Eötvös Loránd Tudományegyetem

Természettudományi kar

Agygráfok gráfelméleti analízise

Szakdolgozat

Írta: Gimes Balázs

Alkalmazott matematikus MSc

Számítástudomány szakirány

Témavezető:

Grolmusz Vince

egyetemi tanár

Számítógéptudományi Tanszék

Budapest, 2016

Tartalomjegyzék

Köszönetnyilvánítás **2**

1. Bevezetés **3**

1.1. Konnektomika 3

1.2. Konnektómok készítése 4

2. Vizsgálatok **10**

2.1. Első vizsgálat 11

2.2. Második vizsgálat 17

2.3. Harmadik vizsgálat 22

3. Következtetések **25**

Ábrák jegyzéke **27**

Táblázatok jegyzéke **27**

Hivatkozások **29**

Köszönetnyilvánítás

Szeretnék köszönetet mondani témavezetőmnek, Grolmusz Vincének a szakdolgozat elkészítésében nyújtott segítségéért. A konzultációkon felmerült ötletek, észrevételek és a munka gondos felügyelete nélkül ez a dolgozat nem készülhetett volna el.

1. fejezet

Bevezetés

1.1. Konnektomika

Az agygráfok kutatása -a szakirodalomban elterjedt elnevezés szerint *konnektomika, connectomics*- az utóbbi években egyre növekvő népszerűségnek örvend. Agygráfnak, illetve konnektómnak nevezzük egy élőlény agyának kapcsolati hálózatát. Ha a legalacsonyabb szinten tekintjük a kapcsolatrendszert, akkor világos, hogy mi lesz a gráf: csúcsai az egyes idegsejteknek felelnek meg, élet pedig akkor húzunk be két csúcs közé, ha a megfelelő két idegsejt között van szinapszis. A tudományterület "Szent Grálja" az emberi agy ilyen szintű feltérképezése (lásd 2.1 ábra).

A jelenlegi helyzet viszont ettől az ideális állapottól messze van: mindössze egyetlen élőlénynek, a *Caenorhabditis elegans* féregnek létezik ilyen mélységben feltárt konnektómja, neki 302 idegsejtje, és körülbelül 5000 szinapszisa van. Az embernek ezzel szemben becslések szerint százmilliárd idegsejtje, és százbillió szinapszisa van. Ennek feltérképezése jelenleg létező eszközökkel lehetetlen, továbbá már a féreg esetében is a konnektóm megléte pusztán egy kezdő lépés az agyműködés megértése felé, nem önmagában álló magyarázat.

Az ember esetében tehát nem ezen a szinten vizsgáljuk a konnektómot, hanem makro-szinten. Csúcsnak valamilyen agyterület felel meg. Az élek alapján két fő típust különíthetünk el: *strukturális konnektóm*nak nevezzük azokat a konnektómokat, melyekben az élek a két agyterület közötti fizikailag meglévő

kapcsolatot jelentik. A *funkcionális konnektóm* élei ezzel szemben azt jelentik, hogy két összekötött agyterület együtt aktivizálódik valamilyen tevékenység közben.

Egy értelmes csúcshalmaz lehet az anatómiailag meghatározott nagyobb agyterületek halmaza, ez jellemzően 60-120 agyterületet jelent. Ez egy viszonylag alátámasztott választás, hiszen meghatározott funkcióval rendelkező, lényegében egységként elkülöníthető részek az anatómiai agyterületek. Azonban foglalkoznak ennél finomabb felbontású konnektómokkal is, ekkor már kevésbé egyértelmű, hogy milyen egy jó csúcsválasztás, több különböző elven működő algoritmus állít elő MRI-vel készült képek alapján csúcsokat, a felső határ a csúcsok számára körülbelül 1000. Jelenleg még nincs általánosan elfogadott legjobb módszer, a különböző módszereknek mind vannak előnyei és hátrányai is.

Egy ideális makro-konnektóm olyan lenne, ami egybefoglalja a strukturális és funkcionális információkat, és irányultsági adatokat is tartalmazna (honnan hova terjed az információ az agyban) - ez az *effektív konnektóm*. Ez egy irányított éleket tartalmazó konnektóm lenne, ami azon felül, hogy pontosabb modellje lenne a valódi agytevékenységnek, bizonyos hálózati elemző módszereket is a jelenleginél hatékonyabbá tenne. Ilyen konnektóm nagyon kevés létezik jelenleg. A konstruálás fő nehézségét az okozza, hogy oksági relációt nehéz idősorból kiolvasni.

1.2. Konnektómok készítése

Ebben a szakaszban ismertetem, hogy milyen folyamat során állítják elő a konnektómokat, röviden bemutatva az egyes állomásokat is.

Adatgyűjtés - MRI

A konnektómok előállításának első lépése a nyers adatok gyűjtése az agyról, ami elsősorban MRI-vel történik. Az MRI (nuclear magnetic resonance imaging, mágneses magrezonancia képalkotás) egy nem-invazív, és ismereteink

szerint káros hatással nem járó eljárás, mellyel jó minőségű képek állíthatók elő az agyról. Fizikai alapja a (nevében is szereplő) magrezonancia jelensége. A vizsgálat során a pácienst befektetik az MRI-készülékbe, mely erős homogén mágneses mezőt generál (tipikusan szupravezető mágnest használnak). Az emberi testben található hidrogén-atommagok (protonok) mágneses momentumukból következően a generált mágneses tér iránya mentén rendeződnek, és az irány, mint tengely körül precesszálnak. Két állapotban lehetnek, vagy a tér irányába állnak, vagy azzal ellentétesen. Egy pici többségben lesznek azok a magok, melyek irányban állnak (ez alacsonyabb energiájú állapot). Ha megfelelő hullámhosszú rádióhullámmal sugározzák be a testet (a gyakorlatban ezt vékony rétegenként teszik), akkor az eredő mágneses vektor kibillen, és amikor ez az impulzus leállítása után fokozatosan visszaáll, a protonok visszasugározzák az energiát, ezt méri az MRI rendszere, és ezeknek az információknak alapján állítja elő a képet.

