_v ako čas, kedy sa do siete pripojil, stupeň vrcholu v v čase t , $d(v)_t$, potom môžeme vyjadriť ako:

$$d(v)_t = m \left(\frac{t}{t_v} \right)^\beta, \quad (20)$$

kde $\beta = \frac{1}{2}$. Z toho vyplýva, že staršie vrcholy (s menším t_v) budú mať viac hrán ako tie vrcholy, ktoré prišli do siete neskôr (väčšie t_v). Staršie vrcholy tak zväčšujú svoju pravdepodobnosť, že sa k nim pripojí ďalší vrchol, čo modeluje ďalší fenomén z reálneho sveta – „bohatý sa stáva bohatším“, čo priamo súvisí s mocninovou funkciou. Tento fenomén značí to, že čím viac susedov má vrchol, tým je väčšia pravdepodobnosť, že sa k nemu pripoja nové vrcholy. Môžeme to vidieť napríklad v sieťach z reálneho sveta, ako je napr. sieť kontaktov na sociálnych sieťach, sieť citácií vedeckých článkov, alebo aj v sieti webových prepojení [15].

1.3. Graflety

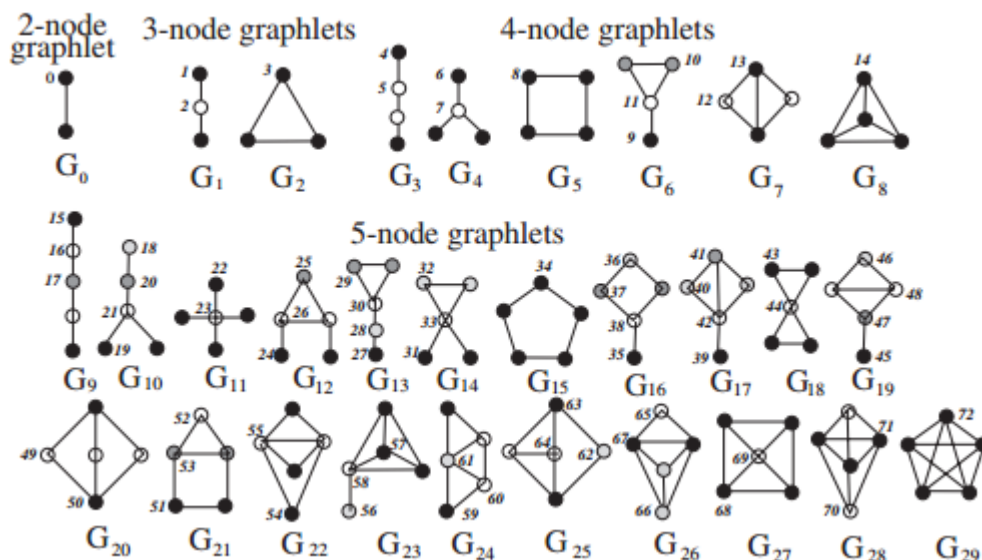
Zisťovanie podobností veľkých sietí je často pamäťovo či časovo náročné, preto vznikajú stále nové prístupy, ako ich porovnávať. Jedným z nich je grafletový popis sietí [5].

Pojem **graflety** bol zadefinovaný po prvýkrát v článku Nataše Pržulj [16]. Cieľom je jednoznačne rozlíšiť všetky malé podgrafy a motívy siete.

Motívy siete¹ sú špeciálne podgrafy siete (hovorí sa o nich aj ako o vzoroch, patternoch), ktoré sa vyskytujú v sieti s pravdepodobnosťou väčšou ako pravdepodobnosť výskytu daných podgrafov v náhodných sieťach [17]. Teda je jasné, že sa zanedbávajú málo frekventované podsiete, alebo podsiete s priemerným výskytom. Z toho dôvodu nie sú motívy veľmi dobre použiteľné na meranie podobnosti dvoch sietí.

Graflety sú malé neizomorfné indukované podgrafy veľkej siete [16]. Na rozdiel od motívov siete, pri grafletoch nesledujeme pravdepodobnosť ich výskytu v sieti. Malé motívy sietí sú teda aj grafletmi, ale nie každý graflet je aj motívom siete. Základné graflety stupňa 2 až 5 vidíme na Obr. 10.

¹ z anglického *network motifs*



Obr. 10 - Graflety (prevzaté z [5])

Z obrázku je jasné, že sa jedná o všetky spojité grafy s daným počtom vrcholov. Najjednoduchší graflet G_0 predstavuje akúkoľvek hranu v grafe. Pre tri vrcholy už máme dva graflety – G_1 a G_2 – prvý sa zvykne označovať ako 2-cesta, t. j. cesta dĺžky 2, druhý ako trojuholník. Všeobecne vieme povedať, že graflety G_0 , G_2 , G_8 a G_{29} sú kompletne grafy daného stupňa (K^2 , K^3 , K^4 resp. K^5), grafy G_2 , G_5 a G_{15} n-uholníky (t. j. C^3 , C^4 a C^5), graflety G_0 , G_1 , G_3 a G_9 n-cesty a graflety G_4 a G_{11} k-hviezdy (S^3 a S^4).

Použitím grafletov vzniklo niekoľko prístupov k skúmaniu vlastností siete. Medzi tri najrozšírenejšie patria: počet grafletov, stupeň grafletov a distribúcia stupňa grafletov.

Nasledujúce metódy budeme opisovať pre G_1 až G_{29} tak, ako v citovaných článkoch. Samozrejme, všetky tieto metódy je možné upraviť pre menší alebo aj pre väčší počet grafletov.

1.3.1. Počet grafletov

Základnou metódou využívajúcou graflety, je spočítanie **výskytov** jednotlivých grafletov v sieti [16]. Označme počet výskytov grafletu G_i v grafe G ako $N_i(G)$. $T(G)$ použijeme na označenie počtu výskytov všetkých grafletov dokopy v grafe G . Platí nasledovný vzťah:

$$T(G) = \sum_{i=1}^{29} N_i(G) \quad (21)$$

Na zistenie podobnosti dvoch sietí na základe počtu grafletov sa používa **relatívna vzdialenosť frekvencií grafletov**² D . Podobnosť dvoch sietí tak nezávisí od počtu vrcholov, či hrán. Z toho dôvodu a zároveň preto, aby najviac vyskytujúce sa graflety neskreslili výsledky, je použitá logaritmická škála:

$$F_i(G) = -\log \frac{N_i(G)}{T(G)} \quad (22)$$

$F_i(G)$ predstavuje logaritmus pomeru počtu výskytov grafletu G_i v grafe k výskytu všetkých grafletov. Relatívnu vzdialenosť frekvencií grafletov pre siete G a H potom vieme vyjadriť ako:

$$D(G, H) = \sum_{i=1}^{29} |F_i(G) - F_i(H)| \quad (23)$$

Čím menšie číslo vyjde, tým sú si dané siete bližšie – t. j. majú približne rovnaký počet jednotlivých typov grafletov.

1.3.2. Vektor distribúcie grafletov

Ďalšou metódou porovnávania sietí využívajúcou graflety je **vektor distribúcie grafletov** (GDV³) [18]. Vychádza z toho, že graflety sú neizomorfné podgrafy siete, ktoré musia byť zároveň aj indukovanými podgrafmi. Preto pre každý vrchol siete vieme zistiť, koľkých grafletov rôzneho typu sa dotýka. Dotyk vrcholu v s grafletom G_x znamená, že vrchol v je súčasťou daného grafletu.

Vrchol siete sa môže každého z grafletov dotýkať v inom vrchole grafletu – napríklad pre graflet G_6 je rozdiel, ak sa ho vrchol dotýka vo vrchole na Obr. 10 označenom čiernou farbou, vrchole označenom bielou alebo v jednom z dvojice vrcholov označených sivou farbou. V článku [5] autorka používa pojem **automorfné orbity**⁴ na rozdelenie jednotlivých vrcholov grafletov do rôznych skupín.

² z anglického *relative graphlet frequency distance*

³ z anglického *graphlet distribution vector*

⁴ z anglického *automorphism orbit*

Automorfná grupa grafu G je grupa, ktorú tvoria všetky automorfizmy grafu. Označme automorfnú grupu grafu G ako $Aut(G)$, potom automorfná orbita vrcholu v z grafu G je definovaná ako:

$$Orb(v) = \{u \in V(G) | u = g(v) \text{ pre nejaké } g \in Aut(G)\}, \quad (24)$$

kde $g(v)$ označuje automorfizmus (t. j. bijekciu, ktorá zobrazí vrchol v na iný vrchol grafu).

Na Obr. 10 sú jednotlivé orbity očíslované od 0 po 72, takže existuje 73 rôznych orbít pre graflety stupňa 2 až 5. Vidíme, že vonkajšie hrany grafletu G_l patria do jednej orbity, kým vnútorný vrchol do druhej. Pri grafletoch G_9 a G_{10} nájdeme až štyri rôzne orbity. Vďaka orbitám vieme presne určiť, v ktorom vrchole grafletu dochádza k dotyku.

Vektor distribúcie grafletov preto má 73 súradníc. Prvou je počet dotykov s grafletom G_0 (čiže orbitou 0), z dôvodu, že graflet G_0 je vlastne hrana, hodnota prvej súradnice tak bude rovná stupňu daného vrcholu. Každé ďalšie číslo na pozícii i predstavuje počet dotykov s orbitou i . Tento vektor sa nazýva podpis⁵ vrcholu [18].

Keďže rozdiely medzi počtom niektorých orbít automaticky ovplyvnia počty iných orbít (napríklad počet orbít 0 automaticky ovplyvní počet orbít 1 alebo 4, počet orbít 2 zas ovplyvní počet orbít 7 a 12 atď.), je potrebné tento vektor ováňovať. Na to slúži **váňovací vektor** W . Rovnako ako podpis vrcholu, aj váňovací vektor sa skladá zo 73 súradníc a každý člen w_i obsahuje váňu orbity i . Aby sme tieto členy vypočítali, potrebujeme každej orbite priradiť číslo o_i , ktoré vyjadruje počet orbít, ktoré ho ovplyvňujú. (Každá orbita ovplyvňuje aj samú seba.) Takzvané ovplyvňovanie môžeme vidieť na príklade – orbitu 15 ovplyvňujú okrem jej samej orbity 0, 1 a 4, keďže graflety G_0 , G_1 a G_3 sú podgrafmi grafletu G_9 , do ktorého orbita 15 patrí, preto $o_{15} = 4$.

Váňovací vektor W potom vypočítame pomocou nasledovného vzorca:

$$w_i = 1 - \frac{\log(o_i)}{\log(73)} \quad (25)$$

⁵ z anglického *signature*

Logaritmická škála je použitá preto, aby významnejšie orbity (teda tie, ktoré sú ovplyvňované menším počtom iných orbít) mali vyššiu váhu. Predelenie $\log(73)$ nastavi hodnoty W do intervalu $\langle 0,1 \rangle$. Odčítavanie takto získaného čísla od 1 iba prehodí váhy – najvyššiu budú mať tie vrcholy, ktoré majú najnižšie o_i .

Pre dva vrcholy u a v vieme určiť vzdialenosť ich počtu dotykov s orbitou i . Označíme ho ako $D_i(u, v)$. u_i predstavuje i -tu súradnicu podpisu vrcholu, takže počet dotykov s orbitou i .

$$D_i(u, v) = w_i \times \frac{|\log(u_i + 1) - \log(v_i + 1)|}{\log(\max\{u_i, v_i\} + 2)} \quad (26)$$

Opäť je použitá logaritmická škála, keďže rozdiely medzi jednotlivými vrcholmi môžu byť veľké. Vďaka logaritmom skúmame relatívny a nie absolútny rozdiel. Pridanie jednotiek do logaritmov v čitateli má predísť tomu, aby išli do nekonečna. Z rovnakého dôvodu sa do logaritmu v menovateli pridáva 2. Logaritmus v menovateli slúži na škálovanie do intervalu $(0,1)$.

Úplná vzdialenosť dvoch vrcholov, t. j. pre všetky orbity, sa vypočíta podľa vzorca:

$$D(u, v) = \frac{\sum_{i=0}^{72} D_i}{\sum_{i=0}^{72} w_i} \quad (27)$$

Vieme, že $D(u, v) \in \langle 0,1 \rangle$. Ak je vzdialenosť rovná 0, podpisy vrcholov sú rovnaké. Z toho vyplýva, že podpisová podobnosť dvoch vrcholov, $S(u, v)$ bude vyjadrená ako:

$$S(u, v) = 1 - D(u, v). \quad (28)$$

1.3.3. Distribúcia stupňov grafletov

Distribúciu stupňov grafletov⁶ (GDD) zadefinovala opäť Nataša Pržulj [5] a vychádza z rovnakých vlastností grafletov, ako sme spomínali v predchádzajúcej časti. Stále skúmame graflety vo vzťahu k ich 73 orbitám.

⁶ z anglického *graphlet degree distribution*

Pre každú automorfnú orbitu j si vytvoríme **vzorovú distribúciu počtu vrcholov** grafu G takých, ktoré sa danej orbite dotýkajú r -krát. Označuje sa $d_G^j(r)$. Ako sme si v predchádzajúcej časti vysvetlili, dotyk vrcholu siete s orbitou o znamená, že daný vrchol siete tvorí vrchol grafletu prislúchajúci orbite o . Jeden vrchol však môže patriť do viacerých grafletov rovnakého typu. Aby sa znížil vplyv vysokých stupňov grafletov niektorých vrcholov, predelíme túto distribúciu číslom r :

$$S_G^j(r) = \frac{d_G^j(r)}{r} \quad (29)$$

Po normalizovaní dostávame zlomok celkovej plochy pod distribučnou krivkou patriacou stupňu r .

$$N_G^j(r) = \frac{S_G^j(r)}{\sum_{l=1}^{\infty} S_G^j(l)} \quad (30)$$

Pre dve siete G a H a nejakú konkrétnu orbitu j vieme určiť vzdialenosť medzi ich normalizovanými distribúciami stupňa j :

$$D^j(G, H) = \left(\sum_{k=1}^{\infty} [N_G^j(k) - N_H^j(k)]^2 \right)^{\frac{1}{2}} \quad (31)$$

Vzdialenosť je v intervale $(0,1)$, kde 1 znamená, že distribúcie stupňov grafletu stupňa j sietí G a H sú od seba úplne rozdielne a 0 znamená, že sú identické. Pomocou $D^j(G, H)$ môžeme počítať **j-ty GDD súhlas**⁷, čo je podobnosť dvoch sietí na základe orbity j :

$$A^j(G, H) = 1 - D^j(G, H) \quad (32)$$

Ten vieme vyjadriť pre každé j od 0 po 72, t. j. pre každú jednu orbitu. **Súhlas dvoch sietí** potom dostaneme buď ako aritmetický $A_a(G, H)$, alebo ako geometrický priemer $A_g(G, H)$, všetkých j -tych GDD súhlasov:

$$A_a(G, H) = \frac{1}{73} \sum_{j=0}^{72} A^j(G, H) \quad (33)$$

⁷ z anglického *GDD agreement*

$$A_g(G, H) = \left(\prod_{j=0}^{72} A^j(G, H) \right)^{\frac{1}{73}} \quad (34)$$

Čím bude súhlas dvoch sietí bližšie k 1, tým si budú podobnejšie a naopak.

