



Универзитет „Св. Кирил и Методиј“ во Скопје
**ФАКУЛТЕТ ЗА ИНФОРМАТИЧКИ НАУКИ И
КОМПЈУТЕРСКО ИНЖЕНЕРСТВО**

Откривање на знаење во мозочни мрежи

Докторска дисертација

Илинка Иваноска

Ментор:

Проф. д-р Слободан Калајциски

Скопје, Септември 2021

КОМИСИЈА

Проф. Соња Гиевска, Претседател

Факултет за информатички науки и компјутерско инженерство
Универзитет “Св. Кирил и Методиј” во Скопје

Проф. Слободан Калајциски, Ментор

Факултет за информатички науки и компјутерско инженерство
Универзитет “Св. Кирил и Методиј” во Скопје

Проф. Љупчо Коцарев, Член

Факултет за информатички науки и компјутерско инженерство
Универзитет “Св. Кирил и Методиј” во Скопје

Проф. Кире Триводалиев, Член

Факултет за информатички науки и компјутерско инженерство
Универзитет “Св. Кирил и Методиј” во Скопје

Проф. Сашо Коцевски, Надворешен член

Универзитет “Гоце Делчев” во Штип

Абстракт

Во невронауката, примената на комплексните мрежи е еден од стандардните начини за претставување на динамиката на мозокот. Сложената структура на мозокот лесно може да се визуелизира, доколку се мапираат $\approx 10^6$ воксели од некој типичен fMRI протокол за снимање на цел мозок во јазли од поврзана мрежа. На овој начин се обезбедуваат функционални податоци за невро-визуелизација, а дополнително се нуди и можност за проучување на тополошките својства скриени во меѓусебните односи помеѓу јазлите од кои е составена мрежата. Колку и да изгледа едноставно и лесно изводливо, пресметувањето на поврзаноста меѓу толку голем број јазли во мрежата може да биде голем предизвик од компјутерска и од толкувачка гледна точка. Едно од можните решенија е намалувањето на димензионалноста на мрежата, што вклучува нетривијална ренормализација на анатомскиот простор во функционално одделни региони на интерес (Regions of Interest - ROI). Во зависност од анатомските принципи на парцелација на просторот и правилата за интеракција што се усвоени, овој процес може да се спроведе на различни начини, за што постојат општо прифатени принципи. Дополнително, квантифицирањето на варијациите во поврзаноста помеѓу јазлите во мозокот за состојбите кај популации на пациенти и процената на нивното функционално значење се особено нетривијални задачи. Деталниот преглед на мрежата доаѓа со цена на висока димензионалност, што го попречува и човечкиот капацитет за дешифрирање на главните механизми што постојат кај различните патологии, а истото е пречка и при изведувањето на секоја статистичка задача и/или задача за машинско учење што се користи при обработка на овие податоци. Очигледно, дел од можните решенија, што обезбедуваат подобро користење на мрежните мозочни репрезентации, се методите за избор на врски, нудејќи отстранување на ирелевантни врски во дадено сценарио.

Во оваа докторска дисертација ќе се осврнеме на методите за конструкција и анализа на мозочните мрежи од повеќе аспекти. Прво, реконструираме функционални мозочни мрежи со комбинирање на различни методи за ренормализација на врските. Методите се засновани и на фреквенцистички и на баесови пристапи кои ги предлагаме, како и на стратегии за праг. Потоа, оценуваме до кој степен се засегнати својствата на добиената тополошка мрежа. Овие својства понатаму ги користиме за да ги разликуваме субјектите со невролошки нарушувања од соодветна контролна група. Ќе покажеме дека разликите во рангирањето на врските предизвикуваат специфични геометрии што се појавуваат во еден помал простор. Ќе разгледаме и големо множество статистички методи и методи

за избор на врски од машинско учење и истите ги евалуираме врз реални функционални мозочни мрежи. На крај, ќе го презентираме веб-базираниот систем за обработка на мозочни мрежи кој нуди различни алгоритми за т.н. мрежно вгнездување.

Резултатите од оваа докторска дисертација покажуваат дека различните методи за реконструкција на врските можат да доведат до квантитативни разлики меѓу популациите, но не влијаат на целокупната дискриминаторска моќ. Разликите меѓу популациите се повеќе видливи со пропорционални, наместо фиксни прагови за бинаризација. Тие се специфични за мрежните карактеристики, минимизирајќи го влијанието на избраниот атлас. Исто така, рангирањето на врските може да обезбеди брзи и сигурни критериуми за параметрите за реконструкција на мрежата. Од разгледаните статистички методи и методи од машинско учење за избор на врски резултатите покажуваат дека повеќето методи работат на квалитативно сличен начин. Методите од машинско учење се концептуално посложени од статистичките методи, но нивното користење не дава некоја познателна предност пред статистичките. Истовремено, високата хетерогеност во множеството врски што се задржани од секој од методите, сугерира дека различните методи даваат комплементарен преглед на податоците. Иако, моќта на дискриминација на функционалните мрежи не е глобално засегната, треба да се внимава во изборот на достапни методи и во толкувањето на резултатите што се добиваат со нив. Во дисертацијата се дискутираат и импликациите на овие резултати во невронауката. Конечно, изработениот веб-базиран систем им овозможува на корисниците брзо и лесно да го најдат најдобриот метод за нивниот проблем за анализа на мозочни мрежи и да ја преземат векторската репрезентација на мозочната мрежа за понатамошна анализа.

Клучни зборови:

Мозочни мрежи, функционален конектом, теорија на графови, статистичка анализа, функционална магнетна резонанса.

Abstract

Representing brain dynamics as a complex network is standard practice in neuroscience. In order to provide functional neuroimaging data with network representation, it is generally possible to map $\approx 10^6$ voxels from a typical fMRI protocol to a network and to study the topological properties emerging from the relationships from its nodes. However, calculating connections between large number of units can be challenging from both a computational and an interpretational point of view. The dimensionality reduction involves non-trivial renormalization of the anatomical space into functionally separate regions of interest (ROI). This process can be carried out in different ways, depending on whether it is analyzed from the anatomical principle of the spatial parcelization and the integration rules involved. Additionally, quantifying variations in the connections between different nodes in the brain across different populations and assessing their functional significance are very non-trivial tasks. Also, this advanced network overview entails a high-dimensional cost, which prevents the capability to decipher the main mechanisms behind the pathologies, and the meaning of statistical and/or machine learning tasks used to process the data. The link selection method, which allows the removal of irrelevant links in a given scenario, is a clear solution that can improve the use of these network representations.

In this doctoral dissertation we look at methods of construction and analysis of brain networks from multiple views. First, we reconstruct functional brain networks by combining different methods for links renormalization, based on frequentist and Bayesian proposed approaches and threshold strategies. Then, we asses the extent to which network topological characteristics are affected to distinguish between subjects with neurological disorders and a control group. Moreover, we show that differences in link rankings cause specific geometries in an auxiliary space that can often easily be detected with a simple visual check. We overview and a large set of statistical methods and machine learning methods for link selection and make assessment of real functional brain networks. At the end of the work, a web-based system for processing brain networks with different embedding algorithms is presented.

The results of this dissertation show that different methods for link reconstruction can lead to quantitatively differences between different populations, but do not affect overall discriminatory power. Differences between populations are more visible with proportional, instead of fixed binary thresholds, while the atlas choice is not important. Link ranking can also provide a fast and reliable criterion for network reconstruction parameters for which no theoretical guidance has been proposed. From

the considered statistical methods and machine learning methods for selecting links, results show that the methods act in a qualitatively similar way. While machine learning methods are conceptually more complicated than statistical ones, they do not present a clear advantage. At the same time, the high heterogeneity in the set of links kept by any method, suggests that they offer complementary dataset views. Although power to discriminate functional brain networks is not globally affected, care should be taken in the choice of available methods and in interpreting the results they give. The dissertation also discusses the implications of these results in neuroscience. Finally, the developed web-based system allows users to quickly and easily find the best method for their brain network problem and download the vector representation for further analysis.

Keywords:

Brain networks, functional connectome, graph theory, statistical analysis, fMRI.

Благодарност

Патот до оваа дисертација беше исполнет со поддршка, разбирање, инспирација и верба од повеќе луѓе, на кои сум им искрено благодарна.

Сакам да изразам особена благодарност на мојот ментор. проф. Слободан Калајциски, кој со својот несебичен пристап ме водеше изминатите години низ мојот високообразовен процес - од додипломските студии, преку магистерските студии, до докторските. Охрабрувајќи ме да прифатам големи предизвици на секој чекор и насочувајќи ме низ нив, ми овозможи да прераснам во добар истражувач и професионалец во областа. Огромна благодарност и до др. Масимилиано Занин кој ја имаше улогата на надворешен ментор во текот на изработката на докторската дисертација.

Сакам да им се заблагодарам и на моите колеги за прекрасната соработка во рамките на научно истражувачките проекти на Факултетот проф. Кире Триводалиев, проф. Билјана Ристеска Стојкоска, проф. Соња Гиевска и проф. Георгина Мирчева. Дополнително сакам да им се заблагодарам и на колегите: Фросина Стојановска, Томи Бошковски, Стефан Кочев, Мартин Миленковски и Игор Гешоски со кои работевме на истражувања и публикации кои се дел од оваа дисертација. Исто така, посебна благодарност до др. Весна Прчковска и целиот тим на компанијата QMENTA доо кои беа инспирација и ме воведоа во темата за оваа докторска дисертација.

На крај, им се заблагодарувам и на моите родители, кои ми пружаа неизмерна поддршка во текот на студиите и гордо ги поддржуваа моите академски и научноистражувачки амбиции.

Содржина

Листа на слики	10
Листа на табели	12
1 Вовед во мозочни мрежи	15
2 Реконструкција на мозочни мрежи	21
2.1 Методи и податоци	24
2.1.1 Податочни множества и претпроцесирање	24
2.1.2 Претпроцесирање на податочно множество	25
2.1.3 Методи за реконструкција на функционални мозочни мрежи	25
2.1.4 Анализа на функционални мозочни мрежи	27
2.2 Резултати	29
2.3 Дискусија	33
3 Брза трансформација за евалуација на разлики во групи на мозочни мрежи	39
3.1 Методи и податоци	40
3.1.1 Креирање на репрезентација	40
3.1.2 Податочни множества	42
3.1.3 Реконструкција на функционални мозочни мрежи	47
3.2 Резултати и дискусија	48
4 Статистички методи и методи базирани на машинско учење за селекција на карактеристики кај мозочни функционални мрежи	58
4.1 Методи за селекција на линкови кај функционални мозочни мрежи	61
4.1.1 Статистички, униваријантни (локални) методи	62
4.1.2 Статистички, мултиваријатни (глобални) методи	63
4.1.3 Машинско учење	69
4.2 Податочни множества и класификациски алгоритми	74

4.2.1	Аквизиција на мозочните слики и реконструкција на мозочните мрежи	74
4.2.2	Алгоритми за класификација	75
4.3	Резултати	79
4.3.1	Споредба на перформанси	79
4.3.2	Intra- and inter-стабилност на метод	82
4.3.3	Перформанси	83
4.3.4	Генерализирање на резултатите	85
4.4	Дискусија	86
5	Веб базиран систем за евалуација на техники базирани на мрежни вгнездувања за обработка на мозочни мрежи	92
5.1	Методи	94
5.1.1	Методи за граф трансформација во векторски простор	94
5.1.2	Метод за евалуација - реконструкција на граф	96
5.2	Спецификации на системот	97
5.2.1	Функционалности на системот	97
6	Заклучок	102
A	Дополнителни слики и табели	105
A.1	ABIDE: Сите резултати	107
A.2	ABIDE: Споредба контролни субјекти vs. пациенти	123
A.3	ABIDE: Споредба на различни атласи	131
A.4	COBRE: Сите резултати	135
A.5	COBRE: Споредба контролни субјекти vs. пациенти	139

Листа на слики

2-1	Еволуција на две тополошки метрики (кластерирачки коефициент, леви панели и ефикасност, десни панели) како функција на прагот τ (горните панели) и густината на врската l_d (долни панели). . . .	31
2-2	Споредба на коефициентите на корелација.	32
2-3	Еволуција на две тополошки метрики (кластерирачки коефициент, леви панели и ефикасност, десни панели) како функција на прагот τ (горните панели) и густината на врската l_d (долни панели), за податочното множество COBRE.	33
2-4	Еволуција на кластерирачкиот коефициент (леви панели) и ефикасноста (десни панели) како функција на прагот τ (горни панели) и густината на врската (долни панели).	34
2-5	Еволуција на Коен d за кластерирачкиот коефициент (леви панели) и ефикасноста (десни панели), како функција на прагот τ (горните панели) и на густината на врската (долу панели).	35
2-6	Еволуција на максималната Коен d за осумте тополошки метрики, за апсолутниот (горниот панел) и пропорционалниот (долниот панел) пристапен праг.	36
2-7	Еволуција на максималната Коен d за осумте тополошки метрики, за апсолутниот (горниот панел) и пропорционалниот (долниот панел) пристапен праг, за множеството COBRE.	37
3-1	Графички приказ на предложениот метод.	42
3-2	Анализа на резултатите дадени од збир на синтетички модели. . .	49
3-3	Анализа на резолуцијата на fMRI податоците.	50
3-4	Анализа на резултатите добиени со предложениот метод за множеството за шизофренија EEG.	51
3-5	Анализа на резултатите добиени со предложениот метод за податочното множество Алцхајмерова болест EEG	52
3-6	Анализа на робусноста на методот во однос на додаден шум. . . .	53