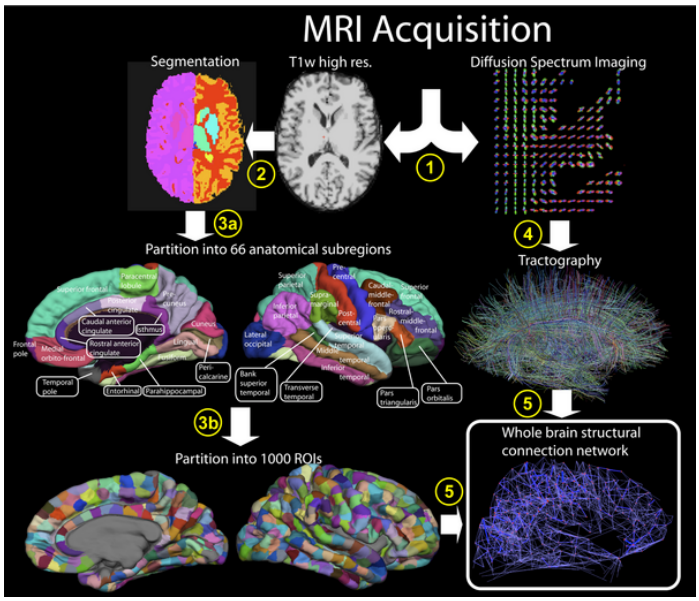
A *diffúziós MRI*, dMRI egy olyan MRI-eljárás, mellyel a strukturális konnektómokat állítják elő. Az alapvető mechanizmus, melyet kihasznál, a vízmolekulák diffúziója az agyban. Az agyban az axonok kötegekben fordulnak elő, a köteg pedig szerkezetéből adódóan erősen gátolja a vízmolekulák oldalirányú diffúzióját: adott idő alatt nagyobb elmozdulás történik a köteg mentén, mint oldalirányokban. Ezt tudja mérni az MRI-készülék, ezzel meg lehet becsülni, hogy az agyban hol és milyen irányban találhatóak axon-lefutások.

A *funkcionális MRI*, fMRI a funkcionális konnektómok előállításában játszik szerepet. Az fMRI vizsgálat során a páciens vagy nyugalomban fekszik (ez a resting state fMRI), vagy megkérik, hogy végezzen ismételten valami egyszerű feladatot (érintse össze két ujját, lassan számoljon magában 10-ig, stb: ez a task-fMRI). Az fMRI alapját képező mechanizmus a hemodinamikus válasz. Régóta megfigyelték, hogy az agyi aktivitás és a vérellátás között összefüggés van: ha egy agyterület aktív, akkor oda oxigéndús vér áramlik. Ezt a következő módon használja ki az MRI: az oxigénnel ellátott és oxigén nélküli hemoglobin koncentrációja megváltozik az aktív agyterületen (illetve annak közelében). Az oxigént hordozó, illetve nem hordozó hemoglobin mágneses tulajdonságai

eltérőek, ebből következően az oxigéndús illetve oxigénben szegény vér mágneses tulajdonságai is, a koncentrációváltozás tehát mérhető változást okoz a mágneses jelben, amit az MRI-készülék képes detektálni.

A továbbiakban a strukturális és funkcionális konnektómok MRI-felvételekből való előállításának főbb lépéseit ismertetem. Kisebb, ám szintén fontos fázisokról (pl. zajszűrés, fej mozgásából eredő hibák korrigálása, stb.) részletesen nem ejtek szót, mivel a folyamat elvi megértéséhez nem elengedhetetlenül szükséges a tárgyalásuk.

Strukturális konnektómok



1.1. ábra. A képfeldolgozás lépései. Forrás: [1] cikk

A képfeldolgozás első fő lépése az agyi felvételek *szegmentálása* fehér- és szürkeállományra. A fehérállomány a központi idegrendszer egyik fő alkotórésze, elsősorban gliasejtekből és mielinált axonokból áll, melyek agykérgi területeket, illetve agykérgi- és alacsonyabb szintű agyi központokat köt össze. A szürkeállomány a másik fő alkotórész, főként az idegsejtek sejttesteiből, axonokból, gliasejtekből és kapillárisokból áll, az agykérget, illetve bizonyos mélyebben fekvő, körülhatárolt agyi magvakat alkotja. A szegmentálásnak kettős célja van: egyfelől a szürkeállományt akarjuk régiókra osztani, hogy legyenek

csúcsaink a gráfhoz, másfelől meg kell becsülnünk a fehérállományban az axonok lefutását, ebből kapjuk majd az éleket.

A második lépés az agykéreg (szürkeállomány) *parcellációja*, részekre osztása. A nagyobb részekre osztás jellemzően valamely digitális atlasz alapján történik (atlasztól függően ez tipikusan 60-120 területet jelent). Az így kapott területek kisebbekre való felosztására több módszer is lehetséges, az egyik módszer vezérelve az, hogy a kis területek lényegében egyforma méretűek legyenek (ez kis terület esetén azt jelenti, hogy egy terület körülbelül 1,5 cm^2 nagyságú lesz). Ennek hátránya, hogy nem vesz figyelembe bizonyos információkat, melyet a felvételek tartalmaznak. Másik lehetőség a klaszterezés, valamilyen megadott tartományhatáron belül (pl. a kapott tartományok száma legyen 400 és 500 között) hozzon létre kis klasztereket, ezzel több képi információt tartunk meg, mintha megkövetelnénk egy egzakt számot a generálandó tartományok számára. Ennek hátránya lehet, hogy csoportos vizsgálat során a különböző személyekhez tartozó gráfoknak eltérő csúcsszáma lesz, aminek következtében az összehasonlítás nehezebbé válik.

A harmadik lépés a dMRI által előállított diffúziós térkép (minden voxelhez, azaz térbeli pixelhez tartoznak a diffúzióra vonatkozó információk) felhasználásával az axonok 3D-s lefutásának minél pontosabb megbecsülése, ezt az eljárást *traktográfiának* nevezik. A konkrét diffúziós technológiától is függ, hogy milyen pontos becslést tudunk végezni, fejlettebb módszerekkel azt is lehet észlelni, ha egy voxelen belül két axon-köteg metszi egymást, illetve több fő diffúziós irányt is meg lehet határozni. A traktográfia lehet determinisztikus vagy sztochasztikus. A traktográfia során az axonokról nyert térbeli elhelyezkedésre vonatkozó információnak a gráfokra vonatkoztatva nincsen jelentősége, ez a lépés azért fontos, hogy a területek közötti kapcsolatok erősségét tudjuk meghatározni.
Egy lehetséges traktográfia: minden voxelre meghatározzák, hogy mely irányokban maximális a diffúzió. Ezt követően minden voxelből minden meghatározott irányban véletlenszerűen választott pozíciókból elindítanak bizonyos számú (pl. 30) szálat. A szálak ezután szomszédos voxelbe lépnek, és akkor

folytatódnak, ha a most aktuális voxelben van olyan irány, hogy arrafelé folytatva egy megadott szögnél kisebb hajlásszögű lesz a keletkező törött-vonal. Akkor is megáll a szál folytatása, ha mindkét irányban elérte a fehérállomány határát.