1.4. Funkčné siete mozgu

V predchádzajúcej časti sme sa venovali komplexným sieťam, ktoré môžu byť dobrým modelom pre siete z reálneho sveta, ako napríklad siete prepojenosti stránok na internete [19], už spomínané siete osobných kontaktov [12], ale aj biologické siete – napríklad siete proteínových interakcií [16], ktoré skúma tím Nataše Pržulj, alebo aj funkčné siete mozgu [20].

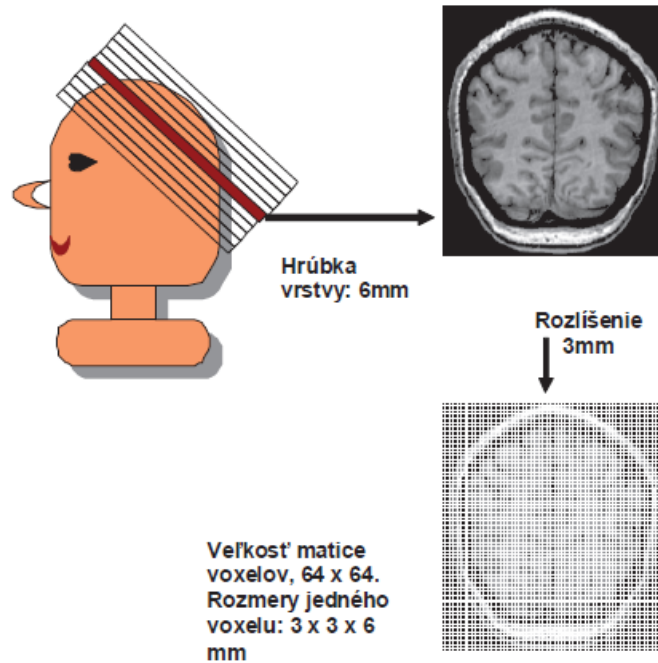
Vytvoriť funkčnú sieť z aktivity ľudského mozgu sa podarilo pomocou fMRI [20]. fMRI funguje na princípe elektromagnetickej rezonancie, ktorá ovplyvňuje vodíky v molekule vody. Pri magnetickej rezonancii sa do tela vyšetrovaného človeka vyšle impulz, na ktorý vodík reaguje signálom s určitou silou. Tá sa líši v závislosti od okolia, preto je možné rozlíšiť šedú hmotu mozgu, bielu hmotu mozgu a mozgovo-miechovú tekutinu na štrukturálnych obrazoch mozgu [21].

Nervové bunky potrebujú na svoje fungovanie kyslík a jeho zvýšená potreba sa prejavuje pri aktivite tej časti mozgu, v ktorej sa nervové bunky nachádzajú. Kyslík sa do buniek dostáva pomocou hemoglobínu v červených krvinkách. V závislosti od aktivity sa mení aj okysličenosť krvi v mozgu, ktoré **fMRI** meria. Táto metóda sa nazýva BOLD zobrazovanie, t. j. zobrazovanie v závislosti na pomere okysličeného a odkysličeného hemoglobínu. Na začiatku aktivity miera okysličenia krvi mierne klesá, následne začne stúpať a svoj vrchol dosiahne približne po 6 sekundách od aktivity.

Aktivita mozgu sa meria pomocou fMRI. Meranej osobe sú zadávané jednoduché kognitívne úlohy, ako napr. zopakovanie slova po výskumníkovi, preklady viet, alebo aj mechanické činnosti ako otváranie a zatváranie konkrétnej dlane, či rytmické klepanie prstami. Taktiež pri niektorých výskumoch nemusia ľudia, ktorých aktivita mozgu je meraná, vykonávať žiadne činnosti, ale iba počúvajú hudbu, či hovorené slovo [15].

Výstupom fMRI je zobrazenie zaznamenatej aktivity mozgu online. fMRI skener je schopný zmerať signál z malej oblasti mozgu zvanej **voxel**. 3D obraz mozgu sa rozdelí pomocou mriežky vo zvislom aj vertikálnom smere na voxely, viď. Obr. 11. Voxel je reálna

časť mozgu o veľkosti približne 3mm^3 , pre ktorú vieme určiť jej presné umiestnenie. Keďže v mozgu zdravého človeka sa nachádza približne 10^{11} neurónov, v každom voxelu sa ich nachádzajú rádovo tisíce.



Obr. 11 - Schéma usporiadania 3D matice voxelov pri fMRI, prevzaté z [16]

Otázkou je, ako z takto nameraných aktivít voxelov vytvoriť funkčnú sieť mozgu. Logicky vyplýva, že vrcholmi siete budú jednotlivé voxely. Ak je v dvoch voxeloch nameraný signál skorelovaný nad zvoleným prahom, dáme medzi ne hranu. Určujeme to pomocou výpočtu korelačného koeficientu.

Označme aktivitu voxelu i v čase t ako $s_i(t)$ a korelačný koeficient voxelov i a j ako $r(i,j)$. Potom, ak $\langle . \rangle$ označuje priemerovanie v čase, vieme **korelačný koeficient** vypočítať ako:

$$r(i,j) = \frac{\langle s_i(t) \cdot s_j(t) \rangle - \langle s_i(t) \rangle \cdot \langle s_j(t) \rangle}{\sigma(s_i(t)) \cdot \sigma(s_j(t))} \quad (35)$$

Kde $\sigma(s_i(t))$ označuje deriváciu aktivity voxelu i v čase t [20]. Na to, aby sme odstránili náhodné korelácie a zároveň znížili počet hrán, sa používa takzvaný **korelačný prah** r_c . Je to hodnota, ktorá určuje minimálnu hranicu korelácie voxelov, z ktorej vznikne v sieti hrana. Čím vyšší je korelačný prah, tým viac skutočných spojení voxelov dostávame, ale zároveň vznikajú veľmi malé siete, na ktorých môže byť problematické skúmať štruktúru a ich štatistické vlastnosti. Znižovaním korelačného prahu dostávame väčšie siete, ale vyskytuje

sa v nich aj viac náhodných korelácií. Optimálne hodnoty korelačného prahu sú tak okolo hodnôt $0,8 - 0,9$. Funkčná sieť mozgu teda bude graf [8]:

$$G = (V, E) = \{(u, v) \mid \forall u, v \in V \wedge r(u, v) > r_c\} \quad (36)$$

Funkčné siete mozgu sú predmetom skúmania už niekoľko rokov. Existujú rôzne prístupy k skúmaniu vlastností sietí, k výberu dát a s tým často súvisia aj rôzne výsledky. Napr. v [15] autori zistili, že funkčné siete mozgu sa správajú ako siete malého sveta a v [22] a [23] sa okrem vlastností sietí malého sveta potvrdila aj bezškálovosť týchto sietí. Z toho dôvodu sa budeme aj v našom výskume venovať porovnaniu reálnych dát s modelmi bezškálových sietí a sietí malého sveta.

2. Ciele práce a dostupné dáta

V tejto kapitole bližšie vysvetlíme, čo je cieľom našej práce a čomu sa budeme pri našom výskume venovať. S tým súvisí aj popis dát, ktoré máme k dispozícii a na ktorých bude výskum prebiehať.

2.1. Ciele práce

Cieľom našej práce je zanalyzovať funkčné siete mozgu vytvorené na báze reálne nameraných dát z hľadiska ich grafletovej štruktúry. Máme k dispozícii tri skupiny dát – funkčné siete mozgu mladých zdravých participantov, funkčné siete mozgu starých zdravých participantov nad 65 rokov a poslednú skupinu tvoria funkčné siete mozgu participantov nad 65 rokov, ktorým bola diagnostikovaná Alzheimerova choroba. Bližší popis funkčných sietí sa nachádza v nasledujúcej podkapitole.

Samotné graflety a s nimi spojené výskumné a porovnávacie metódy sú pomerne nové a ešte nie veľmi rozšírené. S tým súvisí aj fakt, že takýto prístup ešte nebol použitý v žiadnom výskume funkčných sietí mozgu. Vychádzajúc z výskumov zaoberajúcich sa grafletmi, ktoré sme uviedli v predchádzajúcej časti, môže však grafletová štruktúra priniesť nové poznatky do oblasti výskumu funkčných sietí mozgu.

Dôležitou súčasťou analýzy nám dostupných funkčných sietí mozgu bude skúmanie, či grafletová štruktúra vie dostatočne odlíšiť jednotlivé skupiny ľudí, a či z tohto hľadiska existujú nejaké význačné vlastnosti funkčných sietí participantov s Alzheimerovou chorobou.

Ďalším z cieľov je pomocou grafletovej štruktúry sietí zistiť, aké vlastnosti majú funkčné siete mozgu. Pomocou porovnaní s niektorými vybranými modelmi komplexných sietí, ako sú napríklad siete malého sveta, budeme skúmať ich podobnosti a odlišnosti s danými modelmi a v rámci skúmania sa sústredíme aj na rozdiely medzi jednotlivými skupinami. Takýto výskum bude užitočný v kontexte už existujúcich výskumov, z ktorých sme niektoré spomenuli v predchádzajúcej kapitole.

2.2. Reálne dáta

Dáta sme získali vďaka prof. Ľubici Beňuškovej a jej kolegyni Liz Franz z University of Otago, Dunedin na Novom Zélande. Dáta sa skladajú z troch skupín – mladí zdraví participantí (do 25 rokov), starí zdraví participantí (nad 65 rokov) a starí participantí (tiež nad 65 rokov), ktorým bola diagnostikovaná Alzheimerova choroba. Mali sme k dispozícii

tri typy sietí podľa korelačného prahu (vid'. 1.4), kvôli veľkému množstvu dát sme si vybrali iba jednu skupinu, a to tú, kde mal korelačný prah najvyššiu hodnotu $r_c=0.962249$. Siete vytvorené podľa tohto korelačného prahu sme zvolili z dôvodu, že čím vyšší korelačný prah je použitý, tým viac zmysluplných spojení sa vo funkčných sieťach mozgu nachádza. Pre každú skupinu boli k dispozícii dáta z mozočku (cerebellum), záhlavného laloku (occipital) a z celého mozgu. Na analýzu sme vybrali dáta z celého mozgu.

Každá sieť je zapísaná v dvoch textových súboroch. Jeden obsahuje hrany - je tvorený dvojicami čísel (jeden riadok je jedna dvojica), čísla sú indexy vrcholov a každá dvojica reprezentuje hranu. Druhý obsahuje súradnice jednotlivých vrcholov. Pri našej analýze boli potrebné iba súbory obsahujúce hrany, keďže sme sa nezaoberali reálnym umiestnením vrcholov (tj. voxelov) v skene mozgu. V tabuľkách vidíme základné informácie o jednotlivých sieťach, ktoré sme analyzovali.

Pri 11 participantoch, ktorým bola diagnostikovaná Alzheimerova choroba (Tabuľka 1), sú veľké rozdiely v počtoch hrán. Priemerný počet hrán je najnižší spomedzi všetkých troch skupín, 24 936 hrán, so štandardnou odchýlkou 14 894 hrán. Rovnako je najnižší aj priemerný počet vrcholov – 6 118 so smerodajnou odchýlkou 911 vrcholov.

| Názov súboru | Počet vrcholov | Počet hrán |
|--------------|----------------|------------|
| awith_2 | 5 746 | 13 340 |
| awith_6 | 5 988 | 12 738 |
| awith_9 | 4 740 | 27 840 |
| awith_12 | 7 305 | 20 938 |
| awith_30 | 7 806 | 64 365 |
| awith_31 | 5 746 | 25 065 |
| awith_33 | 5 673 | 23 174 |
| awith_35 | 6 972 | 33 837 |
| awith_36 | 5 598 | 13 928 |
| awith_37 | 5 353 | 26 393 |
| awith_40 | 6 372 | 12 675 |
| PRIEMER | 6 118 | 24 936 |

Tabuľka 1 - Základné vlastnosti sietí zo skupiny starých participantov s Alzheimerom

Spracovali sme 14 funkčných sietí mozgu starých zdravých participantov (Tabuľka 2). Priemerný počet hrán bol najvyšší zo všetkých skupín – 46 689 hrán, so štandardnou odchýlkou 28 182 (ktorá bola zo všetkých troch skupín najväčšia). Priemerný počet vrcholov bol len o 250 vrcholov väčší ako priemerný počet vrcholov sietí participantov s Alzheimerom – 6 368 so smerodajnou odchýlkou 706 vrcholov.

| Názov súboru | Počet vrcholov | Počet hrán |
|--------------|----------------|------------|
| awout_4 | 6 997 | 64 597 |
| awout_5 | 6 676 | 86 646 |
| awout_7 | 7 084 | 22 211 |
| awout_8 | 6 169 | 87 281 |
| awout_10 | 5 772 | 77 125 |
| awout_11 | 7 743 | 28 401 |
| awout_13 | 5 113 | 33 434 |
| awout_14 | 5 790 | 64 578 |
| awout_15 | 6 823 | 12 600 |
| awout_19 | 6 275 | 18 097 |
| awout_24 | 6 829 | 13 203 |
| awout_27 | 5 816 | 41 458 |
| awout_28 | 6 441 | 76 698 |
| awout_38 | 5 629 | 27 310 |
| PRIEMER | 6 368 | 46 689 |

Tabuľka 2 - Základné vlastnosti sietí zo skupiny starých zdravých participantov

Priemerný počet hrán v 13 sieťach mladých zdravých participantov (Tabuľka 3) je 41 809 hrán so smerodajnou odchýlkou 12 093 (ktorá bola najnižšia zo všetkých skupín). Priemerný počet vrcholov je najvyšší zo všetkých skupín – 7 394 vrcholov so smerodajnou odchýlkou 1 177.

| Názov súboru | Počet vrcholov | Počet hrán |
|--------------|----------------|------------|
| young_16 | 7 349 | 45 446 |
| young_17 | 6 449 | 42 267 |
| young_18 | 6 104 | 46 965 |
| young_20 | 8 529 | 35 746 |
| young_21 | 8 864 | 50 842 |
| young_22 | 6 488 | 43 899 |
| young_23 | 6 694 | 28 688 |
| young_25 | 9 003 | 33 803 |
| young_26 | 6 124 | 59 355 |
| young_29 | 8 826 | 39 428 |
| young_32 | 6 936 | 64 917 |
| young_34 | 8 579 | 28 712 |
| young_41 | 6 187 | 23 454 |
| PRIEMER | 7 394 | 41 809 |

Tabuľka 3 - Základné vlastnosti sietí zo skupiny mladých participantov

3. Implementácia výskumných metód

Pri prvotných testovaniach dát sme zistili, že dostupné funkčné siete mozgu obsahujú veľké množstvo hrán a v prípade, kedy som chceli používať všetky grafletové metódy popísané v časti 1.3., boli by porovnávaná pomerne časovo a aj pamäťovo náročná. Zvolili sme preto iba dve metódy – počet grafletov (1.3.1) a distribúciu stupňov grafletov (1.3.3). Postupným testovaním sme zistili, že naša výpočtová sila nám neumožňuje pri analýze sietí používať graflety s piatimi vrcholmi, a pri distribúcii stupňov grafletov sme sa museli vzdať použitia aj takmer všetkých grafletov so štyrmi vrcholmi, kvôli veľkej pamäťovej náročnosti. Konkrétne zmeny a úpravy oproti originálnym metódam popíšeme v tejto kapitole.