3-7	Примери за примена на предложената методологија на други сложени системи.	54
3-8	Графичка претстава дадена од предложената методологија за деветте функционални и структурни податочни множества за мозочни мрежи разгледани овде.	55
3-9	Примери за алтернативна употреба на предложената методологија.	57
4-1	Преглед на методите за избор на карактеристики, класифицирани според нивната природа.	61
4-2	Најдобар резултат за класификација, во смисла на AUC (лев панел) и F1 (десен панел), добиен со методите за избор на врски, овде разгледани.	80
4-3	(Лево) Број на врски задржани со секој метод (со исклучок на трите со најмалку перформанси). (Десно) Број на задржани врски како функција на резултатот (AUC) постигнат со секој метод. . .	81
4-4	Еволуција на резултатот на класификација, во смисла на AUC, како функција на големината на примерокот (лев панел) и на пропорционалниот праг што се применува на мрежата (десен панел).	81
4-5	Интра-стабилност D за секој метод.	83
4-6	Интер-стабилност на секој пар методи.	84
4-7	Графички приказ на врските кои се заеднички и ексклузивни за трите најдобри методи според AUC.	85
4-8	Пресметковна анализа.	86
4-9	Најдобри оценки за класификација, во однос на метриката F1, за податочните множества PRURIM (лев панел) и PSP (десен панел).	87
4-10	Интра-стабилност D од секој метод, за податочните множества PRURIM (лев панел) и PSP (десен панел).	87
5-1	Почетен поглед на веб-системот за мрежни вгнездувања за логирање.	98
5-2	Поглед на веб-системот за мрежни вгнездувања за граф трансформации со сите опции откако ќе се логира корисникот.	98
5-3	Поглед на веб-системот за мрежни вгнездувања за граф трансформации за преземање на векторската репрезентација.	99
5-4	Поглед на веб-системот за мрежни вгнездувања за реконструкција на граф за избор на хиперпараметрите на алгоритмот HOPE. . .	99
5-5	Поглед на веб-системот за мрежни вгнездувања за реконструкција на граф со внесена мозочна мрежа.	100

A-1	ABIDE, AAL 1mm, τ , контролни субјекти.	107
A-2	ABIDE, AAL 1mm, l_d , контролни субјекти.	108
A-3	ABIDE, AAL 1mm, τ , пациенти.	109
A-4	ABIDE, AAL 1mm, l_d , пациенти.	110
A-5	ABIDE, AAL 2mm, τ , контролни субјекти.	111
A-6	ABIDE, AAL 2mm, l_d , контролни субјекти.	112
A-7	ABIDE, AAL 2mm, τ , пациенти.	113
A-8	ABIDE, AAL 2mm, l_d , пациенти.	114
A-9	ABIDE, Brainnetome, τ , контролни субјекти.	115
A-10	ABIDE, Brainnetome, l_d , контролни субјекти.	116
A-11	ABIDE, Brainnetome, τ , пациенти.	117
A-12	ABIDE, Brainnetome, l_d , пациенти.	118
A-13	ABIDE, HO, τ , контролни субјекти.	119
A-14	ABIDE, HO, l_d , контролни субјекти.	120
A-15	ABIDE, HO, τ , пациенти.	121
A-16	ABIDE, HO, l_d , пациенти.	122
A-17	ABIDE, AAL 1mm, τ , контролни субјекти <i>vs.</i> пациенти.	123
A-18	ABIDE, AAL 1mm, l_d , контролни субјекти <i>vs.</i> пациенти.	124
A-19	ABIDE, AAL 2mm, τ , контролни субјекти <i>vs.</i> пациенти.	125
A-20	ABIDE, AAL 2mm, l_d , контролни субјекти <i>vs.</i> пациенти.	126
A-21	ABIDE, Brainnetome, τ , контролни субјекти <i>vs.</i> пациенти.	127
A-22	ABIDE, Brainnetome, l_d , контролни субјекти <i>vs.</i> пациенти.	128
A-23	ABIDE, HO, τ , контролни субјекти <i>vs.</i> пациенти.	129
A-24	ABIDE, HO, l_d , контролни субјекти <i>vs.</i> пациенти.	130
A-25	ABIDE, Сите атласи, τ , контролни субјекти.	131
A-26	ABIDE, Сите атласи, l_d , контролни субјекти.	132
A-27	ABIDE, Сите атласи, τ , пациенти.	133
A-28	ABIDE, Сите атласи, l_d , пациенти.	134
A-29	COBRE, AAL 2mm, τ , контролни субјекти.	135
A-30	COBRE, AAL 2mm, l_d , контролни субјекти.	136
A-31	COBRE, AAL 2mm, τ , пациенти.	137
A-32	COBRE, AAL 2mm, l_d , пациенти.	138
A-33	COBRE, AAL 2mm, τ , контролни субјекти <i>vs.</i> пациенти.	139
A-34	COBRE, AAL 2mm, l_d , контролни субјекти <i>vs.</i> пациенти.	140

Листа на табели

3.1	Главни карактеристики на податочните множества.	43
4.1	Рангирање на најдобрите 5 методи, според нивната AUC, стабилност D и пресметковна цена за COBRE податочното множество. .	88
A.1	Синтеза на сите дополнителни слики вклучени овде.	106
A.2	Листа на субјекти од ABIDE II податочното множество влучени овде.	141

Глава 1

Вовед во мозочни мрежи

Човечкиот мозок се смета за едно од најкомплексните нешта на светот. Како објект на испитување, мозокот претставувал енигма за истражувачите со векови. Долго време биле развивани различни техники и теории за разбирање на мозокот, почнувајќи од груба посмртна анализа, како и развој на соодветна апаратура за подобро анализирање на посмртни остатоци, па сè до развојот на не-инвазивни техники кои овозможуваат анализа на мозокот од живи субјекти. Современата наука нуди можности за детално проучување на мозокот на живи субјекти со прибирање на различни типови на податоци, најчесто во облик на визуелни снимки, за различни состојби на субјектите. Невроснимањето (neuroimaging) или снимање на мозокот вклучува употреба на различни техники за директно или индиректно снимање на структурата и функцијата на нервниот систем и е релативно нова дисциплина во медицината, невронауките и психологијата. Целокупниот напредок на ова подрачје доведува до развој и прифаќање на една нова теорија, според која човечкиот мозок претставува мрежа со неверојатна комплексност, составена од милијарди неврони меѓусебно поврзани со уште поголем број синапси. Начинот на кој е поврзана мозочната мрежа е од голема важност за сите функции на мозокот. Клучна парадигма во испитувањето на поврзаноста во мозокот е таканаречениот „конектом“, што претставува детален опис за тоа како неврните и регионите во мозокот се меѓусебно поврзани [1, 2, 3, 4, 5]. Невроснимањето може да се подели на две големи категории: Структурно снимање (Diffusion Tensor Imaging - DTI), кое се занимава со структурата на нервниот систем од кое произлегува „конектот“ на структурната поврзаност [4]; и функционално снимање (functional Magnetic Resonance Imaging - fMRI), кое се занимава со функционалната зависност на компонентите во нервниот систем од кое произлегува „конектот“ на функционалната поврзаност [4].

Графовите се природен начин на претставување на резултатите дадени со

анализа на невровизуелизација. Графовите (мрежите) се важни структури за репрезентација на податоците што се користат во широк спектар на реални апликации, меѓу кои и репрезентацијата на поврзаноста помеѓу регионите во мозокот. Еден граф G може да се дефинира како подредено множество $G = (V, E)$, каде што V е множество на јазли, а E е множество на рабови. Јазлите на графот се дискретни објекти, додека рабовите се асоцијации на два објекта на јазли. Многу проблеми што опфаќаат широк спектар на научни области може да се дефинираат со користење на концептот на мрежи на меѓусебно поврзани објекти, па затоа, застапеноста на проблемот со користење на графови е многу разновидна. Во биологијата, примери за користење на графовидни претстави може да се видат, освен во мозочните мрежи [6] се и мрежите на интеракција на протеините [7]. Во лингвистиката, графовите се користат во конструкцијата на семантички мрежи [8] и во машински превод на јазиците [9]. Во социологијата, графовите се основа за претставување на социјални мрежи [10] и колаборативни мрежи помеѓу луѓе во некоја професија [11]. Во компјутерската наука, тие се користат во широк спектар на проблеми, од моделирање и анализа на комуникациски мрежи [12], до податочно рударење [13][14]. Овие примери опфаќаат само дел од проблемите во дадените области, и само научни области кои ги користат графовите почесто.

Во контекст на значителен напредок на проучување на процесите во мозокот доприносот на теоријата на графови и функционалните невроснимки е неоспорлив. Ваквите снимки овозможуваат креирање на одредени асоцијации помеѓу регионите во мозокот и состојбата (мирување или активност) во која се наоѓа субјектот кој е предмет на снимање. Покрај нормалното асоцирање на процесите во мозокот со одредени активности, постојат останати, можеби уште позначајни, примени на функционалните снимки. Една од најважните примени е запознавање со невронската позадина за многу нарушувања во нормалното функционирање на мозокот. Идејата е дека таквите снимки може да се искористат за детектирање на активности во одредени делови од мозокот кои се различни од нормалната состојба. Вообичаено тоа значи откривање на невообичаена (различна од нормалната) комуникација помеѓу различни делови од мозокот, како и одредување на регионите во мозокот кои се активираат при различни активности. Според голем број истражувачи, постојат одредени шаблони на абнормална активност кои карактеризираат одредено ментално нарушување [15, 16]. Со ваквиот научен пристап се овозможува не само дијагноза на болеста, туку и следење на ефективноста на лекувањето.

Менталните нарушувања претставуваат голема и разнолика група од состојби кои влијаат на однесувањето на личностите. Некои од најчестите ментални

нарушувања се: биполарно нарушување, шизофренија, депресија, опсесивно компулсивно нарушување, аутизам итн. Симптомите на менталните нарушувања се разнолики во зависност од условите, што во суштина значи дека различни луѓе може да имаат различни манифестации на истите ментални нарушувања. Лекувањето на нарушувањата поврзани со мозокот се различни во зависност од нарушувањето. Дел од нив се третираат со медикаментозна и психотерапија (што е карактеристично за ментални болести), додека други (како што се некои невродегенеративни болести и трауматични мозочни повреди) подразбираат третман кој и помога на болната личност да живее со болеста и да одржува нормален живот, толку колку што е можно.

Сите истражувања кои се прават, лесниот начин за прибирање на снимките и подобрувањата во технологијата, ги прават функционалните невроснимки добра алатка за дијагностика. Но, еден голем проблем што постои кај ваквиот пристап е веродостојноста на методологијата за примена на снимките во дијагностицирањето на одредена болест. Овој проблем се јавува поради лажните позитиви, кои претставуваат резултати што покажуваат присуство на одредена состојба, која всушност не постои. За негово решавање се применуваат различни статистички методи [17] кои дозволуваат дополнителна анализа на податоците и резултатите, како и крајна верификација. Со тоа, полето на статистика игра интегрална улога во напредокот на истражувањата на функционално невроснимање и има потенцијал значително да ја зголеми значајноста на техниките на невроснимање во клиничката пракса.

Во овој докторски труд главниот фокус ќе биде на структурната и функционалната поврзаност како методи кои се добро проучени [18], развиени и применувани врз целосни нервни системи. Една заедничка работа за сите типови на поврзаност е идејата дека целиот нервен систем може да се претстави како мрежа. Идејата дека мозокот може да се набљудува како мрежен систем не е нова, но примената на теориите од математички мрежи за анализирање на карактеристиките на мозочниот систем е релативно нов пристап. Структурата на графовите или мрежите може да се анализира со помош на голем број мерки од теорија на графови за комплексни мрежи. Мрежите може да се карактеризираат на различни нивоа, од карактеристики на целата мрежа на глобално ниво, па сè до карактеристики на мрежни компоненти на локално ниво [19]. Во зависност од анализата и типот на технологија која се користи, возможно е да се генерираат мрежи кои содржат различни типови на информации. Различните типови на мрежи, може да се комбинираат за градење на покомплексни модели од кои ќе се извлекува знаење за регионите и подмрежите на мозокот кои се зафатени

при различни услови.

За мрежите на функционална поврзаност во овој труд ќе се разгледува техника на функционалната поврзаност во мирување (resting state functional Magnetic Resonance Imaging – rs-fMRI) која го мери сигналот кој зависи од нивото на кислород во крвта (BOLD) [20, 21, 22, 23], заедно со најчесто користените алатки кои ги анализираат податоците генерирани од таа техника [24, 25]. Голем број на истражувања своето внимание го насочуваат кон спонтаните мозочни активности за време на ментален одмор, односно активности кои не се поврзани со некаков посебен стимул или однесување [23, 26, 27, 28].