Utolsó lépésben az előző lépések eredményeit felhasználva végül el tudjuk készíteni a strukturális konnektómot. A második lépésben keletkezett régiók lesznek a csúcsok, u és v régiókat akkor kötjük össze egy $e(u, v)$ éllel, ha köztük legalább egy (vagy küszöb megadása esetén valamely kis számú), a traktográfia során kapott szál található. Fel kell tüntetnünk azt is a gráfban, hogy milyen erősek a kapcsolatok (mennyi szálat találtunk két terület között), vagyis súlyoznunk kell az éleket. Az anatómiai felosztáskor keletkező területek különböző méretűek, ezzel a különbséggel is szokás normálni az élsúlyokat.

Funkcionális konnektómok

A funkcionális konnektómok készítésénél tipikusan fMRI-felvételekből indulnak ki, a cél az, hogy a régiók közötti funkcionális kapcsolatokat térképezzék föl. Ehhez a strukturális konnektómokhoz hasonlóan szükség van csúcsok, és élek létrehozására. A csúcsok jellemzően itt is agyterületeknek felelnek meg, az élek pedig két agyterület együttes aktivitásának a mértékét fejezik ki.

A csúcsok létrehozásához használható egy ismert anatómiai atlasz, vagy a régiók definiálhatók közvetlenül az fMRI-felvételek adataiból. A pontos módszer használata és a létrehozott csúcsok száma függ attól, hogy milyen vizsgálatot szeretnénk végezni, jelenleg nincs olyan kitüntetett eljárás, amit minden esetben előnyben kellene részesítenünk.

Az élek létrehozásához először meg kell határozni minden egyes csúcshoz egy, az adott csúcsot jellemző idősort. Ez megtehető például úgy, hogy az adott régióhoz tartozó voxelek idősorait átlagoljuk. Ezután meg kell határozni a területeknél kiszámolt jelek közötti statisztikai összefüggést, amely megtehető a korrelációs mátrix kiszámításával, de ennél jobb módszernek tekintik az inverz-kovariancia mátrix meghatározását, mert az figyelembe veszi a parciális korrelációkat is a változók között. A funkcionális konnektómokról és azok

előállításáról a [5] cikkben találhatóak további részletek.

A konnektóm elemzése

A képfeldolgozás során nyert konnektómokból szeretnénk nem triviális gráfelméleti információkat kinyerni, melyek segítségével megállapításokat tehetünk az agyról, mint hálózatról. A törekvések közé tartozik annak megállapítása, hogy milyen hatékonyan képes terjedni információ az agyban, milyen mértékben összefüggő az egész hálózat; milyen kisebb részhálózatok alkotják a teljes hálózatot, mik ezeknek az elemei; továbbá fontos csúcsok keresése, azaz olyan területek azonosítása, melyek feltehetően valamilyen fontos, akár központi szerepet játszanak információ továbbításában, illetve integrálásában. Fontos csúcs lehet olyan, ami egy részhálózaton belül tölt be jelentős szerepet, illetve olyan, mely részhálózatok közötti kommunikációért lehet felelős. Az ilyen módon nyert információk (esetenként) összevethetők más, például anatómiai úton, vagy egyéb méréstechnikákkal már megismert tényekkel.

2. fejezet

Vizsgálatok

A dolgozatban három vizsgálat eredményét írom le. Mindhárom analízis során korábban létrehozott, szabad hozzáférésű konnektómokat elemeztem. Az első vizsgálat gráfjainak forrása a [8] honlap, melyről a "Human cortex connectivity dataset" bekezdés alatti linkről tölthetőek le. Ebben az elemzésben fontos csúcsok keresése történik.

A második, kiterjedtebb elemzésben konnektómok elemzésére eddig nem használt módszereket próbáltam ki. Két csoportot hasonlítottam össze: az egyik csoport tagjai hordoznak egy olyan génváltozatot, melyet összefüggésbe hoztak időskori Alzheimer-kórral, a másik csoport tagjai az ártalmatlan génváltozatot hordozzák. Ezek a gráfok a [9] honlapon találhatóak meg, az "UCLA_CCN_APOE_DTI" tanulmány tartalmazza őket. Célom az volt, hogy valamely gráfparaméterre összefüggést találjak az életkor, és a betegséggel kapcsolatos génváltozat között.

A harmadik vizsgálat során funkcionális konnektómokat elemeztem, amelyeket ADHD-vel diagnosztizált, illetve egészséges gyerekekről és fiatalokról készült felvételekből állítottak elő. Különböző, a második vizsgálatban nem érintett gráfparaméterek, és az ADHD közötti kapcsolat felderítése volt a célom, ebben az elemzésben arra is kitértem, hogy az egyes paraméterek hogyan viselkednek külön a bal-, illetve a jobb agyféltekében.

2.1. Első vizsgálat

Ebben a vizsgálatban az [1] cikk eredményeire támaszkodtam. A szerzők célja az volt, hogy hálózati elemző módszerek segítségével lokalizálják a strukturális konnektóm központi részét (a strukturális magot), egy olyan részhálózatot, melynek tagjai szorosan összefüggenek egymással, és az egész hálózatban kiemelt fontosságú szerepet tölt be. A fontos csúcsok kereséséhez több olyan módszert is találtam, melyet a cikk szerzői nem próbáltak ki, ez adta az ötletet arra, hogy megvizsgáljam a gráfokat ezek közül kettővel: a PageRank algoritmussal, illetve sajátvektor-centralitással is. A cél annak kiderítése volt, hogy ezek a módszerek mennyiben eltérő területeket eredményeznek azokhoz képest, amelyeket a szerzők találtak az általuk használt módszerrel, a betweenness-centralitással. A cikkhez tartozó gráfok 998 csúcsúak, egy nagyobb agyterület kisebb részeinek felelnek meg, címkéi egy agyterületből és egy számból állnak (pl.: Precuneus 12).

A *betweenness-centralitás* értékét a gráf egy v csúcsára a következő formula adja meg:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_s t(v)}{\sigma_s t},$$

ahol $\sigma_s t$ jelöli a legrövidebb st-utak számát, $\sigma_s t(v)$ pedig azon legrövidebb st-utak számát jelöli, melyek áthaladnak a v csúcson.

A *sajátvektor-centralitást* a következő módon határozzák meg: tekintsük a sajátvektor-egyenletét a gráfnak, melyről feltesszük, hogy egyszerű, és összefüggő (A az adjacencia-mátrix): $Ax = \lambda x$. Ennek az egyenletnek több megoldása is lehet, de ha megköveteljük, hogy a vektor komponensei nem-negatívak legyenek, akkor a Perron-Frobenius tételt felhasználva azt kapjuk, hogy λ a legnagyobb sajátérték lehet csak. Ha azt is előírjuk, hogy a sajátvektor úgy legyen normálva, hogy a legnagyobb komponense 1 legyen, akkor egyértelmű vektort kapunk. Az így kapott egyértelmű vektor lesz definíció szerint a sajátvektor-centralitások vektora, a centralitások pedig ennek a sajátvektornak a komponensei. Ez a mérőszám kifinomultabb, mint a fokszámeloszlás, mivel nem csak a kapcsolatok számát, hanem azoknak minőségét is figyelembe veszi (mennyi

fontos csúcshoz kapcsolódik egy adott csúcs).