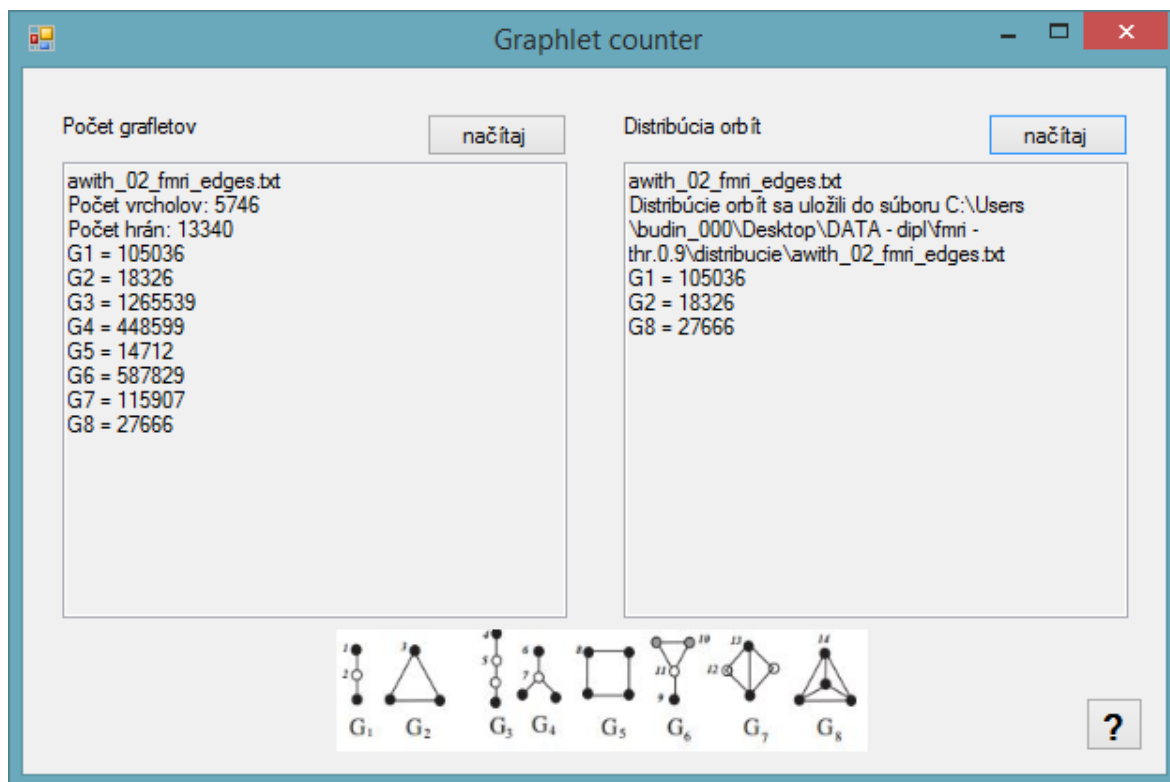
Implementáciu sme rozdelili na dve časti – prvá sa venuje hľadaniu grafletov v sieťach, druhú sme zamerali na analýzu sietí z hľadiska nami vypočítanej grafletovej štruktúry.

3.1. Hľadanie grafletov

Prvá časť implementácie bola naprogramovaná v jazyku C#, v prostredí Microsoft Visual C# 2010 Express. Program dokáže načítať sieť z textového súboru, ktorý v každom riadku obsahuje dvojicu čísel x, y oddelených medzerou. Tieto čísla predstavujú indexy vrcholov, medzi ktorými v sieti existuje hrana. Z takto načítaného grafu vie program vypočítať počet jednotlivých grafletov a vypíše ich. Rovnako vie zistiť distribúciu orbít grafletov stupňa tri a štvorkličky. Výsledok uloží do nového priečinku do súboru, ktorý sa volá rovnako ako vstupný súbor.

Graf je v programe ukladaný do slovníkovej štruktúry, kde kľúčom je číslo vrcholu a hodnotou je zoznam vrcholov, ktoré s ním susedia. Táto premenná má v programe názov *susednosti*.

Na Obr. 12 je zobrazené grafické používateľské rozhranie programu s ukázkou vypočítaných grafletov pre jednu sieť. Ovládanie programu je jednoduché – užívateľ načíta príslušnú sieť a následne sa mu zobrazí počet grafletov, ktoré sa v sieti hľadali. Pri zisťovaní počtu grafletov sa jednotlivé hodnoty uložia aj do schránky a dajú sa tak ľahko vložiť do iného programu alebo súboru. Pri zisťovaní distribúcie orbít grafletov sa súbor uloží do vytvoreného priečinku „*distribucie*“, ktorý sa nachádza v rovnakom umiestnení ako vstupný súbor.



Obr. 12 - Ukážka programu

3.1.1. Grafletové vzory

Marcus a Shavitt sa vo svojej práci RAGE [24] venujú distribúcii troj- a štvorvrcholových grafletov pre jednotlivé vrcholy siete. Ich algoritmy sú efektívne, pomerne rýchle⁸ a ponúkajú presné výsledky. Existuje viacero algoritmov na približné počítanie grafletov, ako napríklad GRAFT [25], cieľom našej práce bolo ale testovanie reálnych dát a ich porovnanie, z toho dôvodu sme použili algoritmy RAGE. Postup hľadania grafletov je nasledovný: najprv sa v sieti nájdu grafletové vzory – sú to podgrafy siete, ktoré majú rovnakú štruktúru ako nami zadané graflety, ale nemusia byť indukované. Následne, pomocou autormi zadaných vzťahov medzi počtami jednotlivých grafletových vzorov, vieme spočítať výskyt indukovaných podgrafov – t. j. grafletov.

Tieto algoritmy sú stavané na zisťovanie distribúcie grafletov pre jednotlivé orbity. Keďže sme v tejto časti implementácie potrebovali zistiť iba počty výskytov jednotlivých grafletov pre celý graf a nezaujímali nás počty dotykov vrcholov s orbitami, mnohé

⁸ V testovacej fáze tejto práce sme našli nový článok, v ktorom autori predstavujú ešte rýchlejšie algoritmy založené na paralelnom počítaní [29]

z algoritmov sme upravili a zjednodušili. V nasledujúcej časti tak opíšeme jednak RAGE algoritmy, ale aj niektoré naše úpravy a vlastné postupy.

Takmer všetky funkcie na hľadanie grafletov stupňa 3 a 4 opísané v [24] používajú funkciu *Merge* (Algoritmus 1):

```
private int merge(int v, int u, ref Dictionary<Int32, Int32> merged, out
List<int> list) {
    list = new List<int>();
    foreach (int w in susednosti.Keys) {
        merged[w] = 0;
    }
    foreach (int w in susednosti[u]) {
        if (w != v) {
            merged[w]=1;
        }
    }
    foreach (int w in susednosti[v]) {
        if (w != u) {
            if ((merged.ContainsKey(w)) && (merged[w] == 1)){
                merged[w] = 3;
                list.Add(w);
            }
            else {
                merged[w] = 2;
            }
        }
    }
    return list.Count;
}
```

Algoritmus 1 – Funkcia Merge, z [24]

Princíp funkcie *Merge* je založený na spájaní podľa hrán do slovníkovej štruktúry (*merged*), v ktorej kľúčmi sú indexy všetkých vrcholov grafu a hodnotou je číslo 0, 1, 2 alebo 3. Parametrami funkcie sú indexy dvoch vrcholov, *u* a *v*, referencia na už používaný *merged* a výstupný zoznam *list*, čo je zoznam všetkých spoločných susedov vstupných vrcholov, ktorý funkcia vráti. Ak majú vrcholy *u* a *v* spoločných susedov, v *merged* sa táto informácia uloží ako hodnota 3 a zároveň sa ich spoločný sused uloží do premennej *list*. Ak vrcholy susedia iba s prvým vstupným vrcholom, v premennej *merged* sa na ich indexoch nachádza číslo 2. Ak susedia s druhým vstupným vrcholom, je tam 1. Funkcia vracia výstupnú premennú *list* a na štandardnom výstupe vráti aj veľkosť tejto premennej, t. j. počet všetkých spoločných vrcholov vstupných vrcholov *u* a *v*.

Pri počítaní grafletových vzorov⁹ M_2 , alebo jednoducho povedané trojuholníkov, prejdeme všetky hrany siete a pre každú pustíme funkciu *merge*. Spočítame tak počet všetkých spoločných susedov vrcholov danej hrany. Výsledný počet trojuholníkov pre každý vrchol musíme predeliť dvomi, keďže v trojuholníku je každý vrchol spojený s dvomi hranami. [24] Pre naše účely – spočítania všetkých trojuholníkov v sieti – sa suma počtu trojuholníkov pre všetky vrcholy vydolí tromi, keďže každý trojuholník bol zarátaný trikrát - raz pre každý jeho vrchol.

Na počítanie grafletových vzorov M_1 , troj-ciest, sme použili vlastný algoritmus. Trojcestu vieme vytvoriť tak, že k nejakej hrane (u, v) pridáme ešte jednu hranu – buď z vrcholu u alebo z vrcholu v . Tento vzťah vieme zapísať ako:

$$\text{pre } \forall (u, v) \in G: M_1(v) = (|sus(u)| - 1) \wedge M_1(u) = (|sus(v)| - 1), \quad (37)$$

kde $M_1(v)$ označuje počet dotykov grafletového vzoru M_1 s vrcholom v a $sus(u)$ označuje množinu susedov vrcholu u . Takto vypočítané čiastkové počty trojciest pre jednotlivé vrcholy zosumujeme a následne predelíme dvoma, keďže každý grafletový vzor sme zarátali raz pre každú z dvoch hrán.

Algoritmus na hľadanie grafletového vzoru M_3 , štvorcesty, v sieti je v [24] rozšírený pre jednotlivé orbity. Na účely spočítania grafletov v celej sieti sme použili iba prvú časť, v pôvodnom algoritme použitú na počítanie výskytu orbity číslo 5 (viď. Obr. 10). V takto upravenom algoritme pre každú hranu (u, v) bude počet grafletových vzorov M_3 pre vrchol u (a rovnako aj pre vrchol v) vypočítaný podľa:

$$M_3(u) = |sus(u) - \{v\}| \cdot |sus(v) - \{u\}| - |merged(u, v)|, \quad (38)$$

kde $merged(u, v)$ označuje množinu všetkých spoločných susedov vrcholov u, v . Počet grafletových vzorov M_3 v celkom grafe bude potom suma grafletov incidentných s každým vrcholom siete delené dvoma, keďže každá štvorcesta má dva vrcholy orbity 5.

Grafletový vzor M_4 , inak nazývaný aj 4-hviezda, má dve orbity. Jedna z nich, orbita 7, patrí iba strednému vrcholu, t. j. vrcholu, z ktorého idú hrany do všetkých ďalších vrcholov

⁹ grafletové vzory budeme označovať ako M_x , pričom tým budeme myslieť podgraf s rovnakou štruktúrou ako graflet G_x na Obr. 10

grafletového vzoru. Na spočítanie počtu výskytov grafletového vzoru M_4 v sieti tak stačí zistiť počet výskytov orbity 7. Služi na to jednoduchý vzťah, opäť prevzatý z [24]:

$$M_5(v) = \binom{|sus(v)|}{3} = \frac{|sus(v)| \cdot (|sus(v)| - 1) \cdot (|sus(v)| - 2)}{6} \quad (39)$$

Takto vypočítané počty dotykov s grafletovým vzorom M_4 pre každý vrchol stačí iba zosumovať. Keďže v danom grafletovom vzore sa vyskytuje iba jeden vrchol orbity 7.

Štvoruholník, t.j. grafletový vzor M_5 , má všetky vrcholy v jednej orbite. Na ich spočítanie použijeme nasledovný algoritmus, kde $pocStvor1[v]$ je to isté ako $M_5(v)$:

```
foreach (int v in susednosti.Keys){
    foreach (int u in susednosti[v]) {
        if (v < u) {
            mer = merge(v, u, ref mergedL, out list);
            foreach (int w in susednosti[v]) {
                if (w != u) {
                    foreach (int t in susednosti[w]) {
                        if (mergedL.ContainsKey(t) &&
                            ((mergedL[t] == 1) || (mergedL[t] == 3))){
                            pocStvor1[v]++;
                            pocStvor1[u]++;
                        }
                    }
                }
            }
            foreach (int w in susednosti[u].Union(susednosti[v])) {
                mergedL[w] = 0;
            }
        }
    }
}
```

Algoritmus 2 - Hľadanie štvoruholníkov, podľa [24]

Každý štvoruholník zarátame dvakrát – pre oba smery, preto výsledný počet musíme predeliť dvoma.