Со употреба на оваа техника може да се набљудува мозочната мрежа во предефинирана состојба (Default Mode Network - DMN) што е најактивна додека субјектите се во состојба на мирување и не се вклучени во никаква активност или задача [29]. Ако субјектите почнат да прават некакви активности или задачи, тогаш оваа мрежа се деактивира. Голем дел од областите во мозокот се вклучени во мозочната мрежа во предефинирана состојба. Бидејќи DMN е најактивна во состојба на мирување и поради структурите кои се вклучени во неа, во некои студии заклучено е дека таа е поврзана со активности како мечтаење или навраќање на некои спомени [30, 31]. Други студии пак, предлагаат дека активноста е поврзана со физиолошки процеси независни од некаква друга специфична активност [31]. Промените во DMN се поврзуваат со голем број болести, како што се Алцхајмеровата болест [32], аутизмот [33], шизофренијата [34], биполарното нарушување [28], како и депресијата [35]. Овие болести може да предизвикаат премала или преголема активност во мозочните региони на DMN.

Од досегашните истражувања во оваа област, посебно се истакнуваат неколку трудови кои имаат импакт во мотивацијата на истражувањата во оваа докторска дисертација. Во [36, 37] е искористен мрежен пристап за покажување на хиерархиската структура на активностите во мозокот во состојба на мирување за здрави субјекти со употреба на кластерирачки алгоритми. Во [38] се истражуваат разликите на особините на “мал свет” на мозочната мрежа [39] од fMRI податоци при ментален одмор меѓу здрави субјекти и субјекти со биполарно нарушување. Преглед на техниките од машинско учење кои се користат кај мозочни мрежи е даден во [40], а додека пак во [41], е направен преглед на техники од длабоко учење кои може да се искористат за анализа на мозочни мрежи. Во [42] е предложена архитектура со конволуциска невронска мрежа за извлекувања на карактеристики од мрежата. Преглед на различните статистички тестови кои се користат при споредба на две групи мозочни мрежи е даден во [43]. Алгоритми за мрежни вгнездувања при извлекување на карактеристики од мозочната мре-

жа се користат во [44]. Карактеристично за сите овие истражувања е тоа што кај сите се користи само по еден тип на мрежна поврзаност. Постојат многу малку истражувања кои ги интегрираат двата типови (модалитети) на нервно-снимање во мрежи од кои може да се добијат заклучоци за различни невродегенеративни болести [45, 46] што е главен мотив во ова истражување.

Општиот предмет на истражувањето во оваа докторска дисертација ќе бидат мрежите на мозокот кои се добиваат од податоци за субјекти со некое нарушување на функцијата на мозокот. Подетално во овој труд ќе се разгледуваат различните типови на мрежи извлечени од снимки на субјекти со мозочни нарушувања. Од добиените мрежи ќе се извлекува знаење за субјектите со употреба на предложена методологија.

Во оваа дисертација ќе се евалуираат разликите кај мозочната поврзаност со помош на различни репрезентации на мозочни мрежи кај субјекти кои имаат одредено мозочно нарушување во однос на здрави субјекти. За таа цел предмет на дисертацијата ќе биде обработката и анализата на податоците добиени од различни модалитети на нервно-снимање. За да се добијат веродостојни резултати од анализата на податоците ќе се направи претпроцесирање на податоците за да се отстранат артефактите и шумовите кои постојат во нив. После претпроцесирањето ќе се утврдат различните типови на мозочна поврзаност кај секој субјект и ќе се креираат модели за различна репрезентација на мрежната поврзаност. Со примена на најсовремени граф базирани методи од комплексни мрежи и статистички методи врз различните репрезентации ќе се утврдат разликите што постојат кај различните типови нарушувања во различни делови од мозокот. Може да се направи и невролошко толкување за овие разлики и за издвојување можни биомаркери за болестите. Вака добиените разлики и биомаркери служат за дизајн на класификатори кои ќе прават автоматска класификација на субјектите според нивното нарушување.

По делови во оваа докторска дисертација ќе се осврнеме на методи за конструкција и анализа на мозочни мрежи од повеќе аспекти. Прво во Глава 2, ги реконструираме функционалните мозочни мрежи со комбинирање на различни методи за ренормализација на врската, засновани и на фреквентистички и на баесови пристапи кои ги предлагаме, и на стратегии за праг. Потоа оценуваме до кој степен се засегнати својствата на добиената мрежа и понатаму ги користиме за да разликуваме субјекти кои страдаат од невролошки нарушувања и соодветна контролна група. Во Глава 3, покажуваме дека разликите во рангирањето на врските предизвикуваат специфични геометрии во еден помошен простор што честопати лесно се препознаваат при обична визуелна проверка. Во Глава 4

разгледуваме и големо множество на статистички методи и методи за избор на врски за машинско учење и ги проценуваме на реалните функционални мозочни мрежи. На крај, во Глава 5 изработуваме и веб-базиран систем за обработка на мозочни мрежи со различни алгоритми за т.н. мрежно вгнездување, а потоа во Глава 6 правиме сумаризација на заклучоците.

Глава 2

Реконструкција на мозочни мрежи

Невронауката и електрофизиолошките студии објаснуваат дека правилното извршување на голем број когнитивни задачи зависи од динамичката поврзаност помеѓу регионите на мозокот [47, 48, 49, 50, 51, 52]. Динамичката поврзаност дава мрежна структура како резултат, со нетривијални тополошки својства [18]. Овие својства се модулирани од извршувањето на когнитивните задачи [53, 54], во различни фази на човековиот развој [55] и процесот на стареење [56, 57], што укажува на улогата на мрежната топологија во функционалната активност на мозокот.

Реконструкцијата на мозочната мрежа од невросликите е крајно нетривијална задача што вклучува многу чекори, од дискретизирање на просторот и избор на метрики за мрежна сврзаност до толкување на значењето на тополошките својства [58, 59]. Мрежата, во суштина, е збирка на јазли поврзани со рабови. Теоријата на графови овозможува извлекување на многу својства што ги поседува мрежата, но е неутрална за тоа како треба да се дефинираат и јазлите и рабовите.

Мрежните јазли се идентификуваат преку системот што разгледува дадена карактеристика и некоја функција за сврзаност. Изборот на просторот со мрежна структура е дискретен, како и карактеристиката што се користи. Првиот е типично анатомски простор во кој се одвива снимената динамика. Изборот на карактеристиката што се користи е попроблематичен. Уредот што се користи за снимање на мозочната активност (на пример, електро- и магнето-енцефалографија) дава природна микроскопска скала и затоа дава основна поделба на мозочниот простор. Додека за неинвазивните електрофизиолошки техники, микроскопската скала е добиена од сензори и главното прашање е колку добро сензорите го земаат примерокот од основниот динамички систем. За функционална магнетна резонанца (fMRI) ова е поврзано со просторната резолуција на уредот за скенирање,

наречена воксел, што предизвикува природна поделба на анатомскиот простор. Типичен fMRI протокол на целиот мозок дава $\approx 10^6$ воксели. За мрежите со толкав голем број на единици, пресметувањето на тополошките својства од повисок ред може да биде пресметковен предизвик, а и толкувањето на поврзаноста или дури и обезбедувањето графичка претстава може да се покаже како макотршна задача. За да се справиме со овие задачи, честопати е потребен критериум за намалување на бројот на јазли. Ова вклучува разграничување на функционално одделните единици на мозокот, задача што оди под името *парцелација* [60].

На скала која е типична за стандардните неинвазивни невро-слики на ниво на систем, дефиницијата за овие функционални единици е многу помалку интуитивна отколку на пример кај единечните неврони, каде што единиците се јасно дефинирани, и јазлите треба да се идентификуваат на поинаков начин. Ова често вклучува одредена форма на функционална проекција на анатомскиот простор, комплексна и слабо разбирлива мапа, бидејќи набљудуваните динамички темплејти на активност на мозокот резултираат во просторно и временски нелокализиран начин од мозочна поврзаност од сите размери [61]. Парцелацијата бара формирање макројазли, кои се нарекуваат региони на интерес (ROI). Дефинирањето на региони на интерес вклучува комплексен процес на ренормализација на јазлите, кој се состои од: *i*) генерални критериуми за конструкција како правила засновани на анатомија, прибегнувајќи на пример на достапни атласи или на динамички базирани; *ii*) правила за членство, на пр. топографска локализација во анатомскиот простор или, статистички критериуми, како на пр. во методите за кластерирање [62]; *iii*) правила за партиција на простор, на пр. стриктна или нејасна партиција [63], или со преклопување [64] на парцели или правила за големина, на пр. регионите на интерес можат да бидат претставени со еднакво големи мозочни региони; *iv*) геометриски или тополошки метаправила, на пр. поврзаност, компактност или раздвојливост.

Помалку внимание обично се посветува на ренормализација на линковите во мрежната структура. Во својата стандардна форма, линковите во граф поврзуваат суштински јазли. Сепак, за разлика од стандардните јазли во графот, регионите на интерес се просторно проширени. Мрежно-базираното разгледување на анатомскиот простор поделен на региони на интерес, може да заземе структура на заедница или пристап на мрежата на мрежи, при што секој регион на интерес се третира како подмрежа или како посебна мрежа. Сепак, решавањето на целта за намалување на димензиите бара претварање на регионите на интерес во еквивалентни јазли. Постојат, во принцип, повеќе начини на кои може да се сумира или дефинира регион на интерес, од просек на временската серија од сите

воксели во дадена анатомска област, до разгледување на центарот на масата на разгледуваниот регион [60]. За крај, ренормализацијата на поврзаноста е функција од начинот на дефинирање на регионот на интерес, т.е. интеракциите со регионите надвор од него зависат од начинот на кој се пресметува структурата во него. Ова е особено релевантно бидејќи правилата за изградба на регион на интерес не треба, и, генерално, не доведуваат до хомогени региони со едноставна внатрешна структура [60, 65]. Улогата на оваа внатрешна структура може да биде пресметана на различни начини, на пр. со проценка на просечната корелација помеѓу вокселите на два различни региона на интерес [66]. Исто така, коефициентите на корелација што се добиваат може да се усреднат на различни начини, т.е. со класични фреквенционистички или баесови пристапи.

Во отсуство на множество принципиелни критериуми за водење на процесот на реконструкција на мрежата, едно фундаментално прашање е разбирањето дали тополошките својства на реконструираниите мрежи се суштински својства на системот, или барем, се робусни во однос на начинот на реконструкција на мрежите од експериментални податоци [67, 59, 60, 58]. Покажано е дека методите за реконструкција и постапките за ренормализација потенцијално квалитативно влијаат на тополошките својства [68, 69]. Особено, една претходна студија за ренормализација на јазлите сугерира дека глобалните тополошки својства, како што е метриката за мал свет, можат да бидат робустни во однос на техниката за парцелација и вкупниот број на јазли, иако квантитативниот аспект на овие својства може да биде многу малку засегнат [67].

Во оваа дисертација, тестираме до кој степен влијаат различните техники за ренормализација на врската врз својствата на тополошката мрежа. Затоа, ги реконструираме нетежинските функционални мозочни мрежи на групи испитаници со дијагностициран аутизам и шизофренија и на соодветни контролни групи, со четири различни процедури за ренормализација на врската. Последниот Баесов пристап за реконструкција на мозочните мрежи е една од предложените методи во оваа дисертација како придобивка на истата. Потоа, ги мериме квантитативните различни стандардни мрежни тополошки својства и како тие се погодени од избраната постапка за ренормализација. Исто така, се оценуваат влијанието на изборот на праг и изборот на атлас.

Остатокот од текстот е организиран на следниов начин. Сек. 2.1 претставува преглед на податочните множества што се користат во овој дел од дисертацијата, заедно со процесот на нивно предпроцесирање, како и методите за реконструкција и ренормализација на мозочните мрежи од нив. Резултатите се презентирани во сек. 2.2, а на крај, сек. 2.3 дава дискусија на резултатите и извлекува некои

заклучоци од истражувањето.

Оваа глава од дисертацијата е во процес на објавување во трудот Plinka Ivanoska, David Papo, Onerva Korhonen, Javier Cano and Massimiliano Zanin. “From voxels to ROIs: link renormalisation in fMRI functional networks.” Under review at Human Brain Mapping, 2021. IF 4.421.

2.1 Методи и податоци

2.1.1 Податочни множества и претпроцесирање

За оваа студија беа искористени две податочни множества, кои се јавно достапни. Првото е избрано подмножество на fMRI податоци за пациенти со аутистичен спектар (АСД) ABIDE II [70]. Избрани се 38 субјекти (просечна возраст и стандардно отстапување од $10,2 \pm 1,5$ години) од збирката на Кенеди Кригер (ККИ) Институтот [71]. Со цел да се обезбеди хомогеност во податоците, беа избрани 19 пациенти со АСД со комбинирано нарушување на хиперактивноста со дефицит на внимание (АДХД), заедно со ист број на соодветни контролни субјекти (со најблизок IQ и возраст). Списокот на избрани субјекти е вклучен во делот со Додадок.