A *PageRank* algoritmus egy valószínűségeloszlást határoz meg a csúcsokon, mely azt fejezi ki, hogy egy véletlen bolyongás során milyen valószínűséggel jutunk el az egyes csúcsokba. Ezen kívül, hogy ne akadhassunk el a bolyongás során (például egy olyan csúcsba érve, amelyből nem vezet ki él, vagy egy olyan részgráfban, amelyből kis valószínűséggel tudunk kijutni), minden lépésben bizonyos valószínűséggel tetszőleges csúcsba véletlenszerűen teleportálunk. A következő egyenlet határozza meg az i. csúcs PageRank-jét:

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)},$$

$M(p_i)$ azon csúcsok halmaza, melyekből mutat él a p_i csúcsba, $L(p_j)$ azon csúcsok halmaza, melyekre mutat él a p_j csúcsból, N a csúcsok száma, a d pedig egy konstans, tipikus értéke $d = 0,85$. A PageRank algoritmus az alapja a Google keresőrendszerének, továbbá előnyös tulajdonsága, hogy elméletileg igazoltan robusztus, hibatűrő, így a gráf kisebb változásaira várhatóan kevésbé érzékeny más módszereknél.

A vizsgálat során a Gephi [10] nyílt forráskódú hálózati elemző programot használtam. Lefuttattam a hat rendelkezésemre álló gráfra (az egyik alanyról két gráf készült) a két algoritmust, továbbá meghatároztam a fokszámeloszlást. Az algoritmusok kilistázták a csúcsokat fontosságuk szerinti csökkenő sorrendben. A célom az volt, hogy meghatározzam, mely nagy agyterületeket találja legfontosabbnak a két algoritmus. Külön-külön megnéztem, hogy az algoritmusok melyik száz legfontosabb csúcsot találják meg az egyes gráfokban. Táblázatba foglaltam azokat a nagy területeket, melyekből mindegyik gráfban találtam legalább egy fontos kis területet, feltüntetve, hogy az egyes gráfokban hány kis terület tartozik az adott nagy területhez. A táblázatok bal fölső sarkában a vizsgált centralitás-mérték áll, az első oszlop az anatómiai agyterületet jelenti, az első sor, hogy mely alanyokhoz tartoznak az értékek, a számok pedig azt fejezi ki, hogy az adott személynél a megfelelő nagy agyterületből hány kis terület található a legfontosabb száz csúcs között.

Node Degree	A1	A2	B	C	D	E
precuneus	18	13	22	14	17	24
superiorfrontal	4	17	5	12	10	8
superiortemporal	11	4	17	10	7	5
superiorparietal	8	6	7	12	6	7
inferiorparietal	8	6	5	4	11	7
cuneus	5	3	4	9	2	7
posteriorcingulate	4	6	6	6	2	4
middletemporal	5	2	2	6	4	5
isthmuscingulate	4	4	5	2	1	1
lateraloccipital	2	4	2	1	4	4
caudalanteriorcingulate	3	3	1	2	1	3
rostralmiddlefrontal	2	1	1	3	3	1
inferiortemporal	2	1	1	3	3	1

2.1. táblázat. A fokszámeloszlás szerinti legfontosabb csúcsok

Eigenvector Centrality	A1	A2	B	C	D	E
precuneus	27	20	34	24	23	38
posteriorcingulate	12	10	13	7	10	7
isthmuscingulate	11	11	10	8	5	4
superiorfrontal	3	16	6	8	12	3
cuneus	9	4	8	10	2	9
superiorparietal	5	6	6	10	4	8
superiortemporal	6	2	3	11	4	3

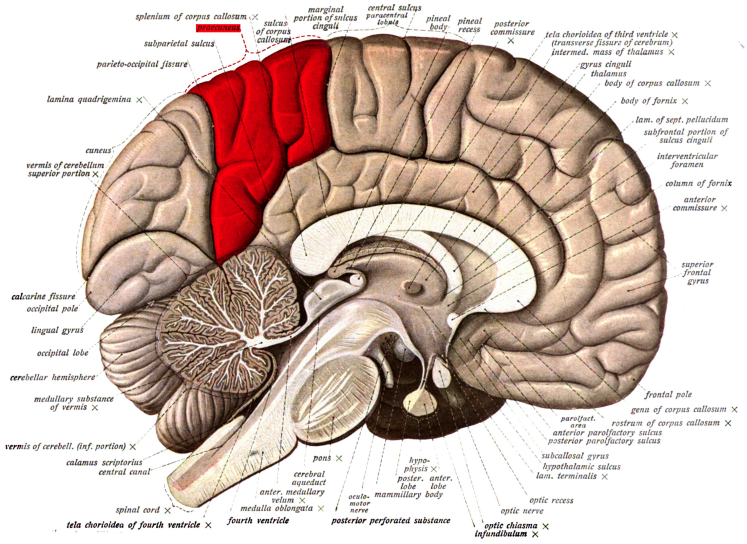
2.2. táblázat. A sajátvektor-centralitás szerinti legfontosabb csúcsok

PageRank	A1	A2	B	C	D	E
superiortemporal	18	11	15	4	7	8
middletemporal	5	8	5	12	12	10
superiorfrontal	4	6	6	12	7	5
precuneus	6	8	6	4	3	11
lateraloccipital	8	5	2	7	5	4
superiorparietal	2	2	1	11	8	7
inferiorparietal	10	5	6	1	3	5
pericalcarine	5	6	9	2	3	4
paracentral	2	2	3	4	8	8
medialorbitofrontal	6	7	4	7	1	1
lateralorbitofrontal	4	3	2	4	9	3
cuneus	5	4	2	6	3	3
bankssts	5	4	6	2	3	2
isthmuscingulate	3	7	2	1	3	3
parstriangularis	3	6	2	3	3	1
rostralmiddlefrontal	1	2	5	3	2	3
postcentral	2	1	4	1	5	2
caudalmiddlefrontal	1	2	4	3	1	3
posteriorcingulate	1	2	1	1	4	5
supramarginal	2	1	2	1	3	1

2.3. táblázat. A PageRank szerinti legfontosabb csúcsok

Figyelemre méltó eredményt adott a sajátvektor-centralitás: összesen hét olyan anatómiai agyterületet talált, melyből legalább egy kis terület szerepelt mindegyik gráf száz legfontosabb csúcsa között. Ráadásul a megtalált területek közül hat terület közös volt a fent említett cikkben talált nyolc, központi fontosságúnak ítélt területtel (a strukturális magot alkotó területekkel), ez alátámasztja, hogy a megtalált területek valóban igen fontosak. A precuneus elnevezésű agyterületből került ki mindegyik gráf esetén a legtöbb csúcs a szá-

zas listára: mind a fokszámeloszlás, mind a sajátvektor-centralitás alapján ez tűnik az egyik legközpontibb agyterületnek. A precuneus az agyműködésben igen fontos szerepet tölt be: összefüggésbe hozzák az epizodikus memóriával, az emberi önreflexióval és bizonyos tudatos funkciókkal is. Az anatómiai vizsgálatok eredményeit a kapott matematikai adatok is megerősítik: valóban centrális fontosságú, integratív szerepű agyterület.