Algoritmus na zistenie počtu grafletových vzorov M_6 , v [24] nazvaných ako trojuholníky s chvostom, sme opäť modifikovali, keďže v tomto grafletovom vzore sa nachádzajú až tri orbity. My sme si zvolili orbitu 11, čo je vrchol trojuholníka, z ktorého vychádza ešte jedna hrana (tzv. „chvost“). Práve kvôli dotyku s trojuholníkom, čo je grafletový vzor M_2 , platí nasledujúci vzťah:

$$M_7(v) = \max\{0, M_2(v) * (|sus(v)| - 2)\} \quad (40)$$

Od počtu všetkých susedov vrcholu v odčítujeme 2 preto, aby sme nezaráтали vrcholy nachádzajúce sa už v trojuholníku.

Grafletový vzor M_7 je štvoruholník s jednou uhlopriečkou. Dva vrcholy sú stupňa 2, zvyšné dva vrcholy majú zas stupeň 3. Upravíme algoritmus tak, aby počítal iba výskyt orbity 13 (čo sú vrcholy so stupňom 3). Pre každú hranu zistíme pomocou funkcie *merge* počet spoločných susedov oboch vrcholov hrany. Na štvoruholník potrebujeme ešte ďalšie dva vrcholy, tak ich vyberieme z množiny spoločných vrcholov (v rovnici (41) označené ako *merged(u,v)*). Keďže sa nemôžu opakovať a nezáleží nám na poradí, jedná sa o kombinácie bez opakovania, čo vieme spočítať ako:

$$\forall w \in \text{merged}(u, v): G_7(w) = \binom{|\text{merged}(u, v)|}{2} \quad (41)$$

Posledným grafletovým vzorom, ktorého výskyt sme počítali, je 4-klika, teda grafletový vzor M_8 . Všetky vrcholy 4-kliky majú rovnaký stupeň vrcholu a to 3. Patria preto do jednej orbity. Algoritmus je podobný tomu na hľadanie štvoruholníka, v jednom z vnútorných cyklov však prechádzame zoznam spoločných vrcholov, kým v štvoruholníku sú to ďalší susedia jedného zo vstupných vrcholov.

```
foreach(int v in susednosti.Keys){
    foreach (int u in susednosti[v]) {
        if (v < u) {
            mer = merge(v, u, ref mergedL, out list);
            foreach (int w in list) {
                foreach (int t in susednosti[w]) {
                    if ((w < t) && (mergedL.ContainsKey(t) && mergedL[t] == 3)){
                        pocKlik[v]++;
                        pocKlik[u]++;
                    }
                }
            }
            foreach (int w in susednosti[v].Union(susednosti[u])) {
                mergedL[w] = 0;
            }
        }
    }
}
```

Algoritmus 3 - Hľadanie štvorkliky, podľa [24]

3.1.2. Počet grafletov

Všetky algoritmy spomenuté v predchádzajúcej časti sa zameriavajú na počítanie výskytov nie nutne indukovaných grafletových motívov. V [24] sa nachádzajú aj vzťahy na výpočet počtu výskytov indukovaných motívov - t. j. grafletov. Keďže aj tieto vzťahy autori opisujú pre jednotlivé orbity a nie pre počet grafletov v celej sieti, pre naše účely sme mnohé, rovnako ako v predchádzajúcej časti opísané algoritmy, upravili.

Existujú dva trojvrcholové graflety. Trojuholníky, graflety G_2 , sú úplnými grafmi stupňa tri, preto počet indukovaných trojuholníkov je rovnaký, ako počet grafletových motívov.

Keďže v trojuholníku sa nachádzajú tri neindukované trojcesty, počet indukovaných grafletov G_1 , t. j. trojciest, pre graf S vypočítame ako $(G_x(S))$ označuje počet grafletov G_x v sieti S a $M_x(S)$ počet grafletových motívov v sieti S):

$$G_x(S) = M_1(S) - 3 \cdot M_2(S) \quad (42)$$

Pri grafletoch stupňa 4 je úplným grafom 4-klika, tj. graflet G_8 . $G_8(S)$ sa tak rovná počtu výskytu motívu M_8 v grafe. Pre všetky ostatné štvorvrcholové graflety musíme počet ich indukovaných výskytov vypočítať pomocou počtu iných grafletových motívov. Najjednoduchšie vypočítame počet indukovaných štvoruholníkov s jednou uhlopriečkou, čo je graflet G_7 , keďže v 4-klike vieme nájsť 6 grafletových motívov M_7 .

$$G_7(S) = M_7(S) - 6 \cdot M_8(S) \quad (43)$$

Pre graflet G_5 , štvoruholník, musíme okrem jeho výskytov v 4-klike (3 výskyty), spočítať aj počet výskytov v graflete G_7 – ten je jeden. Počet grafletov G_5 je potom:

$$G_5(S) = M_5(S) - G_7(S) - 3 \cdot M_8(S) \quad (44)$$

Čo môžeme upraviť podľa vzťahu (43) tak, aby sme počítali iba s grafletovými motívmi:

$$\begin{aligned} G_5(S) &= M_5(S) - (M_7(S) - 6 \cdot M_8(S)) - 3 \cdot M_8(S) \\ &= M_5(S) - M_7(S) + 3 \cdot M_8(S) \end{aligned} \quad (45)$$

Pre graflet G_6 , trojuholník s „chvostom“, rovnako musíme zistiť počet výskytov v štvoruholníku s jednou uhlopriečkou (graflet G_7) a štvorklike (graflet G_8). Pre graflet G_7 je to 4 – keďže zostane uhlopriečka a vyberieme jednu zo zvyšných hrán, ktorá nezostane, čo sú 4 možnosti. Pre graflet G_8 treba odobrať 2 hrany pre nejaký z vrcholov, čo je $4 \cdot \binom{3}{2} = 12$. Výsledný vzťah potom vyzerá:

$$G_6(S) = M_6(S) - 12 \cdot M_8(S) - 4 \cdot G_7(S) \quad (46)$$

Použitím vzťahu (43) dostávame:

$$\begin{aligned} G_6(S) &= M_6(S) - 12 \cdot M_8(S) - 4 \cdot (M_7(S) - 6 \cdot M_8(S)) \\ &= M_6(S) + 12 \cdot M_8(S) - 4 \cdot M_7(S) \end{aligned} \quad (47)$$

Štvorhviezda, G_4 , sa nachádza okrem štvorkliky aj v grafletoch G_6 a G_7 . V graflete G_8 ju môžeme nájsť v každom vrchole, teda sa v 4-klike vyskytuje 4-krát. V graflete G_7 , štvoruholníku s uhlopriečkou, je to pre každý vrchol stupňa tri, takže 2-krát. A v graflete G_6

je iba jeden vrchol stupňa 3, preto sa v ňom štvorhviezda nachádza iba raz. Dostávame tak vzťah:

$$G_4(S) = M_4(S) - 4 \cdot M_8(S) - 2 \cdot G_7(S) - G_6(S) \quad (48)$$

Použijeme vzťahy (43) a (47) a upravíme:

$$\begin{aligned} G_4(S) &= M_4(S) - 4 \cdot M_8(S) - 2 \cdot (M_7(S) - 6 \cdot M_8(S)) \\ &\quad - (M_6(S) + 12 \cdot M_8(S) - 4 \cdot M_7(S)) \\ &= M_4(S) + 2 \cdot M_7(S) - M_8(S) - M_6(S) \end{aligned} \quad (49)$$

Posledným grafletom, pre ktorý musíme vypočítať počet jeho indukovaných výskytov, je graflet G_3 , štvorcesta. Tú vieme nájsť v grafletoch G_5 až G_8 . V štvoruholníku (graflet G_5) stačí odstrániť 1 hranu, aby vznikla 4-cesta. Z toho vyplýva, že sa v ňom nachádza 4-krát. V graflete G_6 môžeme nájsť 4-cestu 2-krát. Pre graflet G_7 máme dve možnosti hľadania štvorcesty – ak odoberieme uhlopriečku, je počet štvorciet rovnaký ako v štvoruholníku, teda 4, ak uhlopriečku necháme, máme dve možnosti. Čo je dokopy 6 výskytov štvorcesty v graflete G_7 . A v štvorklike 4-cesta vznikne ako akákoľvek permutácia všetkých vrcholov, keďže sú spojené hranou, čo je $4!$, aby sme však nezarátali každý graflet dvakrát, musíme to predeliť dvoma, čo je teda 12.

$$G_3(S) = M_3(S) - 4 \cdot G_5(S) - 2 \cdot G_6(S) - 6 \cdot G_7(S) - 12 \cdot G_8(S) \quad (50)$$

Na úpravu použijeme vzťahy (43), (45) a (47):

$$\begin{aligned} G_3(S) &= M_3(S) - 4 \cdot (M_5(S) - M_7(S) + 3 \cdot M_8(S)) \\ &\quad - 2 \cdot (M_6(S) + 12 \cdot M_8(S) - 4 \cdot M_7(S)) \\ &\quad - 6 \cdot (M_7(S) - 6 \cdot M_8(S)) - 12 \cdot M_8(S) \\ &= M_3(S) - 12 \cdot M_8(S) + 6 \cdot M_7(S) - 2 \cdot M_6(S) - 4 \cdot M_5(S) \end{aligned} \quad (51)$$

3.1.3. Distribúcia orbít grafletov

Kvôli povahe sietí a pre pomerne veľkú pamäťovú náročnosť sme sa pri distribúcii orbít grafletov sústredili iba na graflety stupňa tri (trojuholník a trojcestu) a štvorkliku, ktorá pomerne dobre ukazuje rozdiely medzi reálnymi a umelo vytvorenými sieťami. Štvorklika sa napr. v náhodných sieťach a sieťach vytvorených podľa BA modelu nenachádza, alebo nachádza pomerne málo v porovnaní so zvyšnými grafletmi (v závislosti od nastavenia vstupných parametrov na vytvorenie BA modelu).

Pri grafletoch G_2 a G_8 , čo je trojuholník a štvorklika, je distribúcia pomerne jednoduchá, keďže obsahujú iba jednu orbitu. Algoritmus na zistenie dotykov vrcholu s danými grafletovými motívami sme opísali v predchádzajúcej časti a keďže oba graflety sú kompletne grafy daného stupňa, nemusíme sa zaoberať zisťovaním počtu indukovaných grafletov, lebo všetky sú indukované.

V graflete G_l sa nachádzajú dve orbity – jedna zahŕňa vonkajšie vrcholy (orbita 1) a druhá vnútorný vrchol (orbita 2), preto musíme pre každý vrchol zistiť počet dotykov s oboma orbitami. Ako sme už v predchádzajúcej časti opísali, počet dotykov s vonkajšou orbitou v grafletových motívoch zistíme podľa vzťahu (37). Aby sme pre každý vrchol siete zistili presný počet dotykov orbity 1 z grafletu G_l , musíme od počtu grafletových motívov odpočítať dvakrát počet trojuholníkov, keďže z jedného vrcholu vieme v trojuholníku vytvoriť dve trojcesty. Počet dotykov s orbitou 1 $O_1(v)$ potom môžeme zapísať ako:

$$O_1(v) = M_1(v) - 2G_2(v) \quad (52)$$

Na zistenie počtu dotykov vrcholu v s orbitou 2 využijeme kombinatoriku. Vrchol v môže byť stredným vrcholom trojcesty pre každú kombináciu dvoch jeho susedov. Aby sme získali počet dotykov s iba indukovanými grafletmi, musíme odpočítať počet dotykov s trojuholníkom, keďže takto vytvorená trojica vrcholov sa môže nachádzať maximálne v jednom trojuholníku. Dostávame tak vzťah:

$$O_2(v) = \binom{|sus(v)|}{2} - G_2(v) \quad (53)$$

Takto vypočítané distribúcie stupňov grafletov sa ukladajú do špeciálneho priečinku do textového súboru s rovnakým názvom, ako mal vstupný súbor. Textový súbor sa skladá z riadkov obsahujúcich čísla oddelené medzerou – 1 riadok je distribúcia jednej orbity v celej sieti, a každé číslo znamená počet dotykov vrcholu (podľa poradía) s danou orbitou.

3.2. Analýza sietí

Druhú časť, v ktorej sa venujeme porovnávaní sietí na základe počtu grafletov, alebo distribúcie stupňov grafletov, sme naprogramovali v jazyku GNU Octave [26], čo je voľná verzia programovacieho jazyka podobného Matlab-u. Octave sme si zvolili z dôvodu, že je vhodný na náročné výpočty, rýchlo a ľahko sa v ňom pracuje s maticami a vektormi, a dáta vieme jednoducho vykresliť do grafu ľubovoľného typu pomocou jeho súčasti Gnuplot [27].

Na analýzu sietí sme si zvolili metódy popísané v častiach 1.3.1 a 1.3.3, t. j. počet grafletov respektíve distribúciu stupňa grafletov. Analýzu sme rozdelili na štyri časti – základná analýza grafletov, relatívna vzdialenosť frekvencií grafletov v reálnych dátach, distribúcia stupňov grafletov v reálnych dátach a práca s umelými dátami.

Základná analýza spočíva v načítaní súborov s grafletmi (súbory *youngData.txt*, *oldData.txt*, *oldADData.txt*, ktoré sme získali pomocou výpočtov opísaných v predchádzajúcej časti) a následne je pre každú skupinu vytvorený graf, kde sú znázornené logaritmy počtu jednotlivých grafletov pre každú sieť. Priemerný počet jednotlivých grafletov je v grafe znázornený hrubšou prerušovanou čiarou. Rovnaké dáta je možné zobrazit' aj ako súbor čiarových grafov, pričom priemer sa zobrazuje so znakovým označením jednotlivých hodnôt. Ďalej skript vykreslí do samostatného grafu priemerné počty grafletov pre každú zo skupín.