Невросликите за подмножеството ABIDE II се добиени на еден од двата скенери 3T Филипс (Phillips Healthcare, Бест, Холандија) користејќи или 8-канален ($N = 28$) или 32-канална ($N = 10$) фазна секвенцијална намотка. Тие вклучуваат анатомско MRI T1-тежинско скенирање (MPRAGE) ($TR = 8300$ ms, $TE = 3,7$ ms, $FA = 9^\circ$, $FOV = 256$ mm, големина на воксел $1 \times 1 \times 1$ mm³) и функционално MRI во состојба на мирување (rs-fMRI) ($TR = 2500$ ms, $TE = 30$ ms, $FA = 90^\circ$, $FOV = 224$ mm, големина на воксел $3,5 \times 3,5 \times 3,5$ mm³) [71].

Второто податочно множество што се користи во оваа дисертација е COBRE [72], кое опфаќа податоци за пациенти со шизофренија (SZ) и здрави контролни субјекти. Избрани се 144 субјекти (просечна возраст \pm с.д. од $36,7 \pm 12,6$; 70 SZ и 74 контроли); од кои два субјекти од оригиналното множество беа отфрлени заради грешки во временските серии. Податоците се стекнати на 3T Siemens Trio скенер и се состојат од анатомски слики како мулти-ехо MPRAGE секвенца ($TR / TE / TI = 2530 / [1,64, 3,5, 5,36, 7,22, 9,08] / 900$ ms, $FOV = 256 \times 256$ mm, $FA = 7^\circ$, големина на воксел $1 \times 1 \times 1$ mm³ [72]) и rs-fMRI податоци добиени со користење на едноснимена целосна слика на k-простор (EPI) ($TR = 2$ s, $TE = 29$ ms, $FA = 75^\circ$; $FOV = 240$ mm, големина на матрица = 64×64 , големина на воксел: $3,75 \times 3,75 \times 4,55$ mm³ [72]).

2.1.2 Претпроцесирање на податочно множество

Пред-обработката на податоците за fMRI е извршена со користење на статистичко параметричко мапирање (SPM12) [73] во MATLAB R2018b со верзијата 18b со употреба на алатката CONN [74]. Чекорите за пред-обработка вклучуваат стандардни чекори со преуредување и одвиткување, корекција на временската серија, откривање на outlier-и со ART [75], сегментација на сива материја, бела материја и цереброспинална течност (CSF) за отстранување на временски збунувачки фактори, нормализирање на сликите до стандардниот образец на невролошкиот институт во Монреал (MNI) и измазнување со целосна ширина од 8мм на половина кернел (FWHM). Извршен е чекор за отстранување на шум со алатката CONN, вклучително и корекција на збунувачките ефекти на белата маса и CSF [76]. Конечно, беше извршено филтрирање со пропусен опсег со прозорец со фреквенција од 0,008-0,09 Hz. Анатомскиот атлас (AAL) [77] (и шаблоните со резолуција од 1мм и 2мм), атласот Харвард-Оксфорд (HO) се искористени како шаблони за парцелација на мозокот во региони од интерес, соодветно анатомски (AAL, HO) и врз основа на поврзување (Brainnetome).

За секој регион на интерес, временските серии на воксели беа извлечени со помош на *nilearn* Python библиотека [78]. Потоа, функционалните мозочни мрежи (ROI-to-ROI конектоми) за секој субјект беа пресметани со користење на сите пристапи на методите за реконструкција на функционалните конектоми, објаснети подолу.

2.1.3 Методи за реконструкција на функционални мозочни мрежи

По претходно опишаниот чекор за пред-обработка, секој субјект е опишан со множество временска серија $X_i(t)$, каде i е индекс што итерира низ сите достапни воксели. Бидејќи бројот на воксели е од ред на милион, реконструкцијата на комплексната мрежа директно од X е пресметковно предизвикувачка, а резултатот е тежок за толкување. Затоа, вообичаено е динамиката да се примени на множество од $Y_i(t)$, каде што секоја временска серија Y ја синтетизира динамиката во еден ROI дефиниран според специфичен атлас. Овие временски серии потоа се користат за пресметување на тежинска матрица на поврзување W со големина $N \times N$, каде N е бројот на региони на интерес.

Во оваа дисертација разгледуваме и споредуваме четири различни начини на реконструкција на W .

Корелација на средни вредности (Correlations average CA)

Наједноставниот метод за реконструкција вклучува пресметување на $Y_i(t)$ како просек од сите временски серии $X_j(t)$ што припаѓаат на i -иот ROI. $w_{i,j}$ потоа се пресметува како апсолутна вредност на коефициентот на линеарна корелација помеѓу временските серии Y_i и Y_j .

Средна вредност на корелации (Average of correlations AC)

Вториот метод се базира на концептот на конзистентност на динамиката на регионите од интерес, како што е предложено во [66, 79]. Поточно, забележано е дека два региони на интерес може да имаат голема корелација помеѓу нивната динамика, кога вторите се добиваат преку едноставен просек од соодветните $X(t)$, иако нивната внатрешна динамика е слабо конзистентна. Со други зборови, просечната динамика на воксел може да воведо лажни корелации на ниво на регион на интерес. За да се избегне овој проблем, прво се пресметуваат сите корелации по парови помеѓу временските серии што одговараат на два региона на интерес i и j . После тоа, $w_{i,j}$ се пресметува како просек на апсолутната вредност на овие корелации. Забележете дека пресметковната цена на овој втор метод е значително поголема од онаа на првиот метод: наместо да се пресмета една единствена корелација, треба да се проценат $K_i \times L_j$ корелации, каде што K_i и L_j се соодветно бројот на воксели во регионот од интерес i и j .

Средна вредност на корелации со Fisher Z-трансформација (ACZ)

Просечните коефициенти на корелација воведуваат пристрасност, бидејќи распределбата на овие коефициенти не е нормална [80]. За да се избегне ваквата пристрасност, овој трет метод ги следи истите чекори како и претходниот; сепак, наместо да се пресмета просекот на коефициентите на корелација r , ова се трансформираат со Fisher-Z трансформација:

$$z = \frac{1}{2} \ln \frac{1+r}{1-r} = \operatorname{arctanh}(r). \quad (2.1)$$

Потоа, сите z s се усреднуваат, а резултатот конечно е обратно трансформиран во интервалот $[-1, 1]$.

Баесов метод (BM)

Овој последен метод кој го предлагаме во оваа дисертација се базира на Баесов заклучок. Следниве чекори се потребни за да се пресмета елементот w_{ij} во

матрицата на конективност W . За секој пар на воксели (k, l) , $k = 1, \dots, K_i$, $l = 1, \dots, L_j$ кои припаѓаат на i -иот и j -иот ROI, соодветно, ги користиме соодветните временски серии $X_k(t)$ и $X_l(t)$ за да ја пресметаме posterior веројатносна распределба на коефициентот на корелација на популацијата $\rho_{k,l}$, следејќи ја постапката опишана во [81]. За да провериме дали анализираните fMRI податоци ја исполнуваат потребната претпоставка за биваријантна нормалност, го извршивме тестот Henze-Zirkler [82] за секој пар временски серии од податочното множество, добивајќи нормалност за 92,7% од нив ($\alpha = 0,001$). Понатаму, заради едноставност, не претпоставуваме претходно знаење за параметрите на основниот биваријантен нормален модел, иако е едноставно (сепак евентуално пресметковно захтевно) да се вклучат сите претходни достапни информации во форма на претходни распределби на вклучените параметри.

Откако $K_i \times L_j$ posterior распределби $p(\rho_{11}|\text{data}), p(\rho_{12}|\text{data}), \dots, p(\rho_{K_i, L_j}|\text{data})$ се добиени, ги комбинираме преку нивната мешавина

$$p(\rho_{ij}|\text{data}) = \frac{1}{K_i \times L_j} \sum_{k=1}^{K_i} \sum_{l=1}^{L_j} p(\rho_{kl}|\text{data}),$$

каде што на сите компоненти на мешавината $p(\rho_{kl}|\text{data})$ им се доделени еднакви тежини. Важно е да се напомене дека таквата мешавина ги задржува сите достапни информации за поврзаноста помеѓу кој било пар воксели што припаѓаат на регионите на интерес i и j . Потоа, семплираме доволно голем број на случајни вредности $\{r_{ij}\}_{m=1}^M$ (тука, $M = 100$) од posterior распределбата $p(\rho_{ij}|\text{data})$, и ги сумаризираме преку средната вредност на нивните апсолутни вредности за да добиме проценка на поврзаноста помеѓу регионите на интерес i и j :

$$w_{ij} = \frac{1}{M} \sum_{m=1}^M |r_{ij}|.$$

2.1.4 Анализа на функционални мозочни мрежи

Тежинските целосно поврзани графови W добиени во претходниот чекор обично се трансформираат во бинарни и ретки мрежи. Потоа, оваа мрежа е целосно дефинирана со матрица за соседство A со големина $N \times N$, каде што N е бројот на јазли (овде, ROI), со елементот $a_{i,j}$ еднаков на еден ако постои врска помеѓу јазлите i и j (и $a_{i,j} = 0$ инаку). Потребата за бинаризација на тежинските мрежи произлегува од фактот дека повеќето граф тополошки метрики достапни во литературата бараат мрежи без тежина; работата со тежина би го намалила

спектарот на достапни анализи [83].

Второ, ретките мрежи се претпочитаат, бидејќи вклучувањето на најслабите и со тоа потенцијално помалку веродостојни врски може да воведат пристрасност во последователните метрики [84]. Ние овде користиме два вообичаени пристапа за бинаризација на мрежата, т.е. *апсолутно* и *пропорционалното ограничување на прагот*. Во првиот пристап, сите врски што надминуваат апсолутен праг τ се задржани. Иако е едноставен и интуитивен, овој метод дава мрежи со различен број врски помеѓу субјектите и што е најважно, низ групите (на пример, помеѓу контролните субјекти и пациентите). Ова може да доведе до статистички значајни разлики во мрежните метрики помеѓу двете групи, дури и кога тие не се должат на основните тополошки разлики поврзани со болеста. Како таков, се претпоставува дека овој пристап е понеповолен за студии за контрола на случаи [85, 86]. Вториот пристап делумно го надминува ова прашање, со вклучување во секоја мрежа фиксен број на најсилни врски [87]. Овој пристап честопати се нарекува во литературата како анализа во која *густината* [88] или *мрежната цена* [89] се одржува константна.

Откако ќе се добијат бинаризираните мрежи, од нив се извлекуваат збир од осум тополошки метрики, избрани заради нивната важност во обработката на информации за мозокот [18, 90]:

- *Кластерирачки коефициент (Clustering coefficient CC)*. Кластерирачкиот коефициент, исто така познат како *транзитивност*, го мери присуството на триаголници во мрежата [91]. Дефиниран е како врска помеѓу бројот на триаголници во мрежата (т.е. множества од три темиња со рабови помеѓу секој пар од нив) и бројот на поврзани тројки (т.е. множества од три темиња каде што може да се достигне секое теме едно од друго, директно или индиректно).
- *Ефикасност (Efficiency Eff)*. Ефикасноста на мрежата претставува колку лесно може да се движат информациите помеѓу нејзините јазли [92]. Се дефинира како инверзна на хармоничната средина на растојанијата помеѓу паровите на јазли:

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{i,j}}, \quad (2.2)$$

$d_{i,j}$ е растојанието меѓу i и j .

- *Асортативност (Assortativity Ass)*. Коефициент на корелација на Пирсон на степени на јазли на двата краја на врските на мрежата [93]. Позитивните вредности се поврзани со асортираните мрежи, во кои јазлите имаат

тенденција да се поврзуваат со нивните соседи за поврзаност; а од друга страна, негативните вредности се наоѓаат во неасортатираниите мрежи, каде што јазлите со низок степен се поверојатно поврзани со оние кои се многу поврзани.

- *Содржина на информации (Information content IC)*. Метричко оценување на присуството на правилности во матрицата на соседството на мрежата. Се пресметува како количина на информација кодирана во матрицата на соседството, така што малите вредности одговараат на регуларна топологија, а големите вредности на рандом структура [94].
- *Максимален степен (Max. degree MaxD)*. Степен (т.е. број на врски) на најповрзаниот јазол во мрежата.
- *Големина на гигантската компонента (Size of giant component GC)*. Големина на најголемата компонента на графот, т.е. на подмножеството на јазли поврзани едни со други со најмалку една патека.
- *Модуларност (Modularity Mod)*. Метричко оценување на присуството на структура на заедница во мрежата, т.е. групи на јазли густо поврзани меѓу нив, но лабаво поврзани со другите заедници [95, 96]. Модуларноста тука е пресметана со помош на добро познатиот Лувен алгоритам [97], поради неговите добри перформанси [98] особено во анализа на мозочните мрежи [99].
- *Вгнездување (Nestedness Nes)*. Метричко мерење на тенденцијата јазлите да комуницираат со подмножества на партнери за интеракција на подобро поврзани јазли [100].

2.2 Резултати

Сите резултати од овој дел за реконструкција на мозочни мрежи се додадени во Додатокот на оваа дисертација и вклучуваат еволуција на сите осум тополошки метрики, како функција на густината и прагот на врската, и за различни атласи и состојби на пациентите. Од друга страна, и заради синтеза, сл. 2-1 дава подмножество на тие метрики. Поточно, ги вклучува кластерирачкиот коефициент (леви панели) и ефикасноста (десни панели), двете најчесто изучувани метрики и понатаму основата на концептот за *мал-свет* [101, 102, 103, 39]. Дополнително, горните панели ја даваат еволуцијата на метриката како функција на прагот τ (според апсолутниот праг), а на долниот дел како функција на густината на врската (за пропорционален праг). Да се забележи дека X-оските на графовите што одговараат на апсолутниот праг се обратни, така што бројот на активни

врски секогаш се зголемува одлево надесно. Од овде може да се извлечат неколку интересни заклучоци.