2.1. ábra. Agyterületek, kiemelve a precuneus. Forrás: [13] honlap

A PageRank a sajátvektor-centralitással szemben elég sok agyterületet kiválasztott mindegyik gráfból, így kevésbé kimagaslóak a leginkább jelen lévő agyterületek. Megfigyelhető, hogy az összes olyan agyterületet, melyet a sajátvektor-centralitás megtalált, a PageRank is, viszont a PageRank 5 legfontosabbnak ítélt területéből 3 egyáltalán nem szerepel a sajátvektor-centralitás fontos területei között. A kapott eredményeket nehéz olyan módon értelmezni, mint a sajátvektor-centralitás eredményeit.

Végezetül kiszámoltam a kipróbált centralitás-mértékek korrelációit is, azt találtam, hogy a fokszámeloszlás és a sajátvektor-centralitás nagyon erősen korrelál. Ennek ellenére nem értelmetlen a sajátvektor-centralitás, hiszen lényegesen kisebb listát eredményezett, mint a fokszámeloszlás, továbbá a lista első 20-30 elemébe a fokszámeloszláshoz képest minden esetben megnövelte a precuneushoz tartozó területek koncentrációját. Rövidítések: degree: fokszám-

eloszlás, eigen: sajátvektor-centralitás.

Korreláció	A1	A2	B	C	D	E
Degree-PageRank	0,61	0,55	0,57	0,52	0,46	0,52
Degree-Eigen	0,87	0,90	0,86	0,92	0,93	0,86
PageRank-Eigen	0,45	0,44	0,40	0,44	0,37	0,41

2.4. táblázat. A centralitás-mértékek korrelációi

2.2. Második vizsgálat

A második vizsgálat alapját a [2] cikk és az ahhoz tartozó strukturális konnektómok képezik [9]. Az elérhető adatok súlyozott gráfok adjacencia-mátrixai voltak, melyekben az élsúly az egyes területek közötti kapcsolatok erősségét jelentik. Mindegyik gráf 110 csúcsú, a csúcsok anatómiailag meghatározott agyterületeknek felelnek meg.

55 személyt vizsgáltak, ezek közül 25-en az Alzheimer-kórral összefüggésbe hozható génváltozatot (apolipoprotein E epsilon 4 allél, APOE-4) hordozták, 30-an pedig nem voltak hordozói ennek az allélnak (APOE-3). Rendelkezésre álló adat volt még a génváltozaton kívül az életkor, nem, illetve a használt MRI-készülék típusa (a teljes vizsgálat során két, különböző típusú scanner-t alkalmaztak). A nemre, illetve scanner-re vonatkozó adatokat a cikk példáját követve a korrelációk számításakor figyelembe vesszük, továbbá azt is különkülön megnézem, milyen hatással van a korrelációra a gráfok élszáma, illetve az élek összsúlya, ez parciális korreláció kiszámolásával történik. A parciális korreláció lényege, hogy két változó közötti korrelációt vizsgálunk úgy, hogy bizonyos egyéb tényezők (ezeket nevezem *kontrollált tényezőnek*) hatását kiszűrjük. Az elemzéshez az igraph [11] nyílt forráskódú program R-hez készült változatát használtam, a parciális korrelációk számításához szintén az R-hez készített ppcor csomagot [12]. A számításokhoz használt korrelációs formulák az alábbiak:

-Két változó korrelációja:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu(X))(Y-\mu(Y))]}{\sigma_X \sigma_Y},$$

ahol E a várható érték operátor, $\mu(X)$ az X változó várható értékét, $\sigma(X)$ a szórását jelenti, $cov(X,Y)$ pedig X és Y kovarianciáját.

-Az 1,2 változók parciális korrelációja úgy, hogy egy 3. változót kontrollálunk:

$$\rho_{12.3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23}^2}}.$$

Általában, N változó esetében a parciális korrelációja az első két változónak,

kontrollálva a többi $N-2$-t rekurzívan határozható meg:

$$\rho_{12.34...N} = \frac{\rho_{12.34...(N-1)} - \rho_{1N.34...(N-1)}\rho_{2N.34...(N-1)}}{\sqrt{1-\rho^2_{1N.34...(N-1)}}\sqrt{1-\rho^2_{2N.34...(N-1)}}}.$$

A vizsgált gráftulajdonságok a következők voltak:

- Pontdiszjunkt utak: speciálisan az összes pontpárra nézve a pontdiszjunkt utak számának minimuma (ezt nevezik a gráf kohéziójának is). Kiderült, hogy egy (u, v) pontpárra nézve a pontdiszjunkt utak száma az esetek túlnyomó többségében megegyezik a kisebb fokszámú csúcs fokszámával, vagyis a kohézió értéke egyszerűen azonos a gráf legkisebb fokszámával. Emiatt az egyébként érdekes jelenség miatt ezt a számot a további elemzés során nem vettem figyelembe.

- Klikkszám (a gráf legnagyobb klikkjének mérete).

- Legnagyobb klikkek száma.

- Maximális klikkek (azaz nagyobbá nem kiegészíthető klikkek) száma.

- A gráf átmérője (minden pontpárra nézett legrövidebb utak közül a leghosszabb hossza).

- A gráf függetlenségi száma (a legnagyobb független halmaz számossága, megegyezik a gráf komplementerének a klikkszámával).

- Maximális folyam (azaz minimális vágás) nagysága: a legnagyobb maximális folyam (minimális vágás) értéke az összes lehetséges forrás-nyelő párra nézve.