Funkcie na počítanie relatívnej vzdialenosti frekvencií grafletov sa nachádzajú v súbore *relativneVzdialenosti.m*. Implementovali sme vzťahy (21) - (23), ktoré sa nachádzajú v jednotlivých funkciách. Relatívna vzdialenosť frekvencií grafletov sa najprv počíta v rámci skupín princípom každý s každým. Výsledky skript vykreslí do stĺpcového grafu, kde stĺpec značí jedno porovnanie a pre prehľadnosť sme ich usporiadali podľa veľkosti, priemerná relatívna vzdialenosť pre danú skupinu sa zobrazí v rovnakom grafe ako priamka. Skript taktiež vykreslí trojicu grafov, v ktorých sú znázornené porovnania medzi dvoma skupinami – ako stĺpce sa zobrazujú priemery relatívnej vzdialenosti grafletov siete z prvej skupiny so všetkými sieťami druhej skupiny, ako priamka sa zobrazuje priemer všetkých relatívnych vzdialenosti grafletov (princípom každý s každými medzi dvoma skupinami).

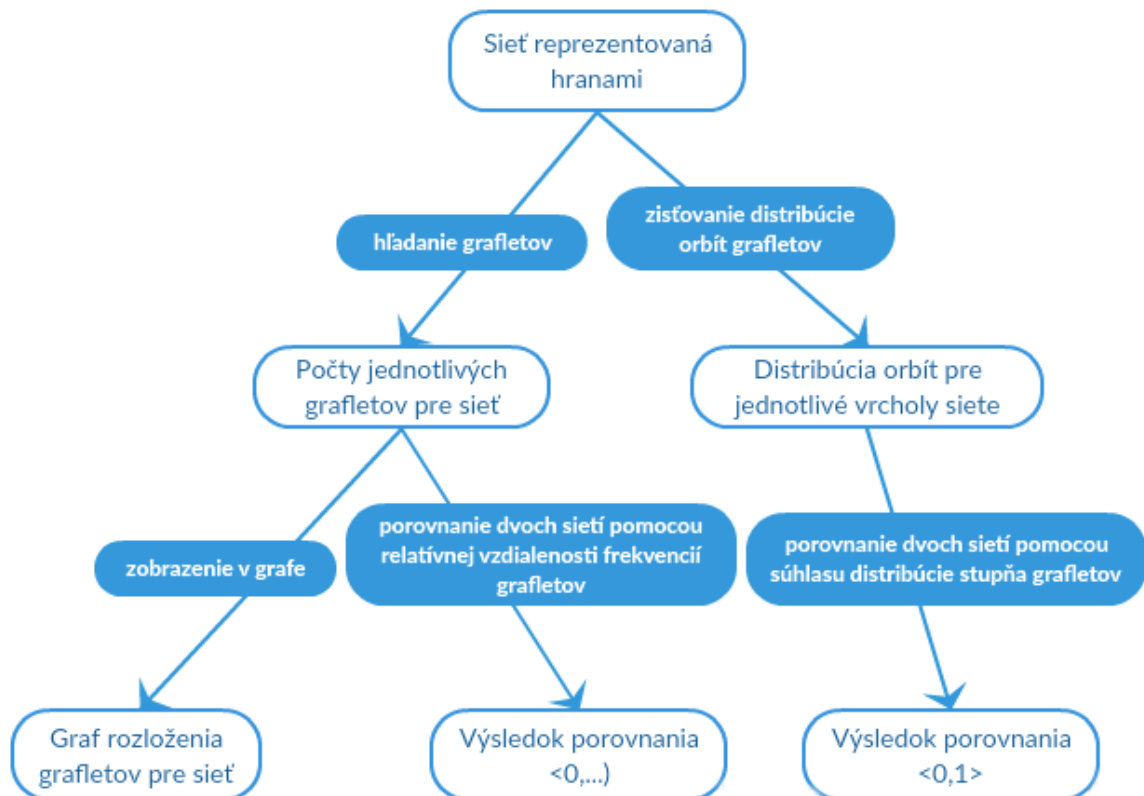
Súbor *distribucieStupnov.m* obsahuje implementované vzťahy (29) - (34). Vieme porovnávať dve siete medzi sebou, jednu sieť so zvolenou skupinou, siete v rámci jednej skupiny, alebo skupiny medzi sebou. Pri každom porovnaní vieme nastaviť, či chceme počítat' geometrický alebo aritmetický priemer všetkých j-tych GDD-súhlasov. Po spustení skriptu sa každú sieť zobrazia tri grafy v jednom okne – postupné porovnanie s jednotlivými skupinami. Grafy sú stĺpcové, na x-ovej osi sa nachádzajú indexy jednotlivých sietí porovnáwanej skupiny, a stĺpec znamená hodnotu súhlasu daných dvoch sietí. Názov porovnáwanej siete je v popise grafu.

Pri práci s umelými dátami používame iba mierne upravené funkcie z predchádzajúcich dvoch častí a pre obe metódy vieme porovnat' umelo vytvorenú sieť s jednotlivými reálnymi

sieťami, so skupinami sietí, ale aj vzájomne dve a viac umelo vytvorených sietí medzi sebou. Vykresľovanie do grafov zostáva rovnaké ako v predchádzajúcich častiach.

Presnejší návod na spúšťanie, ovládanie, prípadne úpravu vstupných parametrov je dostupný v súbore *readMe.txt*, ktorý je súčasťou prílohy na CD.

Obr. 13 zobrazuje celý postup, akým sme pracovali pri výskume. V modrých rámečkoch sa nachádzajú použité metódy, v bielo-modrých rámečkoch je vstup resp. výstup pre danú metódu.



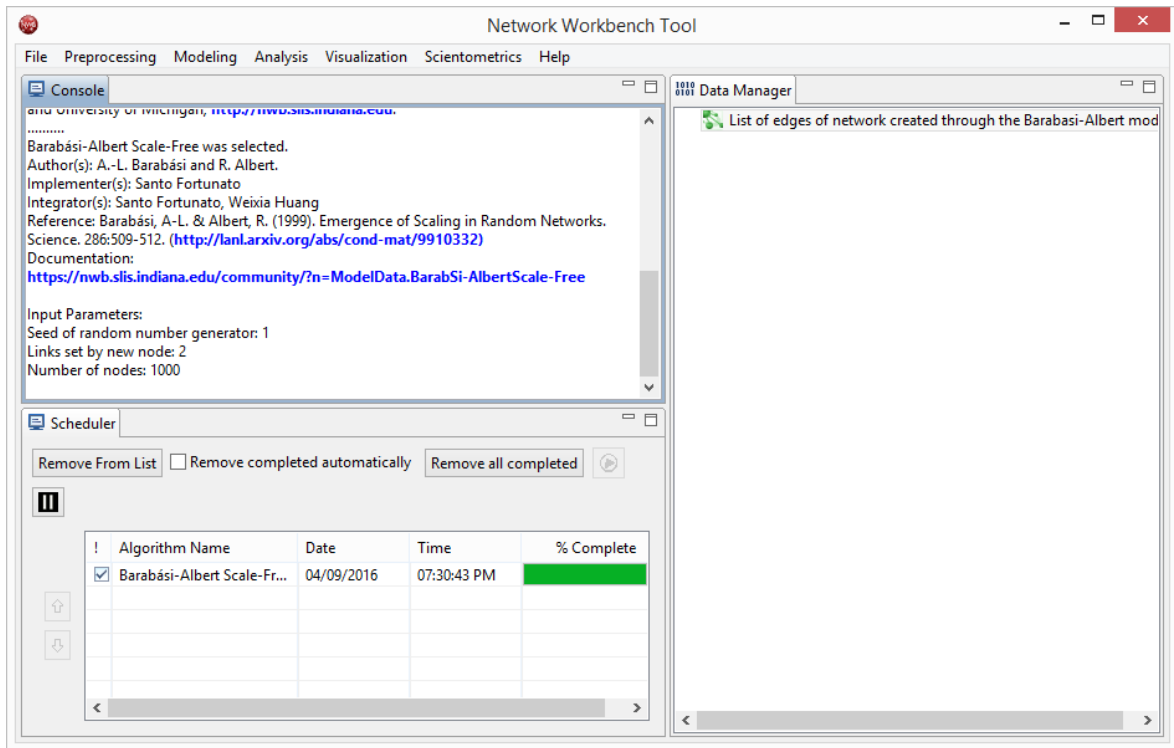
Obr. 13 - Diagram znázorňujúci postup výskumu

3.3. Umelo vytvorené dáta

Dôležitou súčasťou nášho výskumu bolo porovnávanie reálnych sietí s modelmi komplexných sietí. Na vytvorenie umelých dát, ktoré zodpovedajú jednotlivým modelom, sme využili nástroj Network Workbench [28]. Tento program umožňuje analyzovať, vizualizovať, ale aj vytvárať komplexné siete.

Hlavné prostredie Network Workbench sa skladá z troch častí (viď. Obr. 14). Napravo vidíme vygenerované dáta, s ktorými môžeme ďalej pracovať (vizualizovať ich, alebo exportovať). Naľavo sa nachádza konzola, v ktorej sú informácie o nami vykonávaných

úkonoch – ako sú vstupné parametre, ktoré sme zadali, alebo odkazy na články, v ktorých sú vysvetlené algoritmy, ktoré program použil na generovanie. Pod konzolou vidíme priebeh nami spúšťaných operácií.



Obr. 14 - Network Workbench

Pri modelovaní sietí podľa Barabási-Albert modelu môžeme v prostredí Network Workbench nastaviť počet vrcholov siete, počet vrcholov, ku ktorým sa má nový vrchol pripojiť a inicializáciu generátora pseudonáhodných čísel. Na modelovanie sietí podľa Wattz-Strogatz modelu musíme nastaviť počet vrcholov siete, počet počiatočných susedov, pravdepodobnosť, s ktorou sa budú hrany prepájať a opäť, inicializáciu generátora pseudonáhodných čísel. Pri modelovaní náhodných grafov nastavujeme počet vrcholov, pravdepodobnosť pripájania vrcholov a inicializáciu generátora pseudonáhodných čísel.

Vygenerované siete môžeme vyexportovať do niekoľkých formátov, my sme pre kompatibilitu s našim softvérom zvolili klasický model – súbor s vrcholmi a súbor s hranami, pričom pre nás dôležité boli iba súbory obsahujúce popis hrán, ktoré sme následne mohli spracovať pomocou našich programov.

4. Výsledky

V poslednej kapitole bližšie popíšeme, zanalyzujeme a vyhodnotíme výsledky našej práce. Kvôli prehľadnosti je rozdelená na tri podkapitoly, v ktorých sa budeme postupne venovať počtom grafletov vo funkčných sieťach mozgu, následne použijeme prvú z metód porovnávania sietí – relatívnu vzdialenosť frekvencií grafletov a porovnáme jednak funkčné siete mozgu medzi sebou, a taktiež aj reálne dáta s umelo vytvorenými sieťami. Posledná časť sa venuje distribúcií stupňov grafletov, opäť aj v rámci skupiny reálnych dát, aj v rámci porovnania s modelovými sieťami.

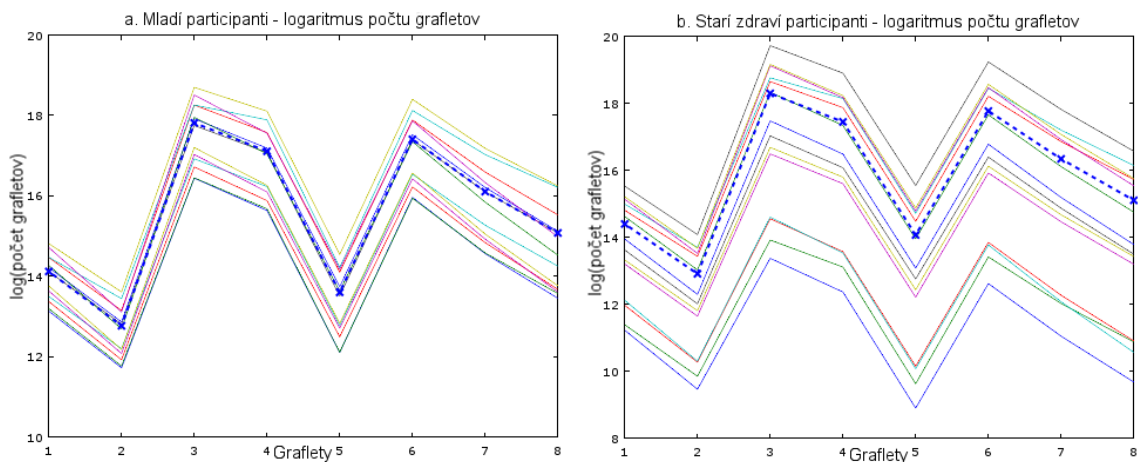
4.1. Počty grafletov vo funkčných sieťach mozgu

Prvá časť nášho výskumu sa zaoberala grafletmi v reálnych funkčných sieťach mozgu. Najskôr sme pre každú z vybraných sietí (popísané v 2.2.) zistili pomocou nášho programu počty grafletov. Tieto údaje sme následne uložili do troch súborov podľa jednotlivých skupín a ďalej s nimi pracovali pomocou skriptov vytvorených v GNU-Octave. Prvotne nás zaujímala štruktúra rozloženia jednotlivých grafletov, preto sme pre každú skupinu vygenerovali graf. Počty grafletov sme zobrazovali v logaritmickej mierke pre lepšiu prehľadnosť.

Graf 1.a znázorňuje počty jednotlivých grafletov pre funkčné siete mozgu mladých participantov. Na osi x sú indexy grafletov, na osi y logaritmy počtu daných grafletov. Počty grafletov pre jednotlivé siete sú znázornené ako farebne odlišené lomené čiary. Priemerný počet jednotlivých grafletov je znázornený hrubšou modrou prerušovanou lomenou čiarou. Počty grafletov pre jednotlivé siete kopírujú rovnaké rozloženie, rozdiely sú pomerne malé. V sieťach sa nachádza najmenej grafletov G_2 a G_5 , naopak obsahujú najviac grafletov G_3 , G_4 a G_6 .

Graf 1.b znázorňuje počty grafletov pre skupinu starých zdravých participantov. Os x opäť reprezentuje indexy grafletov a na osi y sa nachádza logaritmus ich výskytu v sieti. V rámci skupiny starých zdravých participantov môžeme vidieť väčšie rozdiely medzi jednotlivými sieťami. Graflety G_3 , G_4 a G_6 sú tu rovnako ako v skupine mladých participantov zastúpené v najvyššom počte, graflety G_2 a G_5 sa tiež v tejto skupine vyskytujú najmenej. Pre niektoré siete je ale počet grafletov G_2 a G_5 porovnateľný s počtom výskytu grafletu G_8 .