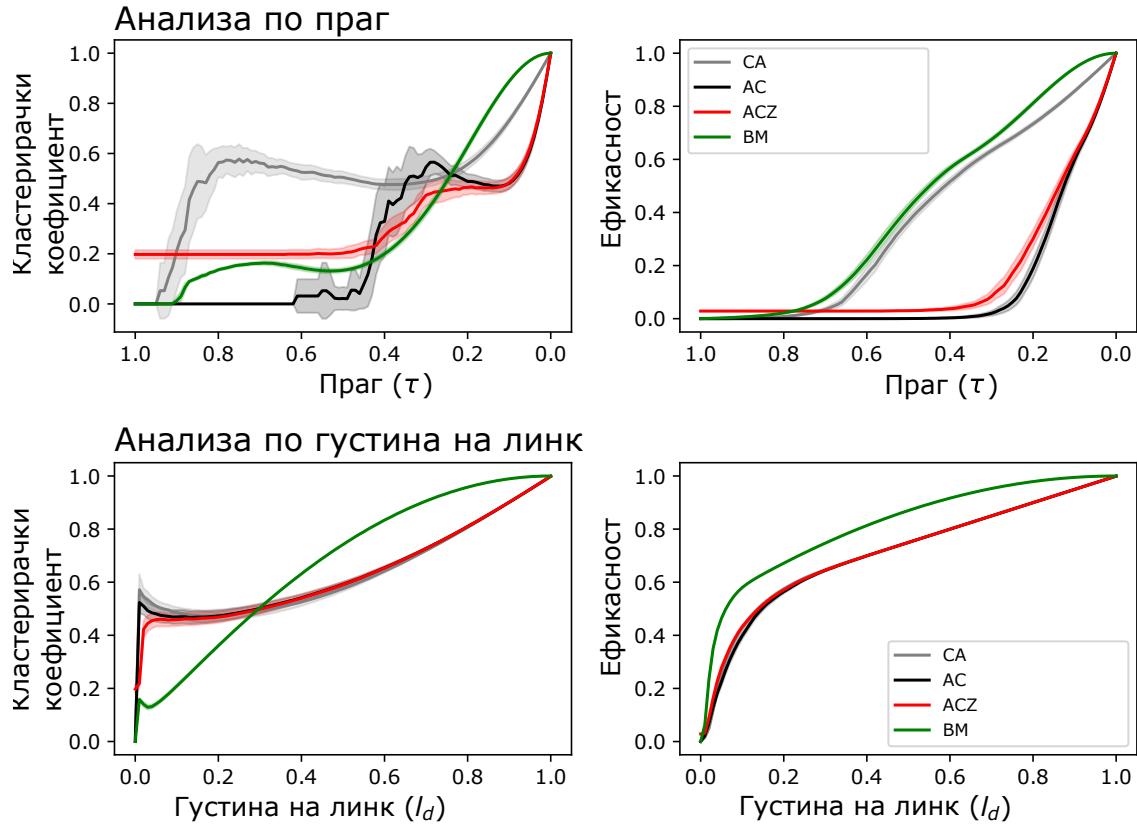
Најголемите разлики предизвикани од различните пристапи за ренормализација може да се забележат при користење на апсолутен праг (горните панели), додека метриците се постабилни во случај на пропорционалниот праг. Причината за ова може да се најде во различната распределба на тежините на врските во четирите методи. Како што е прикажано на сл. 2-2, лев панел, просечниот коефициент на корелација е значително помал во случај на просек на корелации (AC) и на Fisher Z-трансформација (ACZ). Исто така, од десниот панел на сл. 2-2 може да се процени дека првиот метод дава помали вредности од вториот, што објаснува зошто црната (AC) и црвената (ACZ) линија на сл. 2-1 се претежно под зелената (BM) и сивата (CA) - бидејќи се потребни вредности на понизок праг за да се достигнат истите тополошки метрики.

Додека обликот на еволуцијата на метриката е сличен во сите случаи за пропорционалниот праг, Баесовиот метод (BM) дава систематски различни вредности - видете ги долните панели на сл. 2-1. Затоа, иако употребата на кој било од трите фреквенционистички методи е во основа еквивалентна, баесовиот пристап дава поинаква гледна точка.

Разликите помеѓу методите за реконструкција силно зависат од разгледаната метрика. За илустрација, во случај на фиксен праг, корелацијата на просеците (CA) и Баесовиот (BM) метод се суштински еднакви за ефикасноста, но суштински отстапуваат за кластерирачкиот коефициент. Слични резултати даваат и двата методи и во случај на метрика на големина на гигантска компонента, модуларност и вгнездување - видете во Додаток.

Со цел да се исклучи дека претходно прикажаните резултати се должат само на карактеристиките на обработеното податочно множество, сл. 2-3 ја дава еволуцијата на истите метрики како на сл. 2-1 за целото податочно множество COBRE. Може да се процени дека разликите се минимални и во рамките на варијабилноста што се очекува во различни податочни множества.

Потоа го анализиравме влијанието на изборот на атлас врз претходно прикажаните резултати. За тоа, сл. 2-4 ја прикажува еволуцијата на кластерирачкиот коефициент и ефикасноста за корелацијата на методите на средна вредност (CA) и Баесовиот метод (BM) и за четирите разгледувани парцелации. Што се однесува до сл. 2-1, резултатите за сите тополошки метрики се вклучени во Додадокот со дополнителни информации. Разликите предизвикани од различни атласи се мали и во сите случаи се помали од разликите воведени со методите на реконструкција.

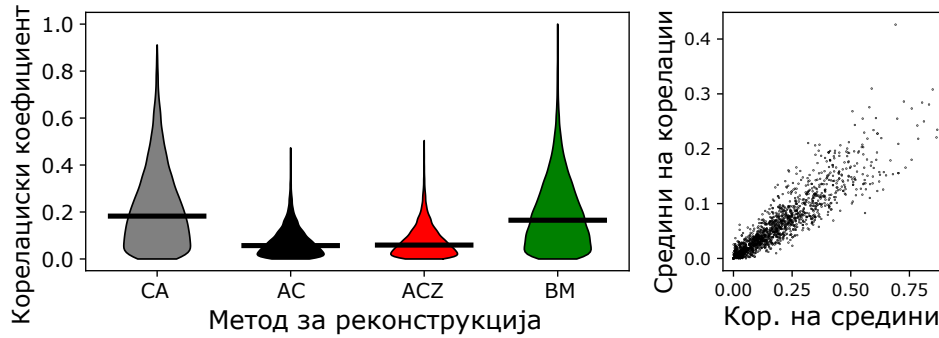


Слика 2-1: Еволуција на две тополошки метрики (кластерирачки коефициент, леви панели и ефикасност, десни панели) како функција на прагот τ (горните панели) и густината на врската l_d (долни панели). Четирите линии одговараат на четирите разгледани методи на реконструкција: корелација на средни вредности (сива, CA), средна вредност на корелации (црна, AC), Fisher-Z коригирана средната вредност на корелациите (црвена, ACZ) и Баесов метод (зелена, BM).

Понатаму оценуваме дали методот на реконструкција може да влијае на присуството на разлики помеѓу контролните субјекти и пациентите. Сл. 2-5 ја покажува еволуцијата на Коен d за четирите методи - како и во претходните случаи, тука се пријавени само резултати за кластерирачкиот коефициент и ефикасноста. Коенската d е најчесто користена метрика за големината на ефектот [104], дефинирана како $d = (\mu_1 - \mu_2)/s$, со μ_1 и μ_2 средните вредности забележани во двете групи и s стандардната девијација:

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}, \quad (2.3)$$

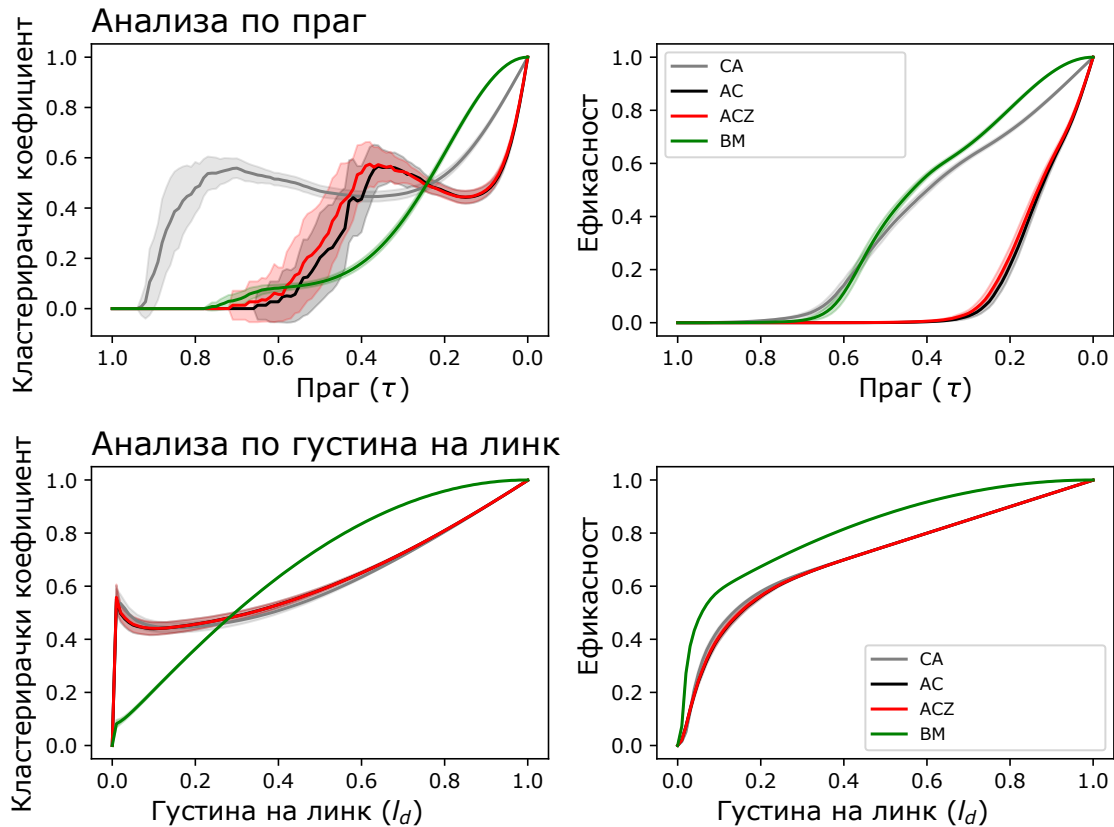
каде n е бројот на субјекти во секоја група. Вредностите над 0,8 обично се



Слика 2-2: Споредба на коефициентите на корелација. (Лево) Веројатностна распределба на коефициентите на корелација на врската, за четирите разгледани методи на реконструкција. (Десно) Однос помеѓу вредностите дадени со методите AC и CA. CA: корелација на средни вредности; AC: средна вредност на корелации; ACZ: Fisher-Z коригирана средната вредност на корелациите; BM: Баесов метод.

толкуваат како голема разлика помеѓу двете групи [104]. Како што може да се забележи на сл. 2-5, секој метод е поврзан со различна крива, т.е. максимумот се постигнува на различни прагови и густини на врската; сепак, максималната вредност е слична во сите случаи, односно методот на реконструкција не влијае на можноста да се дискриминираат пациентите од контролните субјекти.