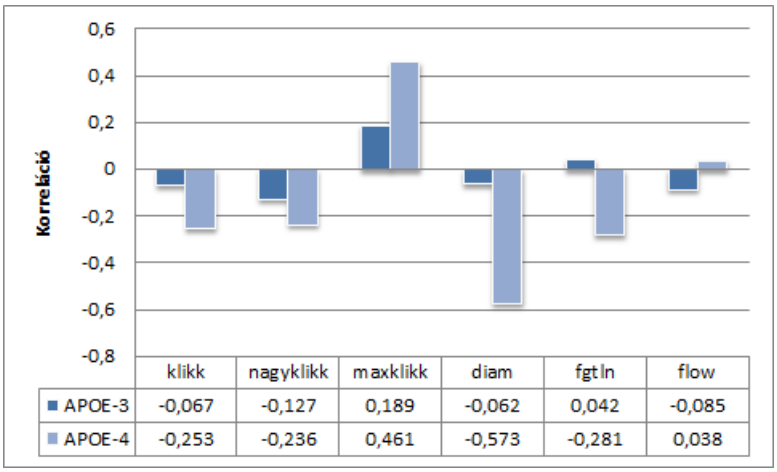
Ezeket a paramétereket minden gráfra meghatároztam, majd mind a két csoportra kiszámítottam a korrelációt az életkor és az egyes paraméterek között. Táblázatban feltüntetem az eredményeket, a korrelációk értékét a paraméterek és az életkor között először egyéb tényezők figyelembe vétele nélkül, majd parciális korrelációkat is:

- Kontrollálva a nemet és a scanner típusát.

- A következő paramétereket ezen felül élszámmal is kontrolláltam, mivel függenek attól: klikkszám, legnagyobb klikkek száma, maximális klikkek száma, függetlenségi szám.

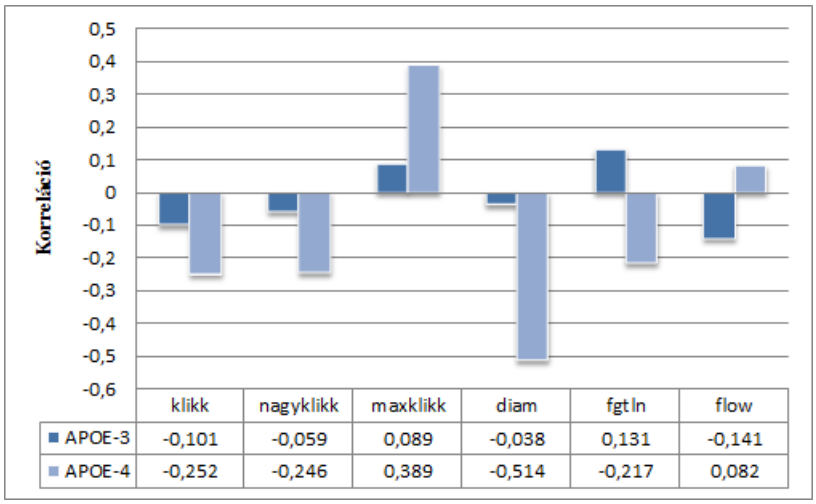
- Végül a nemen és scanner-en felül élszám helyett élsúllyal kontrolláltam az átmérőt, illetve a legnagyobb maximális folyam nagyságát.

A táblázatban az APOE-3 jelenti az ártalmatlan génváltozatot, az APOE-4 a betegségre hajlamosítót. Az alkalmazott rövidítések jelentése: klikk: klikkszám, nagyklikk: legnagyobb klikkek száma, maxklikk: maximális klikkek száma, diam: átmérő, ftln szám: függetlenségi szám, flow: legnagyobb maximális folyam értéke.

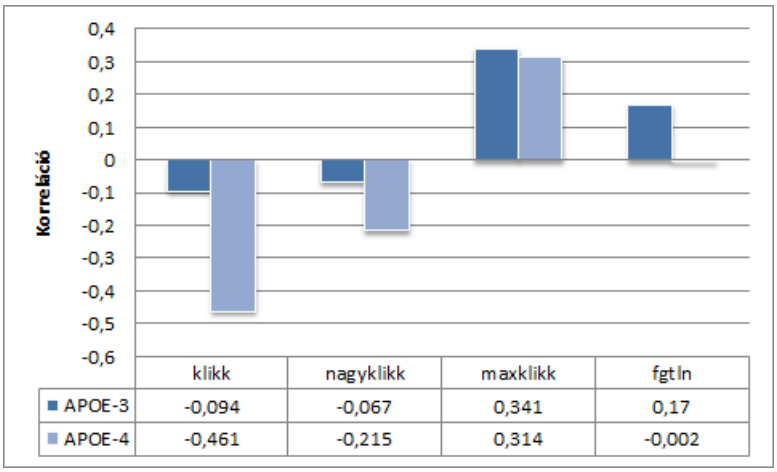


	klikk	nagyklikk	maxklikk	diam	fgtln	flow
APOE-3	-0,067	-0,127	0,189	-0,062	0,042	-0,085
APOE-4	-0,253	-0,236	0,461	-0,573	-0,281	0,038

2.2. ábra. Korreláció az életkor és a paraméterek között



2.3. ábra. Parciális korreláció az életkor és a paraméterek között, kontrollálva a nemet és a scanner-t

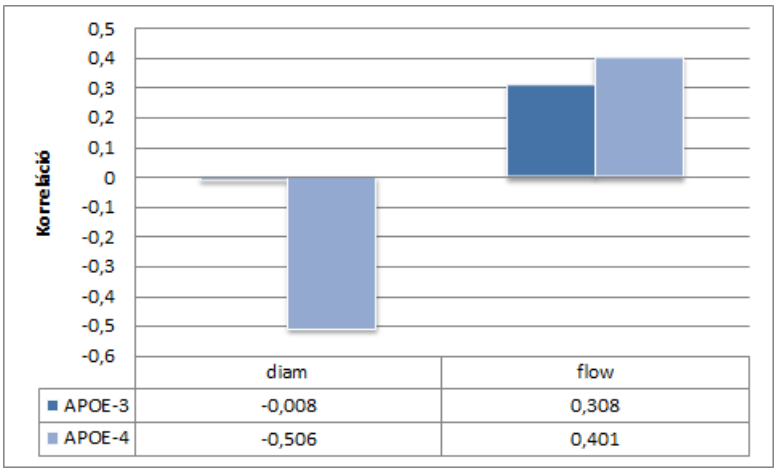


2.4. ábra. Parciális korreláció az életkor és a paraméterek között, kontrollálva a nemet, scanner-t és élszámot

A legfigyelemreméltóbb eredmény az átmérő korrelációja az életkorral az APOE-4 hordozók között: minden esetben nagyon erős negatív korreláció figyelhető meg, míg az allélt nem hordozók között elhanyagolható a korreláció.

Szintén érdekes a klikkszám erős negatív korrelációja a betegséggel abban az esetben, amikor nemre, scanner-re és élszámra is kontrolláltunk. A többi esetben a korreláció csak közepesnek mondható (szintén negatív). A korreláció jelentéktelen az APOE-3 csoportban.

Közepes erősségű negatív korreláció figyelhető meg a legnagyobb klikkek



2.5. ábra. Parciális korreláció az életkor és a paraméterek között, kontrollálva a nemet, scanner-t és élsúlyt

számánál az APOE-4 csoportban.

A maximális klikkek esetén a kontrolláló tényezők figyelembe vételével az APOE-4 csoportban kezdetben megfigyelhető nagyon erős pozitív korreláció fokozatosan eltűnik, az összes tényező figyelembe vételekor már jelentéktelen lesz a különbség a két csoport között.