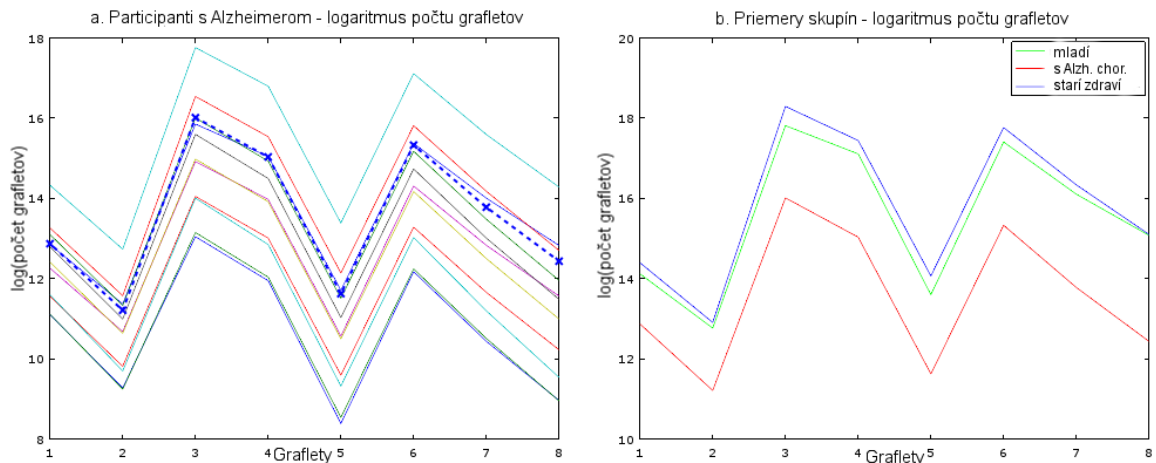
Keď porovnáme skupinu mladých zdravých a skupinu starých zdravých participantov, väčšina sietí má približne rovnaké distribúcie počtu grafletov, okrem štyroch sietí zo skupiny starých zdravých participantov. Môžeme to vidieť už pri počte grafletov G_1 . Kým pre väčšinu sietí zo skupiny starých zdravých participantov platí, že logaritmus výskytu grafletu G_1 má hodnotu medzi 13,2 a 15,6, pre tieto štyri siete platí, že $\log(G_1)$ má hodnotu menšiu ako 12,2.



Graf 1- Distribúcia počtu grafletov – (a) pre mladých participantov, (b) pre starých zdravých participantov

Pri skupine participantov s Alzheimerovou chorobou sú väčšie rozdiely medzi jednotlivými sieťami, vid'. Graf 2.a. Celkovo platí, že počty jednotlivých grafletov sú v rámci tejto skupiny menšie, ako pri ostatných skupinách. Napr. logaritmus výskytu grafletu G_1 má hodnotu v rozmedzí 11 až 14,3 (s tým, že iba jedna sieť mala túto hodnotu vyššiu ako 13,2). Priemer je na grafe opäť znázornený modrou prerušovanou lomenou čiarou.

Priemery jednotlivých skupín sme zobrazili v spoločnom grafe, Graf 2.b. Kým skupiny mladých zdravých participantov a starých zdravých participantov majú priemerné počty grafletov (znázornené v logaritmickej mierke) veľmi podobné, v skupine funkčných sietí mozgu participantov s Alzheimerovou chorobou sa graflety priemerne vyskytujú v menšom počte.



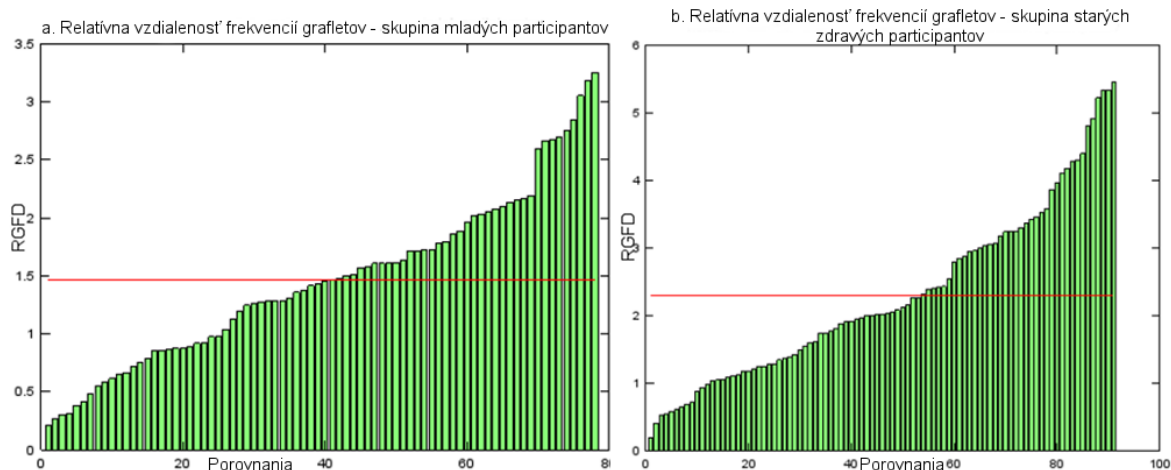
Graf 2 – (a) Logaritmus počtu grafletov pre participantov s Alzheimerovou chorobou, (b) priemery počtu grafletov pre jednotlivé skupiny

4.2. Relatívna vzdialenosť frekvencií grafletov

Relatívna vzdialenosť frekvencií grafletov (RGFD) určuje podobnosť dvoch sietí z hľadiska grafletovej štruktúry s využitím počtu grafletov v sieti. Čím je vzdialenosť menšia, tým viac sú si porovnávané siete bližšie (viac v kapitole 1.3.1.). Podobnosti funkčných sietí mozgu sme zisťovali jednak v rámci jednotlivých skupín, medzi skupinami a následne s umelo vytvorenými modelovými sieťami.

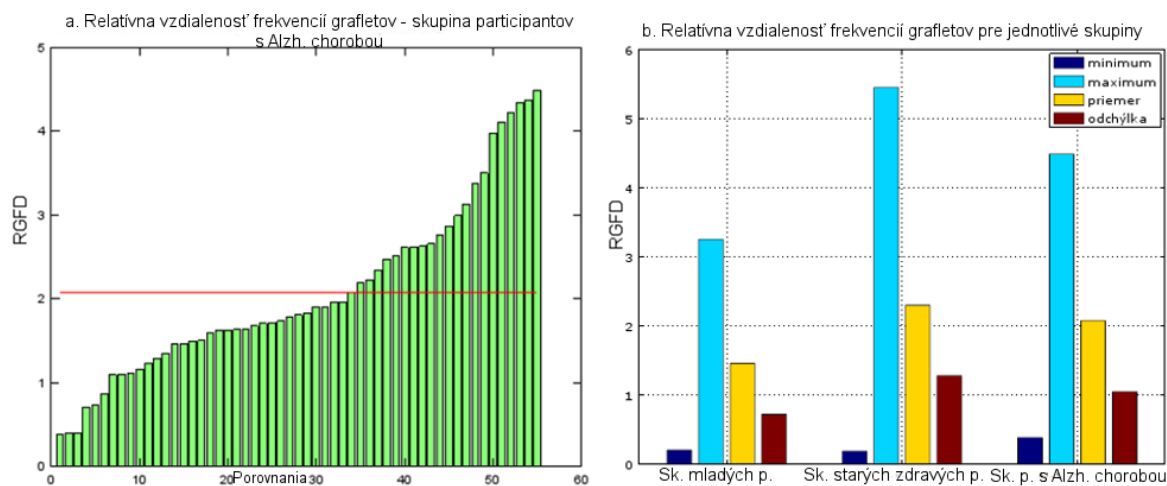
Relatívne vzdialenosti frekvencií grafletov v rámci skupiny mladých participantov znázorňuje Graf 3.a. Podobnosti sietí sme počítali spôsobom každý s každým a keďže v tejto skupine sme mali k dispozícii funkčné siete mozgu 13 participantov, porovnaní bolo 78. Hodnoty sme usporiadali vzostupne. Hodnoty podobnosti dvoch sietí sa nachádzajú na osi y. Priemerná hodnota relatívnej vzdialenosti frekvencií grafletov medzi sieťami tejto skupiny (v grafe znázornená ako červená priamka) je 1,46 so štandardnou odchýlkou 0,73.

V skupine funkčných sietí starých zdravých participantov, ktorých bolo 14, sme dostali 91 porovnaní. Výsledky zobrazuje Graf 3.b. Priemerná podobnosť dvoch sietí v rámci tejto skupiny (v grafe znázornená ako červená priamka) je 2,30 so štandardnou odchýlkou 1,28. Priemerná hodnota je takmer o jedna väčšia ako priemer podobností v rámci skupiny funkčných sietí mozgu mladých participantov.



Graf 3 - Relatívna vzdialenosť frekvencií grafletov - (a) v rámci skupiny mladých participantov, (b) v rámci skupiny starých zdravých participantov

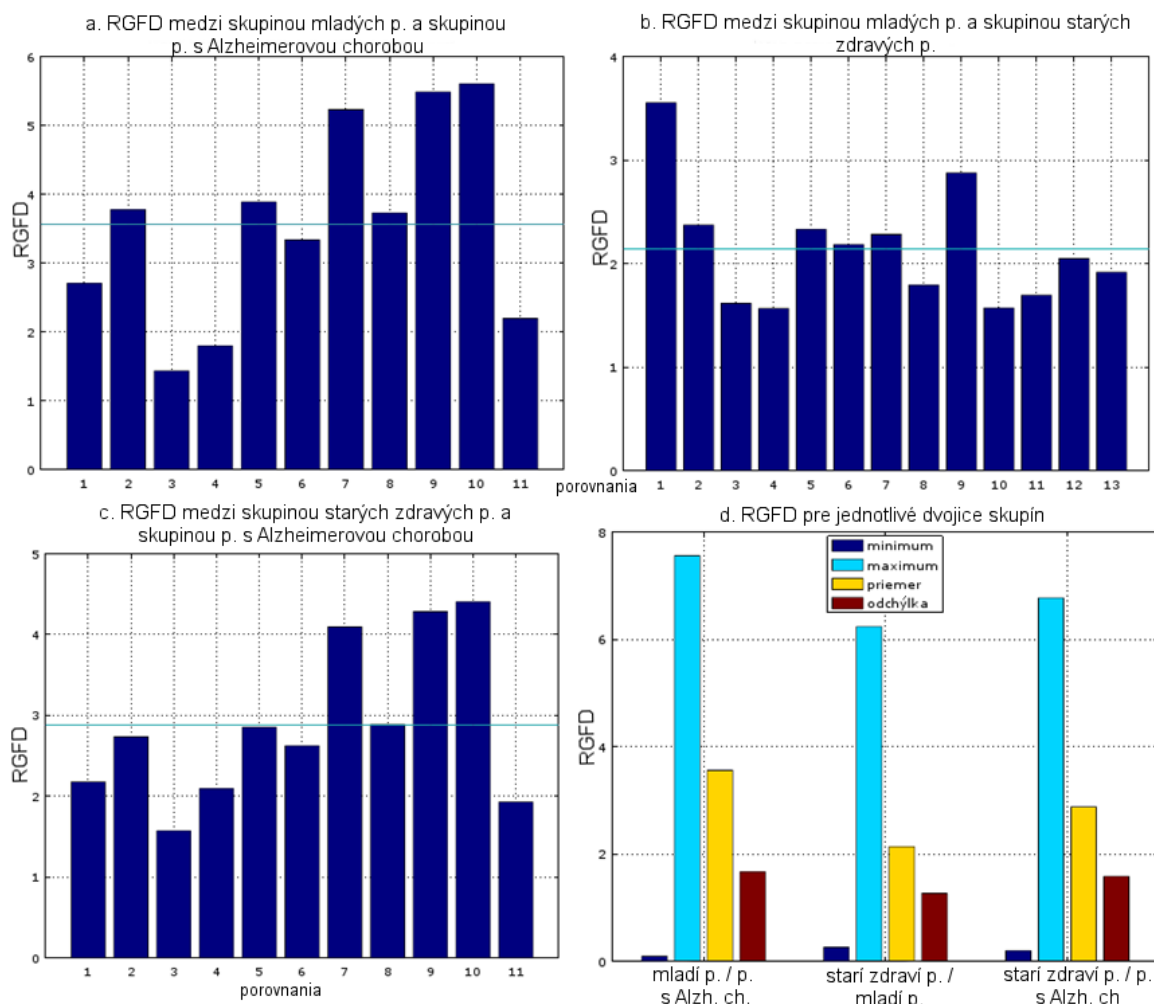
Graf 4.a obsahuje hodnoty podobností sietí v rámci skupiny funkčných sietí mozgu participantov s Alzheimerovou chorobou. Priemerná hodnota relatívnej vzdialenosti frekvencií grafletov (opäť znázornená červenou priamkou) je 2,08 so štandardnou odchýlkou 1,05, čo je veľmi podobný výsledok ako v skupine starých zdravých participantov. Celkové porovnanie pre relatívne vzdialenosti frekvencií grafletov v rámci jednotlivých skupín znázorňuje Graf 4.b.



Graf 4 - Relatívna vzdialenosť frekvencií grafletov - (a) pre skupinu participantov s Alzheimerovou chorobou, (b) súhrnné výsledky pre jednotlivé skupiny

Z tohto grafu vyplýva, že najviac sú si podobné siete v skupine mladých participantov a najmenej zas v skupine starých zdravých participantov. Veľké rozdiely v rámci skupiny starých zdravých participantov boli spôsobené štvoricou sietí, ktoré majú menší počet grafletov ako zvyšok skupiny (viď. Graf 1.b), keďže táto metóda odlišuje siete práve na základe počtu grafletov.

Následne sme skúmali relatívnu vzdialenosť frekvencií grafletov medzi skupinami. Porovnávali sme vždy dve skupiny. Pre každú sieť z jednej skupiny sme vypočítali relatívnu vzdialenosť frekvencií grafletov so všetkými sieťami druhej skupiny a v grafe zobrazujeme priemer týchto porovnaní. Na osi x sa tak nachádza index siete z jednej skupiny a hodnota prislúchajúceho stĺpca je priemer porovnaní tejto siete so všetkými z druhej skupiny.