На крај, ја проучивме максималната вредност на Коен d за сите осум разгледани тополошки метрики. Да се забележи дека негативните вредности d се исто така значајни, бидејќи тие само укажуваат дека метриката е поголема за контролните субјекти отколку за пациентите. За да се земе предвид ова, две ленти се прикажани на сл. 2-6 за секоја тополошка метрика и методот на ренормализација: пополнета, означувајќи ја најголемата вредност во апсолутни износи; и празна, означувајќи ја најголемата вредност со спротивен знак во однос на претходната лента. Како што претходно беше прикажано на сл. 2-5, големината на овој максимум Коен d е слична, независно од методот на реконструкција - со неколку исклучоци, на пр. големината на гигантската компонента во анализата според густината на врската. Во исто време, интересно е да се напомене дека изборот на методот на реконструкција може да влијае на знакот за најдобар d . За илустрација, кога се разгледува кластерирачкиот коефициент и бинаризацијата заснована на праг, трите фреквенционистички пристапи даваат повисоки вредности за пациентите (т.е. $d > 0$), додека баесовиот една повисока вредност за контролните субјекти ($d < 0$). Исто така, кога тополошката метрика дава, на пример, голем позитивен d , другите прагови овозможуваат да се добие слично

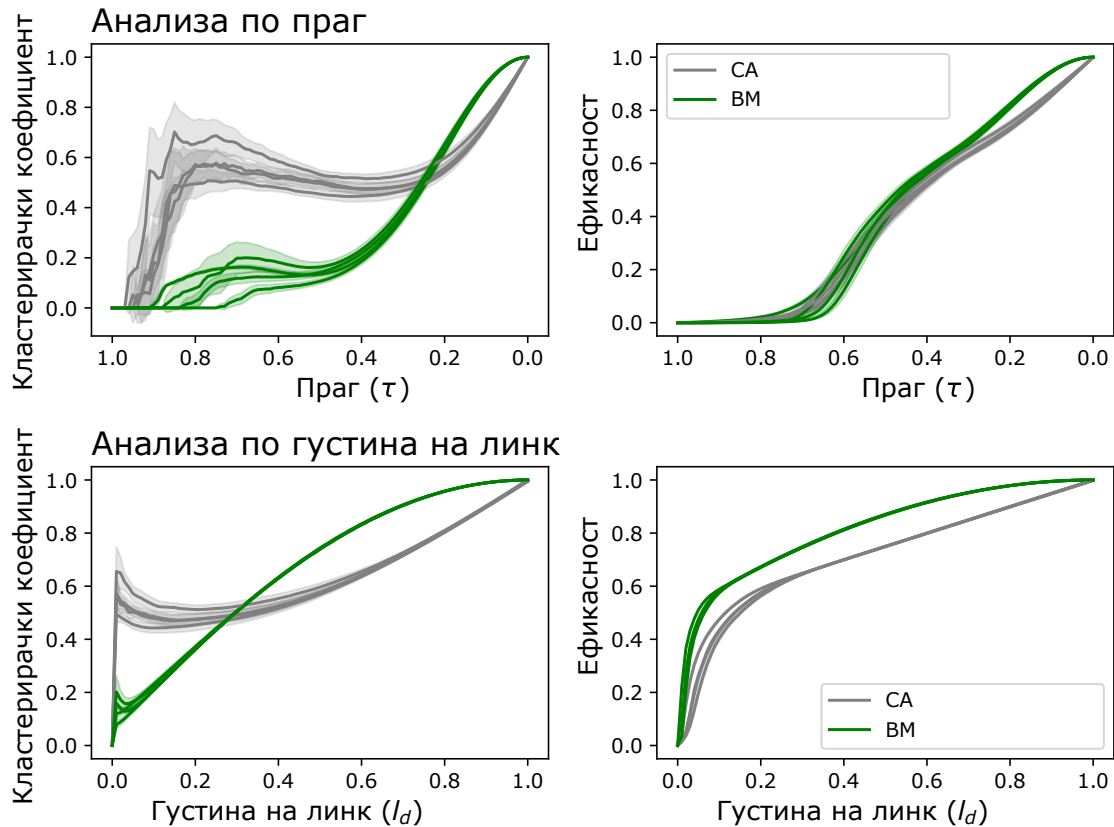


Слика 2-3: Еволуција на две тополошки метрики (кластерирачки коефициент, леви панели и ефикасност, десни панели) како функција на прагот τ (горните панели) и густината на врската l_d (долни панели), за податочното множество COBRE. Значењета на линиите и боите е како на сл. 2-1.

голем негативен d . Како и во претходниот случај, на сл. 2-7 се прикажува истата анализа за множеството COBRE; додека резултатите се различни, може да се забележат истите заклучоци.

2.3 Дискусија

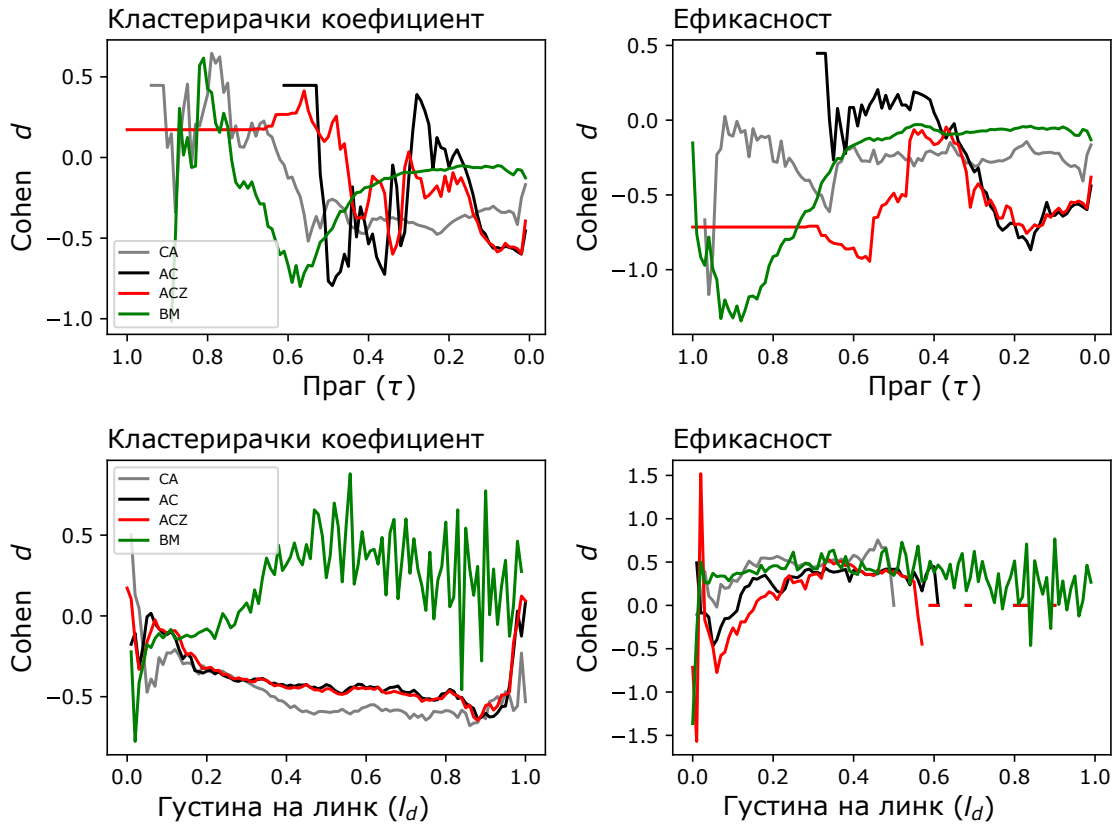
Идеално, на забележана активност на мозокот да се додели мрежна структура треба да се донесе непристрасна претстава на основниот динамички мозочен систем. За својствата на таквата структура да бидат внатрешни својства на системот, тие треба да бидат робусни во однос на конкретниот начин на реконструкција на мрежата од експерименталните податоци. За да се реши овој проблем во оваа дисертација, реконструиравме мрежи поврзани со fMRI активност во време на мирување на луѓе кои страдаат или од нарушување на аутистичен спектар или



Слика 2-4: Еволуција на кластерирачкиот коефициентот (леви панели) и ефикасноста (десни панели) како функција на прагот τ (горни панели) и густината на врската (долни панели). Сивите и зелените линии соодветствуваат на корелацијата на методите на средна вредност (CA) и Баесовата реконструкција (BM). Секоја од четирите линии со иста боја одговара на четирите разгледувани парцелации т.е. атласи, т.е. 1mm AAL, 2mm AAL, Brainnetome и HO.

пизофренија, и соодветни контролни групи, користејќи четири различни парцелации и два пристапи на праг. Особено, оценивме дали промената на правилата за интеракција преку кои се одредува регион на интерес влијае на својствата на тополошката мрежа на добиената мрежа. Нашите резултати генерално покажуваат дека различните техники за ренормализација на врската можат да доведат до скаларни и метрички квантитативни зависности и знаци на разлики меѓу популациите, но може да не влијаат на можноста за дискриминација помеѓу популациите.

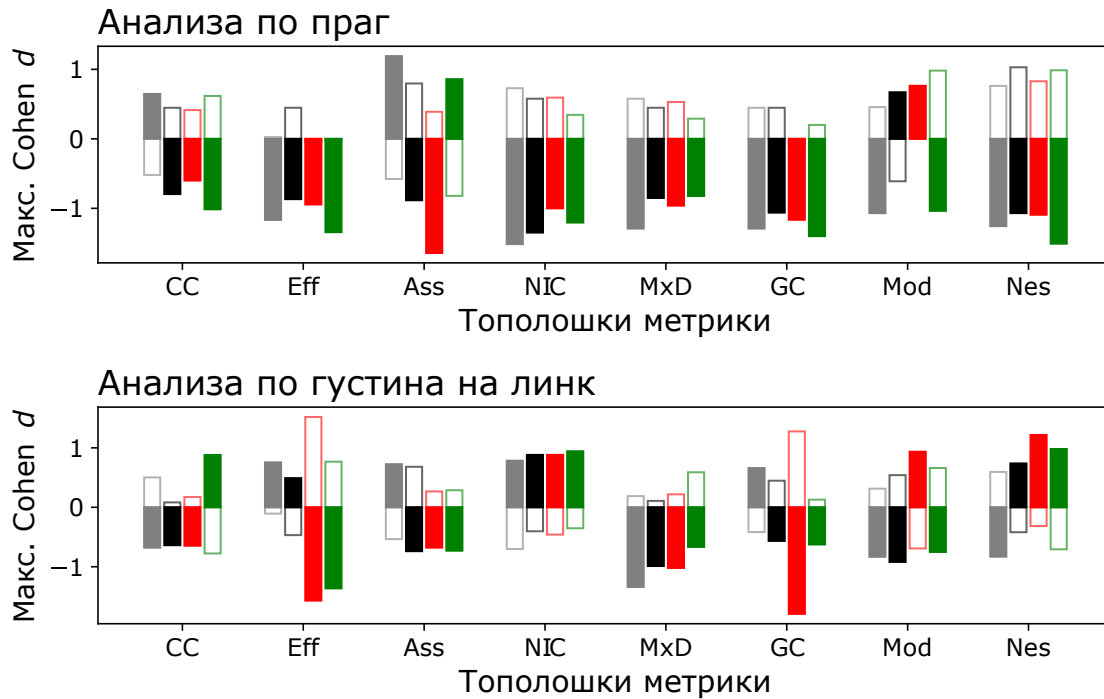
Првиот важен заклучок е дека методите за реконструкција може да се разликуваат во нивната способност да ги обноват специфичните својства на мрежата. Ова сугерира дека на својствата на мрежата може да влијае методот што се користи за реконструкција на мрежата. Исто така, се покренува можноста, ко-



Слика 2-5: Еволуција на Коен d (видете го главниот текст за дефиниција) за кластерирачкиот коефициент (леви панели) и ефикасноста (десни панели), како функција на прагот τ (горните панели) и на густината на врската (долу панели). Четирите редови во секој панел одговараат на четирите овде разгледани методи на реконструкција; кодови на боја како на сл. 2-1.

ја треба посебно да се истражи, некои својства на мрежата да бидат подобро истакнати преку специфични методи на реконструкција.

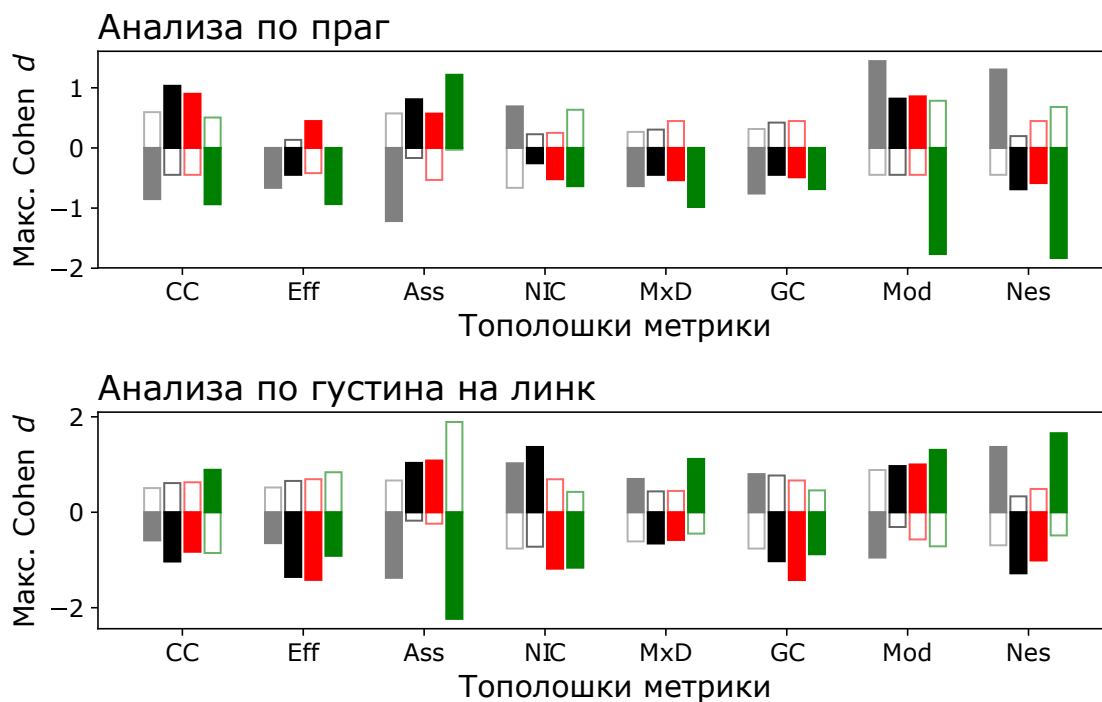
Понатаму, со оглед на multiscale природата на динамиката на мозокот, евалуацијата на постапката за реконструкција подразбира разбирање на нејзиното однесување во опсег на просторни размери. Нашите резултати покажуваат дека процедурите за ренормализација на врските потенцијално можат да го променат протокот на ренормализација на целиот систем, а со тоа и на крај основната физика на системот што може да се извлече од репрезентацијата на поврзаната мрежа. Во нашите податоци, разликите помеѓу методите за реконструкција се поизразени при ниските густини на врските, типични за невронаучните студии. Сепак, разликите се појавуваат и при голема густина или, еквивалентно, при ниски вредности на праг, за различните мрежни својства. Покажано е дека



Слика 2-6: Еволуција на максималната Коен d за осумте тополошки метрики, за апсолутниот (горниот панел) и пропорционалниот (долниот панел) пристапен праг. Четирите колони во секоја група одговараат на четирите методи на реконструкција; кодови на боја како на сл. 2-1.