A függetlenségi számnál a közepesen erős negatív korreláció a kontrolláló tényezők figyelembe vételével gyakorlatilag teljesen megszűnik, míg az APOE-3 csoportban gyenge pozitív korrelációt tapasztaltunk a kontrolláló tényezőkkel.

A legnagyobb maximális folyam értékének korrelációja az életkorral egyéb tényezők figyelmen kívül hagyásával elhanyagolható, és nem túl érdekes akkor sem, ha kontrollálunk a nemmel és a scanner-rel is. Ha az élsúlyt is figyelembe vesszük, akkor közepesen erős, illetve erős pozitív korreláció tapasztalható mindkét csoportban.

2.3. Harmadik vizsgálat

A harmadik vizsgálatban funkcionális konnektómokat elemeztem R-ben az iGraph csomag [11] segítségével, az adatok a [9] adatbázisból származó, "ADHD200_CC200" címkével ellátott gráfok. Az adathalmaz 520 konnektómot tartalmaz, ezek mindegyike egységesen 190 csúcsú gráf, a csúcsok agyterületeknek felelnek meg. A gráfokat ADHD-s, illetve egészséges gyerekek, fiatalok agyáról készült fMRI felvételekből állították elő.

Az 520 konnektómból 330 tartozik egészséges, 190 pedig ADHD-val diagnosztizált személyhez. Az ADHD-nak több fajtája van, ezek a típusok fel vannak tüntetve az egyes gráfoknál, a vizsgálatban viszont a különböző típusokat egyetlen csoportban kezeltem, tehát két csoportba különítettem a gráfokat: ADHD-s és egészséges csoportokba.

A vizsgálat célja az volt, hogy összefüggéseket találjunk bizonyos gráfparaméterek, illetve az ADHD között. A gráfparaméterek és az egyes csoportok összefüggését külön-külön megnéztem a teljes agygráfokban, illetve a bal- és jobb agyféltekére külön-külön. A konnektómok élsúlyozott teljes gráfok voltak, vagyis az összes korrelációt tartalmazták az agyterületek között, a pozitív és negatív értékeket is.

Az adatok előfeldolgozása során először kiszűrtem a nem-pozitív éleket, majd ritkítottam a gráfokat, hogy csak erősebb korrelációkat (nagyobb súlyú éleket) tartalmazzon. Ehhez nagyság szerint növekvő sorrendbe rendeztem a megmaradt éleket, az így kapott listának a felső 40%-át -a legnagyobb súlyú éleket- megtartottam, így körülbelül 20%-os élsűrűségű gráfokhoz jutottam. Ezután a gráfokból előállítottam a bal- és jobb agyféltekéhez tartozó csúcsok által feszített részgráfokat, végül töröltem az élek eltávolítása során keletkezett izolált csúcsokat. Az ADHD csoportban az egyik gráfhoz nem volt definiálva az alany neme, egy másiknál pedig a számítások során nem értelmes adatok keletkeztek, ezeket kihagyva összesen 518 gráfon készült az elemzés. Négy gráfparaméter értékét határoztam meg a teljes konnektómra, illetve a bal- és jobb agyféltekékre külön-külön. Ezen paraméterek vizsgálatához az ötletet a [6] cikk adta.

- Hoffman-korlát, HK

  Legyen λ_{min}, λ_{max} a G gráf szomszédsági mátrixának legkisebb, illetve legnagyobb sajátértéke. Ekkor a Hoffman-korlátot az alábbi képlettel kapjuk: $1 + \frac{\lambda_{max}}{\lambda_{min}}$. Említésre méltó, hogy a Hoffman-korlát alsó becslést ad G kromatikus számára.

- Normalizált legnagyobb sajátérték, NLS

  Az általánosított szomszédsági mátrix olyan nxn-es mátrix, amelyben a sorok és oszlopok a csúcsoknak felelnek meg, a mátrix egy a_{ij} eleme 0, ha az i és j csúcs között nincs él, egyébként pedig a két csúcsot összekötő él súlya. A normalizált legnagyobb sajátértéket ebből a mátrixból kapjuk: a legnagyobb sajátértékét, λ_{max}-ot elosztjuk az átlagos fokszámmal. Mivel a legnagyobb sajátérték az átlagos fokszám és a maximális fokszám közé esik, ezért tekinthető egyfajta "átlagos foknak". Ebből az következik, hogy egy sűrűbb gráfnak lehet nagyobb λ_{max} legnagyobb sajátértéke azért, mert nagyobb átlagos fokszáma van. Ezért kell leosztani a legnagyobb sajátértéket az átlagos fokszámmal.

- Az átmenetmátrix sajátérték-hézagja, SH

  Egy csúcs általánosított foka a csúccsal szomszédos élek összege. Ennek segítségével határozható meg az átmenetmátrix, amelyet úgy kapunk, hogy az általánosított szomszédsági mátrix minden sorát elosztjuk a sorhoz tartozó csúcs általánosított fokával. Amikor egy véletlen bolyongást hajtunk végre egy gráfon, az i, j csúcsokra az átmenetmátrix megfelelő eleme az i csúcsból j csúcsba átlépés valószínűségét írja le. Az átmenetmátrix legnagyobb, és második legnagyobb sajátértékének a különbségét nevezzük sajátérték-hézagnak. Ez az érték jellemzi a gráf expandertulajdonságát: minél nagyobb a sajátérték-hézag, annál jobb expander a gráf.

- Feszítőfák számának logaritmusa, FSL

  Jelölje A a G gráf adjacencia-mátrixát, D azt a diagonális mátrixot, melynek d_{ii} eleme az i csúcs fokszáma. Jelölje $L = D - A$ a G Laplace-

mátrixát. Kirchoff tétele alapján a Laplace-mátrix segítségével meghatározható a gráf feszítő fáinak száma: legyenek $\lambda_1, \lambda_2, \ldots, \lambda_{n-1}$ L nem-nulla sajátértékei. Ekkor a feszítőfák száma a következőképpen adható meg: $t(G) = \frac{1}{n}\prod_{i=1}^{n-1}\lambda_i$. A vizsgálat során az így kapott érték logaritmusát használtam.