Graf 5 - Relatívne vzdialenosti frekvencií grafletov – (a) medzi skupinou mladých participantov a skupinou participantov s Alzheimerovou chorobou, (b) medzi skupinou mladých participantov a skupinou starých zdravých participantov, (c) medzi skupinou starých zdravých participantov a skupinou participantov s Alzheimerovou chorobou, (d) súhrnné výsledky pre jednotlivé porovnania

Najpodobnejšie siete vyšli medzi skupinami mladých a starých zdravých participantov. Priemerná hodnota podobnosti siete v týchto skupinách vyšla 2,14 so štandardnou odchýlkou 1,27 (Graf 5.b). Priemerná hodnota podobnosti sietí zo skupiny starých zdravých participantov so skupinou participantov s Alzheimerovou chorobou je 2,88 s odchýlkou 1,58 (Graf 5.c). Pri porovnaní skupiny participantov s Alzheimerovou chorobou a skupiny mladých participantov sme dostali priemernú relatívnu vzdialenosť frekvencií grafletov rovnú 3,56 so štandardnou odchýlkou 1,67, z čoho vyplýva, že najväčšie rozdiely boli práve

medzi týmito dvoma skupinami (Graf 5.a). Graf 5 zobrazuje všetky výsledky, v podgrafe d sú zobrazené význačné hodnoty pre všetky porovnania medzi skupinami.

Na základe týchto výsledkov môžeme skonštatovať, že existujú rozdiely medzi sieťami z jednotlivých skupín z hľadiska počtu grafletov, ale nevieme pomocou tejto metódy zaradiť sieť do skupiny, alebo dostatočne odlíšiť dve siete z rôznych skupín. Priemerné výsledky síce ukazujú určité väčšie alebo menšie rozdiely v rámci jednotlivých porovnaní, ale pre niektoré konkrétne dvojice sietí (aj z rôznych skupín) vyšli pomerne malé hodnoty relatívnej vzdialenosti frekvencií grafletov.

Rovnaké porovnania sme spravili pre jednotlivé skupiny s umelo vytvorenými sieťami. Porovnávali sme s dvoma typmi sietí – BA modelom a WS modelom, keďže, ako sme v časti 1.4. spomínali, v niekoľkých štúdiách sa potvrdilo, že funkčné siete mozgu sa správajú ako siete malého sveta respektíve majú podobné vlastnosti ako siete BA modelu. Všetky nami vytvorené siete, ktoré v nasledujúcich častiach porovnáваме s funkčnými sieťami mozgu, mali priemerne 6 000 vrcholov, čím sme chceli dosiahnuť čo najväčšiu podobnosť s reálnymi dátami. Skúšali sme, samozrejme, aj iné vstupné parametre a siete, uvádzame však iba skupiny, v ktorých bola nájdená nejaká podobnosť s reálnymi dátami. Výsledky pre vybrané siete modelované podľa Barabási-Albert modelu obsahuje Tabuľka 4.

| Porovnanie so skupinou: | mladých participantov | starých zdravých participantov | participantov s Alzheimerovou chorobou | priemer |
|-------------------------|-----------------------|--------------------------------|--|---------|
| BA model – k = 2 | 27,877 | 28,090 | 28,253 | 28,073 |
| BA model – k = 3 | 31,998 | 31,758 | 31,518 | 31,758 |
| BA model – k = 4 | 28,491 | 28,268 | 28,067 | 28,275 |
| BA model – k = 5 | 26,718 | 26,463 | 26,264 | 26,482 |
| BA model – k = 6 | 25,345 | 25,199 | 25,211 | 25,252 |
| BA model – k = 9 | 22,745 | 22,697 | 23,008 | 22,817 |
| BA model – k = 10 | 22,240 | 22,131 | 22,421 | 22,264 |
| BA model – k = 12 | 21,389 | 21,284 | 21,643 | 21,439 |

Tabuľka 4 - Relatívna vzdialenosť frekvencií grafletov pre porovnanie BA modelov s jednotlivými skupinami funkčných sietí mozgu (k – počet vrcholov, ku ktorým sa pripojí nový vrchol)

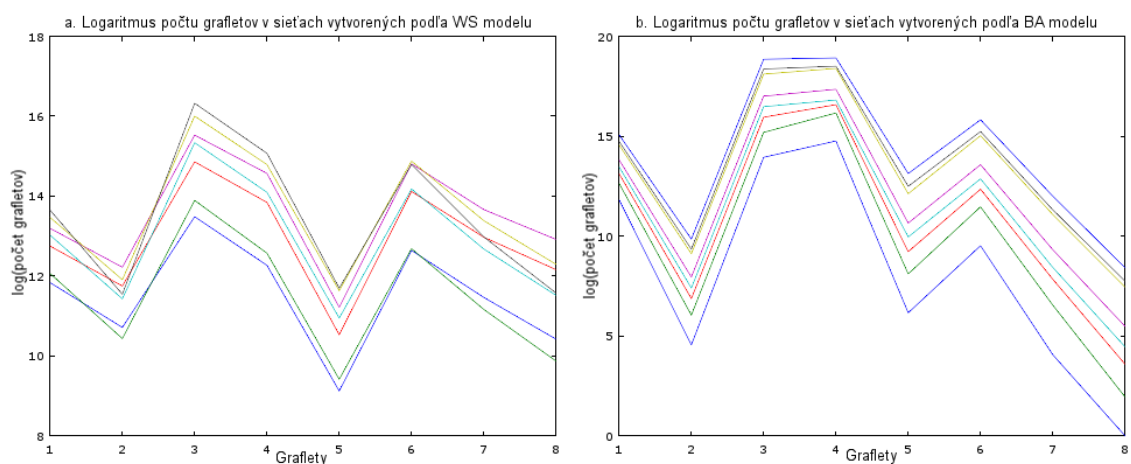
V tabuľke vidíme, že hodnoty relatívnych vzdialeností frekvencií grafletov pre BA modely a funkčné siete mozgu sú pomerne vysoké a medzi jednotlivými skupinami nie sú význačné rozdiely. So zvyšujúcim sa k hodnoty RGF D mierne klesajú, preto sme skúšali vytvárať siete, v ktorých sa nové vrcholy pripájali k väčšiemu počtu existujúcich vrcholov v grafe, ale takto vytvorené siete mali príliš mnoho hrán a bolo časovo a pamäťovo náročné zistiť počty jednotlivých grafletov.

Porovnanie reálnych sietí so sieťami malého sveta dopadlo lepšie. Výsledky zachytáva Tabuľka 5. V tabuľke vidíme, že siete vytvorené podľa WS modelu majú najbližšie k funkčným sieťam mozgu participantov s Alzheimerovou chorobou, dve najnižšie hodnoty sú zvýraznené. Všetky siete mali 6000 vrcholov. V priemere dopadlo najlepšie porovnanie so sieťou, ktorá vznikla z grafu, v ktorom mal každý vrchol 10 susedov a hrany sa prelinkovali s pravdepodobnosťou $p = 0,2$.

| Porovnanie so skupinou: | mladých participantov | starých zdravých participantov | participantov s Alzheimerovou chorobou | Priemer |
|------------------------------|-----------------------|--------------------------------|--|---------|
| WS model – $k = 5, p = 0,1$ | 5,5595 | 5,0436 | 4,1638 | 4,9223 |
| WS model – $k = 5, p = 0,2$ | 6,3591 | 5,1641 | 3,2521 | 4,9251 |
| WS model – $k = 8, p = 0,1$ | 4,5043 | 4,3035 | 3,6200 | 4,1426 |
| WS model – $k = 8, p = 0,2$ | 5,0956 | 3,8916 | 2,3944 | 3,7939 |
| WS model – $k = 10, p = 0,1$ | 4,0560 | 3,9250 | 3,3569 | 3,7793 |
| WS model – $k = 10, p = 0,2$ | 4,4888 | 3,3788 | 2,1287 | 3,3321 |
| WS model – $k = 10, p = 0,3$ | 5,6652 | 4,7202 | 3,0316 | 4,4723 |

Tabuľka 5 – Relatívne vzdialenosti frekvencií modelových sietí s jednotlivými skupinami funkčných sietí mozgu (k – pôvodný počet susedov v WS sieti, p – pravdepodobnosť prelinkovania)

Podobné hodnoty relatívnej vzdialenosti frekvencií grafletov dosahovali aj porovnania sietí jednotlivých skupín medzi sebou, z čoho môžeme usúdiť, že z hľadiska počtu grafletov s tromi a štyrmi vrcholmi, majú funkčné siete mozgu podobnú štruktúru ako siete malého sveta. Dokazuje to aj rozloženie počtu grafletov pre jednotlivé skupiny umelo vytvorených dát, ktoré zachytáva Graf 6.a. Rozloženie počtu jednotlivých grafletov je podobné ako vo funkčných sieťach mozgu (Graf 1 a Graf 2), kým rozloženie počtu grafletov v sieťach vytvorených podľa BA modelu je od nich veľmi odlišné (vid'. Graf 6.b).



Graf 6 - Logaritmus počtu grafletov v sieťach vytvorených podľa - (a) WS modelu, (b) BA modelu (V prvej sieti (najnižšie v grafe) vytvorenej podľa BA modelu bol počet grafletov G_8 rovný 0, pre lepšie zobrazenie je v grafe zobrazený ako $\log(0)$)

4.3. Distribúcia stupňov grafletov

Druhá časť nášho výskumu sa zaoberala distribúciou stupňov grafletov. Sledovali sme všetky orbity grafletov stupňa 3 a orbitu prislúchajúcu grafletu G_8 , t. j. 4-klike. Pomocou tejto distribúcie sme počítali podobnosti sietí – a to aj ako aritmetický, aj geometrický priemer distribúcií stupňov grafletov (GDD) pre všetky orbity (viď. kapitola 1.3.3). Podobnosť sietí pomocou distribúcie stupňa grafletov môže dosahovať hodnoty z intervalu $(0,1)$, kde 1 znamená, že siete majú rovnakú distribúciu a 0 značí úplnú odlišnosť sietí. Na rozdiel od predchádzajúcej metódy na zisťovanie podobnosti sietí, pri tejto nie je tak dôležitý počet grafletov, ale ako už značí jej názov, rovnaká distribúcia vrcholov, ktoré sa dotýkajú jednotlivých orbít.

Najprv sme počítali podobnosti sietí v rámci jednotlivých skupín. Pre aritmetický priemer GDD to bolo v rámci skupiny participantov s Alzheimerovou chorobou 0,922 so štandardnou odchýlkou 0,045, v skupine starých zdravých participantov mali jednotlivé dvojice sietí podobnosť 0,927 so štandardnou odchýlkou 0,052 a v poslednej skupine mladých participantov dosiahla podobnosť sietí číslo 0,920 s odchýlkou 0,046. Vidíme, že podobnosti dosahovali veľmi podobné hodnoty a pre geometrický priemer distribúcií stupňov grafletov boli výsledky na približne rovnakej úrovni ako pre aritmetický priemer. Všetky hodnoty obsahuje Tabuľka 6.

| skupina | minimum | | maximum | | priemer | | štandardná odchýlka | |
|------------------------------------|---------|--------|---------|--------|---------|--------|---------------------|--------|
| | AP | GP | AP | GP | AP | GP | AP | GP |
| Participant s Alzheimerovou ch. | 0,8552 | 0,8542 | 0,9808 | 0,9808 | 0,9217 | 0,9204 | 0,0450 | 0,0461 |
| Starí zdraví participant | 0,8087 | 0,8067 | 0,9811 | 0,9811 | 0,9273 | 0,9265 | 0,0522 | 0,0530 |
| Mladí participant | 0,8392 | 0,8373 | 0,9858 | 0,9858 | 0,9201 | 0,9185 | 0,0458 | 0,0470 |

Tabuľka 6 - Distribúcia stupňov grafletov - porovnanie v rámci jednotlivých skupín (AP – aritmetický priemer GDD súhlasov, GP – geometrický priemer GDD súhlasov)

Rovnakým spôsobom ako v predchádzajúcej časti sme skúšali zisťovať podobnosti sietí medzi dvoma skupinami. Výsledky kopírovali podobnosti v rámci jednej skupiny. Z tohto dôvodu táto metóda nie je vhodná na odlišenie dvoch funkčných sietí mozgu. Môžeme však skonštatovať, že dokáže pomerne presne určiť, či je daná sieť funkčnou sieťou mozgu.

Ďalej nás preto zaujímalo, či táto metóda dokáže odlišiť funkčné siete mozgu a modelové siete – či sa už bude jednať o WS model, alebo BA model. Kým takéto porovnanie

pre predchádzajúcu metódu, relatívnu vzdialenosť frekvencií grafletov, ukázalo, že funkčné siete mozgu (a to hlavne pre participantov s Alzheimerovou chorobou) sú podobné sieťam malého sveta, pri distribúcii stupňov grafletov boli výsledky iné. Siete malého sveta boli s funkčnými sieťami mozgu podobné iba na približne 30% (Tabuľka 8), ale naopak, siete vytvorené podľa BA modelu mali už približne 60 percentnú podobnosť (Tabuľka 7).

| Porovnanie so skupinou: | participantov s Alzheimerovou chorobou | | starých zdravých participantov | | mladých participantov | | priemer | |
|-------------------------|--|--------|--------------------------------|--------|-----------------------|--------|---------------|--------|
| | AP | GP | AP | GP | AP | GP | AP | GP |
| BA model - k = 3 | 0,5611 | 0,4666 | 0,5705 | 0,4786 | 0,5660 | 0,4541 | 0,5659 | 0,5123 |
| BA model - k = 4 | 0,5923 | 0,4847 | 0,5981 | 0,4950 | 0,5896 | 0,4670 | 0,5933 | 0,4822 |
| BA model - k = 5 | 0,5822 | 0,5055 | 0,5915 | 0,5171 | 0,5875 | 0,4958 | 0,5871 | 0,5061 |
| BA model - k = 6 | 0,5996 | 0,5344 | 0,6093 | 0,5465 | 0,6047 | 0,5263 | 0,6045 | 0,5357 |
| BA model - k = 9 | 0,6574 | 0,6018 | 0,6690 | 0,6159 | 0,6575 | 0,5917 | 0,6613 | 0,6031 |
| BA model - k = 10 | 0,6553 | 0,6042 | 0,6686 | 0,6193 | 0,6508 | 0,5915 | 0,6582 | 0,6050 |
| BA model - k = 12 | 0,6648 | 0,6173 | 0,6770 | 0,6316 | 0,6515 | 0,5983 | 0,6644 | 0,6157 |