слабите врски играат важна улога во топологијата на мрежата; нивното вклучување може да предизвика транзиции од фрактални во класи на универзалност на мал свет [105, 69] и да влијае на динамиката на мрежата и процесите што се случуваат во неа [106, 107]. Нашите резултати се во согласност со претходните индикации дека оптималниот дел од врските потребни за карактеризирање на системот може да биде специфичен за состојбата и ова може да одговара на процент на врски значително повисок од оној што обично се задржува во невронаучните студии [59].

Друг важен резултат во оваа дисертација е претставен со разлики поврзани со прагот помеѓу методите за реконструкција. За фиксните прагови (т.е. апсолутен праг), разликите помеѓу резултатите дадени од корелацијата на средни вредности и баесовите методи се мали, во споредба со оние од двата други методи. Ова се должи на суштински различните дистрибуции на тежината на врската во случаите на AC и ACZ (видете сл. 2-2). Од друга страна, за пропорционален праг, еволуцијата на метриците според Баесовиот пристап е квалитативно различна од другите три. За да се разбере оваа специфична разлика на прагот во техни-



Слика 2-7: Еволуција на максималната Коен d за осумте тополошки метрики, за апсолутниот (горниот панел) и пропорционалниот (долниот панел) пристапен праг, за множеството COBRE. Значењето на колоните и боите се како на сл. 2-6.

ките за ренормализација, корисно е да се разгледуваат праговите како посебни намалувања на релевантниот простор и да се разберат аспектите за кои апсолутното и пропорционалното ограничување соодветно ги истакнуваат или не се свесни. На пример, ако даден услов, на пр. патологија, вклучува само промена во динамичката јачина, но нема промена во целокупната топологија на мрежата, апсолутниот праг може погрешно да покаже дека оваа состојба ја модифицира топологијата, кога всушност не е така. Спротивно на тоа, пропорционалниот праг ја истакнува структурата на мрежите со најсилни врски, независно од големина на coupling јачината; како и да е, може погрешно да ги смета функционално нееквивалентните мрежи како неразлични ако функционалната активност на мозокот е нелинеарна функција на coupling јачината.

Додека методите за ренормализација на врските можат да ги променат тополошките својства на реконструираната мрежа, за да се процени функционалното значење на таквите модификации, корисно е да се разбере степенот до кој таквите модификации влијаат на можноста да се прави разлика помеѓу експерименталните услови или популациите. Нашите резултати сугерираат дека ваквата дискриминација е силна во однос на методот на ренормализација и дека

повеќето тополошки метрики можат да се користат како сигурни класификатори. Меѓутоа, особено, промената на методот на ренормализација или прагот што се користи во бинаризацијата на мрежата може да доведе до пресврт на знакот во Коен d , што укажува на тоа дека специфичните состојбени модулации на набљудуваните мрежни својства треба да се толкуваат во смисла на апсолутна промена, наместо во смисла на знак на таквата промена.

Конечно, нашите резултати покажуваат дека изборот на атлас или парцелација не влијае значително на својствата на реконструираната мрежа, без оглед на нејзината анатомска (т.е. AAL и HO) или функционално поврзанана (Brainnetome) природа. Иако овој резултат не гарантира стабилност на алтернативните методи кои не се базираат на атлас, тој укажува на разлики помеѓу атласите кои се доволно суптилни за да не донесат значителни последици врз добиената структура на мрежата.

Глава 3

Брза трансформација за евалуација на разлики во групи на мозочни мрежи

Специфичните анатомски и динамички темплејти на поврзаност се основен дел за здраво функционирање на мозокот [48, 108]. Од друга страна, дисконективноста, т.е. намалената и зголемената поврзаност, сугерира дека постои основа за некои невролошки и психијатриски состојби [109, 110, 111, 112]. Исто така, покажано е дека различните фази на развој [55] и стареењето [113] на човекот, како и разни невролошки и психијатриски патологии [57] влијаат на тополошките својства на мозочните мрежи добиени од анатомска и динамичка поврзаност [18].

Една многу нетривијана задача е квантифицирањето на значајни разлики во мозочната поврзаност во дадени експериментални услови или популации и одредување на кое мрежно својство е важно при нивно идентификување, која бара или софистицирано статистичко тестирање или компјутерски интензивни техники за машинско учење [14]. Една причина за овој проблем се однесува на фактот дека набљудуваните динамички модели на мозочна активност се појавуваат на нетривијален и нелокализиран начин во мозочната поврзаност во сите размери [61]. Исто така, додека топографијата на мозокот игра важна улога во функцијата на мозокот, својствата на тополошката мрежа се во суштина од статистичка природа. Литературата за мрежна невронаука обично ја нагласува мозочната поврзаност и топологија предизвикана од силни врски. Сепак, покажано е дека слабите врски имаат силно влијание врз топологијата на мрежата, каде што нивното вклучување може да предизвика транзиции од фрактални во класи на

мал свет [114], но исто така и на динамиката и процесите што се случуваат на мрежата [106, 107]. Заедно, овие размислувања сугерираат дека експерименталните услови може да се препознаат не само преку структурата предизвикана од силни врски, со можно додавање на слаби, туку преку нивните односи низ целата мрежа на начин што е барем делумно независен од топографската локализација.

Во оваа дисертација, предлагаме пресметковно брз метод за мерење на разликите во мозочната поврзаност во експериментални услови засновани на рангирање на тежината на врската. Врските на еднакви целосно поврзани тежински динамички мрежи на мозочна активност асоцирани со две различни експериментални состојби се рангирани според нивната тежина. Во сите состојби се споредуваат медијаната и стандардната девијација на овие врски. Разликите во рангирањето на врските на сите нивоа на распределба по категорија-тежина предизвикуваат специфични геометрии во соодветен помошен простор, чии оски се соодветно нивната меѓу-групна разлика во медијаната и стандардната девијација.

На крај, податоците се претставени во распрсувачки дијаграм, каде што секоја точка одговара на рангирање на врската, и нејзината позиција до разликата во медијаната/стандардната девијација на врската помеѓу групите. Оваа графичка претстава овозможува лесно да се опише како јачините на поврзаност се менуваат од состојба, како во однос на медијаната, така и во однос на варијабилноста; дали ваквите промени се униформни или се центрирани на силни/слаби врски; и колку се сигурни овие резултати.

Остатокот од текстот од оваа глава е организиран на следниов начин. Сек. 3.1 претставува преглед на методите и податочните множества што се користат во овој дел од дисертацијата. Резултатите се презентирани во сек. 3.2, и на крај, се дава дискусија на резултатите и се извлекуваат некои заклучоци.

Оваа глава од дисертацијата е објавена во трудот Zanin, Massimiliano, Plinka Ivanoska, Bahar Güntekin, Görsev Yener, Tatjana LoncarTurukalo, Niksa Jakovljevic, Olivera Sveljo, and David Papo. “A fast transform for brain connectivity difference evaluation.” *Neuroinformatics* (2021): 1-15. IF 4.085.

3.1 Методи и податоци

3.1.1 Креирање на репрезентација

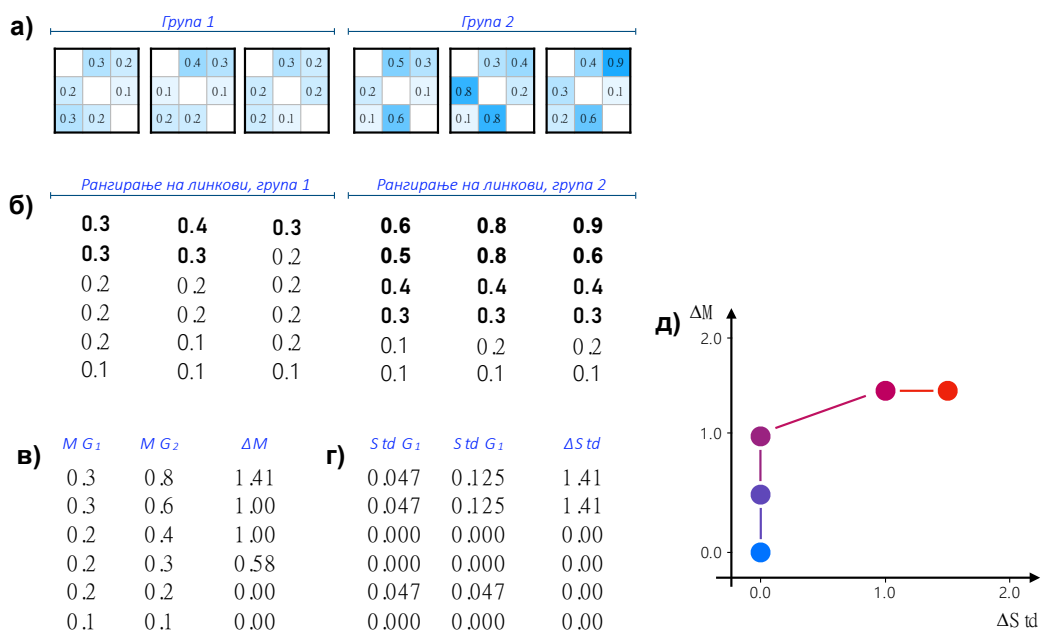
Методот предложен во оваа дисертација, зема како влез збир на тежински матрици на соседство, каде секоја од нив претставува поврзаност помеѓу различните

мозочни региони на даден субјект. Забележете дека не се воведуваат ограничувања на начинот на пресметување на таквата поврзаност, под услов резултатот да добие форма на реален број. Се споредуваат две различни групи на субјекти (на пр. пациенти и контролни), при што секоја матрица припаѓа на една од нив - видете сл. 3-1 (а). Тежините на врските (без јамките) потоа се извлекуваат од секоја матрица и се рангираат во опаѓачки редослед - сл. 3-1 (б). За секоја група и ранг позиција, дополнително се извлекуваат две метрики: медијаната (сл. 3-1 (в)) и стандардната девијација (сл. 3-1 (г)) на тежините на врските во иста група.

Разликата во двете метрики помеѓу две групи, соодветно се пресметува како $\Delta M = \log_2 M_{g1}/M_{g2}$ (сл. 3-1 (в), трета колона) и $\Delta Std = \log_2 Std_{g1}/Std_{g2}$ (сл. 3-1 (г), трета колона), со M што претставува медијана, и $g1$ и $g2$ двете групи што треба да се споредат. И во двата случаи, вредностите од Δ поголеми (соодветно, помали) од нула покажуваат дека мрежите во втората група имаат поголеми (помали) вредности од оние во првата. Да се забележи дека вредностите ΔM и ΔStd сè уште зависат од позицијата на рангирање. Конечно, сите резултати се претставени во дијаграм, каде што точките, што одговараат на рангирачките позиции, се наоѓаат во $\Delta Std - \Delta M$ рамнина (сл. 3-1. Заради јасност, точките се обоени според нивната позиција во рангирањето, од сини (најслаби врски) до црвени (најсилни врски).

Конечната слика, како онаа на сл. 3-1 (д), е лесна за толкување. Точките формираат континуум што се движи од најсилни (црвени) до најслаби (сини) врски, а нивната позиција покажува како тежините се разликуваат помеѓу условите. Позитивните вредности долж Y -оската означуваат дека втората група има посилни врски во просек; и позитивни вредности долж X -оската, дека втората група има поголема варијабилност. Како дополнителен пример, разгледајте го првиот панел на сл. 3-2 (т.е. Модел 1). Тука, посилните врски (т.е. црвените точки) во првата група ја имаат истата јачина на врската на медијаната (како $\Delta \approx 0$) и поголема варијабилност ($\Delta Std \approx 1$) од линковите во втората група. Од друга страна, слабите врски (т.е. сините точки) имаат и помала средна јачина ($\Delta M \approx -1$) и помала варијабилност ($\Delta Std \approx -2$).

Заради комплетност, дополнителна карактеристика се додава на дијаграмот. Полутранспарентна ознака се црта заедно со секоја точка, опфаќајќи се од 16-ти до 84-ти процент на секоја метрика пресметана со земање на половина од достапните матрици по случаен избор. Така, секоја ознака претставува несигурност во положбата на соодветната точка ако беа достапни само половина од податоците.