Elemzés

A csoportok és gráfparaméterek közötti összefüggés elemzéséhez az ANOVA (analysis of variance) módszert használtam. A nullhipotézis az volt, hogy a gráfparaméterek nem különböznek a két csoportban. Az összes vizsgált érték száma 12 volt (4 paraméter a teljes agygráfokra, és a féltekékre külön-külön kiszámolva). Ezek közül a következő attribútumoknál találtam szignifikáns eltérést:

- Hoffman-korlát a jobb agyféltekére nézve
- Legnagyobb normalizált sajátérték a jobb agyféltekére nézve
- Feszítőfák számának logaritmusa a jobb agyféltekére nézve

A következő táblázat tartalmazza ezeket a paramétereket, és a kiszámolt adatokat:

paraméter	**átlag (ADHD)**	**átlag (egészséges)**	**p érték**
HK, jobb agyfélteke	5,284	5,160	0,00278
LNS, jobb agyfélteke	1,226	1,209	0,0503
FSL, jobb agyfélteke	256,232	255,002	0,0279

2.5. táblázat. Szignifikáns különbségek

3. fejezet

Következtetések

Az első vizsgálatban 998 csúcsú strukturális konnektómok fontos csúcsait vizsgáltam, összehasonlítva az eredményeket az [1] cikk adataival.

Azt tapasztaltam, hogy a sajátvektor-centralitás szerinti fontosság-sorrend alapján megtalált hét agyterület közül hat azonos a cikkben talált nyolc agyterülettel, ezzel azt mutattam meg, hogy ez a mérték önmagában nézve is igen jól használható fontos területek megtalálásához. A PageRank-kel talált területek nem mutattak jelentős egyezést, a tapasztalatok alapján a PageRank ezekben a gráfokban más hálózatot talál jelentősnek.

Fontos megjegyezni, hogy a PageRank igazán irányított gráfokon működik jól, és a konnektómok jelenleg irányítatlanok. Érdekes lenne megvizsgálni valamilyen módon megirányított konnektómon is a PageRank viselkedését, ez egy hasznos további kutatás tárgya lehet.

A második vizsgálatban, melynek alapja a [2] cikk, és az azokhoz tartozó konnektómok voltak, eddig figyelmen kívül hagyott, ám érdekes gráfparamétereket vizsgáltam. 55 személyről készültek konnektómok, közülük 25 az APOE-4 génváltozatot hordozza, melyet összefüggésbe hoztak az időskori Alzheimer-kórral, 30-an pedig az ártalmatlan változatot hordozzák. Célom az volt, hogy találjak olyan, a szakirodalomban eddig jellemzően nem vizsgált gráfparamétert, mely jelentős mértékben összefügg az életkor előrehaladottságával az Alzheimer-kórra hajlamosító gént hordozók csoportjában.

A vizsgált mértékek a következők voltak: klikkszám, a legnagyobb klikkek

száma, maximális klikkek száma, a gráf átmérője, függetlenségi szám, illetve a legnagyobb maximális folyam nagysága. A vizsgálat során ügyeltem arra, hogy szűrjem bizonyos tényezők (nem, az MRI-készülék típusa, élszám, élsúly) okozta korrelációs hatást. Az életkor előrehaladásával legjelentősebben összefüggő gráfparaméterek a gráf átmérője, illetve a klikkszám voltak. Az erős negatív összefüggés az életkor és a paraméterek között arra enged követni, hogy ezek a mérőszámok fontos kapcsolatban állnak az APOE-4 hordozók konnektómjával, további kutatási lehetőség lenne ezeket, és a többi mértéket egyéb adathalmazokon megvizsgálni, és jobban megérteni a működésüket.

A harmadik vizsgálatban újabb paramétereket tanulmányoztam egészséges, illetve ADHD-val diagnosztizált fiatalok funkcionális konnektómjaiban azzal a céllal, hogy összefüggést találjak a paraméterek és az ADHD között. Ezúttal a teljes konnektómon kívül azt is megnéztem, hogy külön-külön az egyes agyféltekékben milyen viselkedése van a paramétereknek.

Ebben az elemzésben a következő paraméterek szerepeltek: Hoffman-korlát (HK), normalizált legnagyobb sajátérték (NLS), az átmeneti mátrix sajátértékhézagja (SH), illetve a feszítőfák számának logaritmusa (FSL). Az ANOVA módszert használva azt találtam, hogy az említett paraméterek közül a jobb agyféltekében három szignifikánsan különbözik az ADHD-val diagnosztizáltak csoportjában: HK, NLS, FSL.

Az elkövetkező néhány évben várhatóan továbbra is nagy fontosságúak lesznek a kisebb, körülbelül száz csúcsot tartalmazó konnektómok. A tapasztalat azt támasztja alá, hogy az ilyen mérettartományon meghatározható a klikk- és a függetlenségi szám, mely feladatok általában nehezek (NP-teljes problémák), ezért nagy hálózatokon nem vizsgálhatóak, de ezen a mérettartományon, új kísérletekben megvizsgálható lesz a viselkedésük.

Ábrák jegyzéke

1.1. A képfeldolgozás lépései. Forrás: [1] cikk 6

2.1. Agyterületek, kiemelve a precuneus. Forrás: [13] honlap 15

2.2. Korreláció az életkor és a paraméterek között 19

2.3. Parciális korreláció az életkor és a paraméterek között, kontrollálva a nemet és a scanner-t 20

2.4. Parciális korreláció az életkor és a paraméterek között, kontrollálva a nemet, scanner-t és élszámot 20

2.5. Parciális korreláció az életkor és a paraméterek között, kontrollálva a nemet, scanner-t és élsúlyt 21

Táblázatok jegyzéke

2.1. A fokszámeloszlás szerinti legfontosabb csúcsok 13
2.2. A sajátvektor-centralitás szerinti legfontosabb csúcsok 13
2.3. A PageRank szerinti legfontosabb csúcsok 14
2.4. A centralitás-mértékek korrelációi 16
2.5. Szignifikáns különbségek 24

Irodalomjegyzék

[1] Patric Hagmann et al.: *Mapping the Structural Core of Human Cerebral Cortex*, PLOS Biology, 2008.

[2] Jesse A. Brown et al.: *Brain network local interconnectivity loss in aging APOE-4 allele carriers*, PNAS, 2011.

[3] Alex Fornito, Andrew Zalesky, Michael Breakspear: *Graph analysis of the human connectome: Promise, progress and pitfalls*, NeuroImage, 2013.

[4] Olaf Sporns: *Discovering the Human Connectome*, The MIT Press, 2012.

[5] Gaël Varoquaux, R. Cameron Craddock: *Learning and comparing functional connectomes across subjects*, NeuroImage, 2013.

[6] Balázs Szalkai, Bálint Varga, Vince Grolmusz: *Graph-theoretical Analysis Reveals: Women's Brains are Better Connected than Men's*, PLoS One, 2015.

[7] `http://en.wikipedia.org/wiki/Centrality`

[8] `http://www.cmtk.org/viewer/datasets/`

[9] `http://umcd.humanconnectomeproject.org/umcd/default/browse_studies`

[10] `http://gephi.github.io/`

[11] `http://igraph.org/`

[12] `http://cran.r-project.org/web/packages/ppcor/ppcor.pdf`

[13] `http://en.wikipedia.org/wiki/Precuneus`