Tabuľka 7 – Porovnanie sietí vytvorených podľa BA modelu s funkčnými sieťami mozgu pomocou distribúcie stupňa grafletov (k – počet vrcholov, ku ktorým sa pripojí v sieti nový vrchol, AP – aritmetický priemer GDD súhlasov sietí, GP – geometrický priemer GDD súhlasov sietí)

| Porovnanie so skupinou: | participantov s Alzheimerovou chorobou | | starých zdravých participantov | | mladých participantov | | priemer | |
|----------------------------|--|--------|--------------------------------|--------|-----------------------|--------|---------------|--------|
| | AP | GP | AP | GP | AP | GP | AP | GP |
| WS model – k = 5, p = 0,1 | 0,3677 | 0,3618 | 0,3782 | 0,3707 | 0,3442 | 0,3372 | 0,3634 | 0,3566 |
| WS model – k = 5, p = 0,2 | 0,4111 | 0,4051 | 0,4217 | 0,4151 | 0,3865 | 0,3793 | 0,4064 | 0,3998 |
| WS model – k = 8, p = 0,1 | 0,3789 | 0,3724 | 0,3898 | 0,3814 | 0,3553 | 0,3478 | 0,3747 | 0,3672 |
| WS model – k = 8, p = 0,2 | 0,3819 | 0,3752 | 0,3929 | 0,3843 | 0,3582 | 0,3504 | 0,3776 | 0,3700 |
| WS model – k = 10, p = 0,1 | 0,3836 | 0,3770 | 0,3945 | 0,3861 | 0,3598 | 0,3523 | 0,3793 | 0,3718 |
| WS model – k = 10, p = 0,2 | 0,3850 | 0,3783 | 0,3960 | 0,3874 | 0,3612 | 0,3535 | 0,3807 | 0,3731 |
| WS model – k = 10, p = 0,3 | 0,3891 | 0,3829 | 0,4001 | 0,3922 | 0,3652 | 0,3580 | 0,3848 | 0,3777 |

Tabuľka 8 - Porovnanie sietí vytvorených podľa WS modelu s funkčnými sieťami mozgu pomocou distribúcie stupňa grafletov (k – pôvodný počet susedov v WS sieti, p – pravdepodobnosť prelinkovania, AP – aritmetický priemer GDD súhlasov sietí, GP – geometrický priemer GDD súhlasov sietí)

Rozdiely medzi aritmetickým a geometrickým priemerom distribúcií stupňa grafletov neboli veľké, pre všetky siete však vyšli lepšie porovnania s použitím aritmetického priemeru. Zaujímavosťou je, že tentoraz ako najpodobnejšia skupina dát vyšla skupina funkčných sietí starých zdravých participantov. Z týchto výsledkov môžeme skonštatovať, že funkčné siete mozgu majú distribúciu stupňov grafletov viac podobnú BA modelu, ako sieťam malého sveta, hoci z hľadiska počtu grafletov v celej sieti je podobnosť opačná.

Táto podobnosť môže byť spôsobená fenoménom „bohatý sa stáva bohatším“, ktorý opisujú siete BA modelu. Keďže pravdepodobnosť, že vrchol siete je súčasťou nejakého grafletu, sa zvyšuje s jeho stupňom, v týchto sieťach existujú vrcholy, ktoré obsahujú veľké množstvo dotykov s rôznymi orbitami, a zároveň existuje mnoho vrcholov bez dotykov alebo s minimálnym počtom dotykov s grafletmi. Pri sieťach WS modelu je rozloženie jednotlivých grafletov pre vrcholy siete viac rovnomerné. Podobnosť funkčných sietí mozgu so sieťami vytvorenými podľa BA modelu tak ukazuje, že aj funkčné siete mozgu obsahujú veľa vrcholov s vysokým počtom dotykov a zároveň vrcholy, v ktorých dotyky s grafletmi neexistujú.

Bolo by zaujímavé sledovať podobnosti pre ďalšie vstupné parametre BA a WS modelov, keďže v tabuľkách vidíme, že podobnosť sietí stúpa pre zväčšujúce sa zložitosti modelových sietí. Je preto možné, pre dostatočne veľké umelé siete by sa funkčné siete mozgu podobali distribúciou stupňa grafletov viac aj sieťam malého sveta. Tieto otázky nechávame otvorené pre ďalší výskum.

Záver

Naša diplomová práca sa zaoberala teóriou sietí a metódami využívajúcimi graflety. Teoretické východiská sme zhrnuli v prvej kapitole, kde sme okrem základov teórie grafov a teórie sietí bližšie opísali niektoré z metód využívajúcich grafletovú štruktúru na porovnanie sietí a taktiež sme sa venovali funkčným sieťam mozgu. Druhá kapitola sa venovala cieľom nášho výskumu s čím súvisel aj opis reálnych dát, ktoré sme mali k dispozícii – funkčných sietí mozgu mladých zdravých participantov, starých zdravých participantov a participantov s Alzheimerovou chorobou.

Cieľom práce bolo bližšie skúmať funkčné siete mozgu z hľadiska ich grafletovej štruktúry. Implementovali a otestovali sme dve metódy na porovnanie sietí využívajúce graflety a to relatívnu vzdialenosť frekvencií grafletov a distribúciu stupňov grafletov, s čím súvisela aj implementácia algoritmov na hľadanie grafletov v sieťach. Zamerali sme sa na graflety s tromi a štyrmi vrcholmi a pri druhej metóde sme skúmali distribúciu orbít grafletov pre trojcestu, trojuholník a štvorkliku. Implementáciu sme opísali v tretej kapitole.

Posledná kapitola sa zaoberala výsledkami výskumu. Zistili sme, že relatívna vzdialenosť frekvencií grafletov dokáže nájsť rozdiely medzi jednotlivými skupinami funkčných sietí mozgu. Rovnako sa nám pomocou tejto metódy podarilo nájsť podobnosť s umelo vytvorenými sieťami, a to so sieťami malého sveta. Najviac boli siete WS modelu podobné s funkčnými sieťami mozgu participantov s Alzheimerovou chorobou a hodnoty podobnosti vyšli porovnateľné ako pri porovnávaní jednotlivých skupín funkčných sietí mozgu navzájom.

Distribúcia stupňov grafletov nedokázala zistiť väčšie rozdiely medzi jednotlivými skupinami, ale dokázala presne určiť štruktúru funkčných sietí mozgu. Podobnosti jednotlivých sietí dosahovali hodnoty okolo 90 %. Následné porovnanie reálnych dát s umelo vytvorenými sieťami ukázalo viac ako 60 % podobnosť so sieťami Barabási-Albert modelu, ale podobnosť so sieťami malého sveta bola už len okolo 30 %.

Podarilo sa nám splniť stanovené ciele. Problémy, s ktorými sme sa stretli, sa týkali veľkosti reálnych dát. Hoci sme skúmali dáta s najvyšším korelačným prahom, takže v nami analyzovaných funkčných sieťach mozgu bolo viac zmysluplných spojení a menej šumu, čo znamená aj menej hrán, algoritmy na zisťovanie počtu grafletov a následne podobností sietí z hľadiska grafletovej štruktúry boli pomerne časovo a pamäťovo náročné. S tým súvisí aj

fakt, že sme museli znížiť počet skúmaných grafletov pre jednotlivé metódy. Do ďalšej práce by preto bolo užitočné skúmať aj graflety s 5 vrcholmi, prípadne, pre metódu využívajúcu orbity grafletov, pozorovať aj distribúciu stupňov orbít pre ďalšie štvorvrcholové graflety.

Tieto problémy by mohlo vyriešiť implementovanie nových, rýchlejších, algoritmov, s ktorými sme sa zoznámili až v testovacej časti našej práce [29], prípadne využiť paralelné výpočty na viacerých strojoch, čím by sa zväčšil výpočtový výkon. Samozrejmosťou je testovanie na väčšej skupine reálnych funkčných sietí mozgu, ktoré by mohlo potvrdiť nami zistené podobnosti a rozdiely jednak v rámci jednotlivých skupín a zároveň so sieťami vytvorenými podľa BA a WS modelu.

Napriek týmto problémom považujeme našu prácu za úspešnú, keďže sme sa nestretli s výskumami, v ktorých by sa autori venovali grafletovej štruktúre funkčných sietí mozgu. Naša analýza ukázala, že tento prístup k funkčným sieťam mozgu môže odhaliť existujúce rozdiely medzi rôznymi skupinami participantov a hlavne, môže pomôcť pri ich ďalšom skúmaní a hľadaní modelu, ktorý by popísal ich dynamiku.

Literatúra a internetové zdroje

- [1] M. P. van den Heuvel, C. J. Stam a R. S. H. P. H. E. Kahn, „Efficiency of Functional Brain Networks and Intellectual Performance,“ *The Journal of Neuroscience*, zv. 29, %1. vyd.23, pp. 7619-7624, 10 Jún 2009.
- [2] V. Calhoun, T. Eichele a G. Pearlson, „Functional brain networks in schizophrenia: a review,“ *Frontiers in human neuroscience*, zv. 3, %1. vyd.17, 2009.
- [3] K. Superkar, M. Musen a V. Menon, „Development of Large-Scale Functional Brain Networks in Children,“ *Plos Biology*, zv. 7, %1. vyd.7, 2009.
- [4] D. Tomasi a N. D. Volkow, „Aging and functional brain networks,“ *Molecular Psychiatry*, %1. vyd.17, pp. 549-558, 2012.
- [5] N. Pržulj, „Biological network comparison using graphlet degree,“ *Bioinformatics*, zv. 23, p. e177–e183, 2006.
- [6] D. Reinhard, *Graph Theory*, New York: Springer-Verlag, 2000.
- [7] V. Kvasnička a J. Pospíchal, *Algebra a diskrétna matematika*, Bratislava: Slovenská technická univerzita, 2008.
- [8] P. Náther, *Complex Networks: dynamics and applications (Dizertačná práca)*, Bratislava: Univerzita Komenského, 2010.
- [9] P. Erdős a A. Rényi, „On random graphs I,“ rev. *Publicationes Mathematicae*, Budapest, 1959, pp. 290 - 297.
- [10] B. Bollobas, *Random Graphs*, Cambridge University Press, 2001.
- [11] S. Milgram, „The Small-World Problem,“ *Psychology Today*, zv. 1, %1. vyd.1, pp. 61-67, Máj 1967.
- [12] J. D. Watts a S. Strogatz, „Collective dynamics of 'small-world' networks,“ *Nature*, zv. 393, pp. 440-442, Jún 1998.
- [13] V. Bret, „Collective dynamics of 'small-world' networks (Watts, J. Duncan; Strogatz, S.),“ [Online]. Dostupné na: <http://worrydream.com/ScientificCommunicationAsSequentialArt/ScientificCommunicationAsSequentialArt.pdf>. [Cit. 5. december 2015].
- [14] A.-L. Barabási a R. Albert, „Emergence of Scaling in Random Network,“ *Science*, %1. vyd.286, pp. 509-512, 1999.
- [15] M. Markošová, „Dynamika sietí,“ rev. *Umelá inteligencia a kognitívna veda II*, Bratislava, Slovenská technická univerzita v Bratislave, 2010, pp. 321-379.
- [16] N. Pržulj a e. al, „Modeling interactome: scale-free or geometric?,“ *Bioinformatics*, zv. 20, p. 3508–3515, 2004.

- [17] R. Milošević et al., „Network Motifs: Simple Building Blocks of Complex Networks,“ *Science*, zv. 296, pp. 824-827, 2002.
- [18] T. Milenković a N. Pržulj, „Uncovering Biological Network Function via Graphlet Degree Distribution,“ *Cancer Informatics*, pp. 257-273, Jún 2008.
- [19] M. Faloutsos, P. Faloutsos a C. Faloutsos, „On Power-Law Relationships of the Internet Topology,“ rev. *ACM SIGCOMM'99*, Pittsburgh, 1999.
- [20] D. R. Chialvo, „Critical brain networks,“ *Physica A: Statistical Mechanics and its Applications*, zv. 4, %1. vyd.340, pp. 756-765, 2004.
- [21] Nuffield Department of Clinical Neurosciences, University of Oxford, „Introduction to fMRI,“ 2016. [Online]. Dostupné na: <http://www.ndcn.ox.ac.uk/divisions/fmrib/what-is-fmri/introduction-to-fmri>. [Cit. 29. Február 2016].
- [22] M. van den Heuvel, C. Stam, M. Boersma a H. P. H.E., „Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain,“ *Neuroimage*, zv. 43, %1. vyd.3, pp. 528-539, 15. November 2008.
- [23] V. M. Eguíluz, D. R. Chialvo, G. A. Cecchi, M. Baliki a A. V. Apkarian, „Scale-Free Brain Functional Networks,“ *Physical Review Letters*, zv. 94, %1. vyd.1, p. 018102, 2005.
- [24] D. Marcus a Y. Shavitt, „RAGE - A rapid graphlet enumerator for large networks,“ *Computer Networks*, zv. 2, %1. vyd.56, pp. 810-819, 2. Február 2011.
- [25] M. Rahman, M. Bhuiyan a M. Al Hasan, „Graft: An Approximate Graphlet Counting Algorithm for large graph analysis,“ rev. *Proceedings of the 21st ACM international conference on Information and knowledge management*, Maui, Hawaii, USA, 2012.
- [26] J. W. Eaton, „GNU Octave,“ 1998-2016. [Online]. Dostupné na: <https://www.gnu.org/software/octave/>. [Cit. 3. Január 2016].
- [27] T. Williams a C. Kelley, „Gnuplot,“ [Online]. Dostupné na: <http://www.gnuplot.info/>. [Cit. 13. December 2015].
- [28] NWB Team, „Network Workbench Tool,“ Indiana University, Northeastern University, and University of Michigan, 2006. [Online]. Dostupné na: <http://nwb.slis.indiana.edu/>. [Cit. 25. Január 2016].
- [29] N. K. Ahmed, J. Neville, R. A. Rossi a N. Duffield, „Efficient Graphlet Counting for Large Networks,“ 2015. [Online]. Dostupné na: <http://nickduffield.net/download/papers/ICDM2015-CameraReady.pdf>. [Cit. 25. Marec 2016].

Prílohy

Príloha je v digitálnej podobe a obsahuje:

- Vstupné dáta
- Zdrojové kódy aplikácie a skripty použité pri výskume
- Spustiteľnú aplikáciu na hľadanie grafletov
- Súbory obsahujúce vypočítané počty grafletov a distribúcie orbít pre jednotlivé vstupné dáta