Слика 3-1: Графички приказ на предложениот метод. (а) Почетни матрици за поврзување, за две групи составени од по три мрежи. (б) Рангирање на врските на секоја мрежа. (в) Медијана на врските, за секоја група и секоја позиција на рангирање; и разлика на медијаните на врската, пресметана како логаритам со основа два на односот помеѓу вредностите на двете групи (г) Стандардна девијација на вредностите на врската, за секоја група и секоја позиција на рангирање; и разликата во стандардните девијации на врската. (д) Конечна графичка претстава во рамнината $\Delta M - \Delta Std$, секоја точка претставува позиција за рангирање и е обоена според рангираната позиција (од црвена, највисока, до сина, најниска).

3.1.2 Податочни множества

Девет податочни множества се разгледани во оваа дисертација за потребите на овој дел кои се опишани во следниот преглед на нивните главни карактеристики дадени во таб. 3.1. Освен ако не е поинаку наведено, не се извршени понатамошни чекори за обработка (вклучително и намалување на шум или елиминација на артефакт).

Шизофренија (EEG)

Ова податочно множество вклучува EEG снимки во состојба на мирување од збир на пациенти со шизофренија и соодветни контролни субјекти, како што е опишано во [115] и достапно на <http://dx.doi.org/10.18150/repod.0107441>. 14 пациенти (7 мажи, 27, 9±3, 3 години и 7 жени, 28, 3±4, 1 години) се со критериуми за меѓународна класификација на болести за параноидна шизофренија (катего-

Type	Под. множ.	# Контроли	# Пациенти	# Сензори / ROI	време
EEG	Шизофренија	14	14	19	≈ 15m
	Алцхајмер	38	42	32	≈ 8m
	Паркинсон	22	74	32	≈ 8m
	Алкохолици	8,048	14,066	64	1s
fMRI	Аутизам	592	521	116	5-16 m
	Шизофренија	33	21	116	6m
	Алцхајмер	34	31	116	6m
DTI	Аутизам	43	51	264	<i>n.a.</i>
	Пол	114	80	188	<i>n.a.</i>

Табела 3.1: Главни карактеристики на податочните множества.

рија F20.0). Соодветните 14 здрави контроли се 7 мажи, возраст од $26,8 \pm 2,9$ години и 7 жени, возраст од $28,7 \pm 3,4$ години. Петнаесет минути EEG податоци се снимени за време на состојба на одмор со затворени очи. Податоците се добиени на 250 Hz со користење на стандардна 10-20 EEG монтажа со 19 EEG канали: Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2. Референтната електрода е поставена во FCz.

Алцхајмерова болест (EEG)

Податочното EEG множество за пациенти со Алцхајмерова болест (АД) е снимено во Универзитетската болница Медипол во Истанбул и мултидисциплинарен центар за истражување на мозочна динамика на мозокот во Универзитетот Измир Докуз Ејлул во Измир. Пациентите со АД се дијагностицирани според критериумите на “Националниот институт за невролошки и комуникативни заболувања и асоцијација на мозочен удар-Алцхајмерова болест и сродни нарушувања (NINCDS-ADRDA)” [116, 117] и на мануалот за ментални нарушувања-4-то издание (DSM-IV-TR) [118]. Исто така, скалата за клиничка деменција [119, 120, 121, 122] е искористена за проценка на сериозноста на АД. Вкупно се анализирани 42 пациенти (возраст: 56 – 86, просек од 74; пол: 6 мажи; edu: 0 – 13, просек од 5) и 38 здрави субјекти (возраст: 54 – 70, просек од 62,5; пол: 5 мажи; edu: 0 – 12, просек од 5). Тест за мини-ментална состојба [123, 124] е искористен за проценка на општата когнитивна состојба на сите учесници. Употреба на лекови поврзани со АД не е дадена во групата пациенти и пациентите со АД земале холинергична монотерапија или комбиниран холинергичен третман со мемантин.

EEGто на сите здрави контроли и пациенти со АД е снимено во слабо изолирана просторија со 32-канален системска машина Brain Amp (Brain Product GmbH,

Германија) со 32 различни електроди кои се распоредени според меѓународниот 10/20 систем. Стапката на земање примероци е 500 Hz со ограничувања на опсег од 0,01 - 250 Hz. Сите напони се чувани под 10 Ком и две дополнителни поврзани електроди на ушниот лобус (A1 + A2) служат како референтни електроди. Електроколограмот е снимен со две електроди сместени во средниот горен и страничен орбитален раб на левото око.

Паркинсонова болест (EEG)

EEG податочното множество за пациентите со Паркинсон (ПД) е снимено во Универзитетската болница Медипол во Истанбул. Пациентите со Паркинсон се дијагностицирани според критериумите на “United Kingdom Parkinson’s Disease Society Brain Bank” [125]. Користена е Унифицирана скала за оценување на Паркинсонова болест (UPDRS) [126] со цел да се утврдат клиничките карактеристики на ПД; и скалата Хоен-Јахр [127] е искористена за да се утврди фазата на болеста. Вкупно се анализирани 74 пациенти (возраст од 56 – 86, просек од 74) и 22 соодветни контролни субјекти (возраст од 54 – 89, просек од 67). Сите пациенти со ПД се евалуирани 60-90 минути по утринската доза на леводопа за EEG-снимките. Условите за снимање, опремата и локацијата на електродите се како во податочното множество за Алцхајмеровата болест (EEG).

Алкохоличари (EEG)

Податочното множество за EEG снимки за алкохоличари содржи група алкохолни субјекти и соодветни контроли [128, 129], слободно достапни на <https://archive.ics.uci.edu/ml/datasets/EEG+Database>. Секое испитување одговара на задача за препознавање на предмет, како што е опишано во [130]; и соодветната EEG активност е забележана во текот на една секунда, со стапка на земање примероци од 256 Hz (3,9-ms / епоха) од електроди од 64 лоцирани на стандардни места на скалпот. Достапни се испитувања од 4024 контроли и 7033 пациенти, со вкупно 11057 инстанци.

Аутизам (fMRI)

Првото податочно fMRI множество избрано за овој дел од дисертацијата е ABIDE II [131], достапно на http://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html, кое се состои од збирка од 19 податочни множества на лица со нарушување на аутистичниот спектар (АСД) и типични контроли. Тоа вклучува податоци за функционална магнетна резонанца (rs-fMRI) во состојба на

мирување, DTI слики, фенотипски податоци, како и анатомски податоци. За да се креираат функционални репрезентации на конектоми, ги земавме анатомските и функционалните податоци од сите 19 податочни множества (освен лонгитудалните), за 521 пациент со аутизам и 592 контроли, за вкупно 1113 субјекти. Дополнителни информации во врска со овие податоци и нивните соодветни слики се дадени на http://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html.

Пред-обработката ја извршивме со употреба на статистичко параметричко мапирање (SPM12) [132] во MATLAB 2018b вграден во алатката CONN [133]. Чекорите за пред-обработка вклучуваат стандардни чекори со функционално прередување (проценка и корекција на движење), корекција на временски серии, корегистрирање на соодветните анатомски T1 слики на субјектите со нормализација на стандардниот образец за невролошки институт во Монреал (МНИ), откривање на outlier-и и мазнење со целосна ширина од 8 мм на половина од максимумот (FWHM) кернел. Покрај овие чекори, се користи и сегментација на сива материја, бела материја и области на цереброспинална течност (CSF) за отстранување на временските збунувачки фактори (бела материја и CSF). Покрај тоа, филтрирањето со пропусен опсег беше извршено со прозорец со фреквенција од 0,008-0,09 Hz. За елиминација на движењето на главата и артефактите, временските точки на периоди се идентификувани во параметрите на движење и глобалниот интензитет на сигналот со употреба на ART [134] додаден на стандардниот прозорец во алатката CONN. Матриците за функционална поврзаност за секој субјект се пресметани со помош на CONN со употреба на атласот Anatomical Atlas Labelling (AAL) [135] за мозочна парцелација.

Шизофренија (fMRI)

Второто fMRI податочно множество вклучено во оваа дисертацијата во овој дел е податочното множество COBRE достапно на (http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html), кое се состои од анатомска и функционална магнетна слика во состојбата во мирување (rs-fMRI) за 72 пациенти со шизофренија и 75 типични контроли. Податоците за rs-fMRI се добиени со користење на ехо-планарна слика (EPI) со $TR = 2$ s, $TE = 29$ ms, големина на матрица = 64×64 , број на делови = 32 делови и големина на воксел = $3 \times 3 \times 4$ mm³. Дополнителни информации за податочното множество и фенотипските податоци се достапни на http://fcon_1000.projects.nitrc.org/indi/retro/cobre.htm. Извршената претходна обработка се совпаѓа со онаа на податочното множество за Аутизам (fMRI).

Алцхајмерова болест (fMRI)

Иницијативата за невро-слика на Алцхајмеровата болест (ADNI) (www.adni-info.org) е лансирана во 2003 година со примарна цел да се тестира дали сериско снимање со магнетна резонанца (МРИ), томографија со емисии на позитрони (ПЕТ), биолошки маркери, и клиничката и невропсихолошката проценка може да се комбинираат за да се измери прогресијата на благото когнитивно оштетување (MCI) и раната Алцхајмерова болест (АД).

Во оваа дисертација користевме податочно fMRI множество од базата на податоци ADNI (<http://adni.loni.ucla.edu>). 60-те записи користени во овој дел се од контролните (пред-третман) сесии. Учесниците се категоризирани според клиничките податоци во повеќе групи, од кои овде се користени две групи: група на когнитивно нормални субјекти (37 учесници) и група на Алцхајмерова деменција (АД) (23 учесници). Групата контроли се состои од 17 мажи и 20 жени на просечна возраст 73.65 ± 5.74 , додека АД групата се состои од 10 мажи и 13 жени со просечна возраст 73.17 ± 7.62 . Секој учесник е скениран на MRI скенер „Филипс“ 3.0T. fMRI аксијалните слики се добиени со користење на ехо-планарна (EPI) секвенца со време на повторување (TR) од 3000ms; време на ехо TE = 30ms; агол на превртување (FA) = 800 и 48 делови со дебелина на дел од 3,313 mm. Учесниците имале упатства да се релаксираат и да ги држат очите затворени за време на сесијата за скенирање. Вкупно 140 волуменски податоци се достапни за секој учесник.

Податочното множество е претходно обработено со помош на алатката за обработка на fMRI податоци во состојба на мирување (DPARSF) (<http://www.rfmri.org/DPARSF>) [24]. Првите седум временски точки беа отфрлени за да се обезбеди стабилизирање на магнетното поле. Чекорите за пред-обработка вклучуваат: корекција на времето за делови, преуредување за да се елиминираат артефактите на движењето и нормализирање на просторот до стандардниот образец за EPI. Сликите беа просторно измазнети со Гаусов кернел со целосна ширина на полу-максимум (FWHM) од 6 mm. И линеарните и квадратните трендови беа отстранети, а параметрите за движење на главата, сигналите на церебралната течност и белата маса беа тргнати од податоците. Учесниците со значителни артефакти на движење беа исклучени од понатамошна анализа. Конечно, BOLD сигналите беа извлечени од 116 мозочни региони според AAL атласот [135] за остатокот од учесниците.

Аутизам (DTI)

Првото DTI податочно множество со структурни конектоми за оваа дисертација е базата на податоци UMC UCLA Аутизам колекција [136] земена од базата на податоци за мултимодална мозочна поврзаност UCLA [137], отворено достапна на <http://umcd.humanconnectomeproject.org/>. Колекцијата се состои од однапред конструирани структурни и функционални конектоми за вкупно 175 субјекти, од кои 94 се структурни конектоми за 51 пациент со нарушување на аутизмот (АСД) и 43 контроли, покрај 79 функционални конектоми за 42 пациенти (АСД) и 37 контроли. Во оваа дисертација, ги избравме структурните DTI конектоми 264x264, за откривање на разликите помеѓу АСД и контролите. Фенотипски и демографски податоци за податоците, како и аквизиција на сликите и структурни детали за конструкција на DTI матриците се дадени во [136].

Пол (DTI)

Последното податочно множество искористено овде е добиено од базата на податоци за мултимодална поврзаност UCLA [137] (отворено достапна на <http://umcd.humanconnectomeproject.org/>), и конкретно од колекцијата NKI Rockland. Се состои од 194 структурни и функционални матрици за поврзаност (конектоми) за 194 контролни субјекти земени од Институтот Нејтан Клајн (NKI)/примерок со податоци на Рокланд [138]. Информациите за податоците, претходната обработка на DTI и fMRI, како и изведувањето на структурната и функционалната матрица за поврзување се подетално опишани во [137]. Со цел да ги откриеме разликите во половите (мажи наспроти жени), ние ги користевме само структурните матрици 188x188 конектоми.

3.1.3 Реконструкција на функционални мозочни мрежи

Откако ќе се добие податочно множество на временски серии за секој субјект, во случај на податоци за EEG и fMRI, овие може да се поделени во не-преклопувачки прозорци со големина τ . Освен ако не е поинаку наведено, тука разгледаваме $\tau = 128$. Конечно, реконструирана е тежинската матрица за соседство W за секој прозорец, каде што секој елемент $w_{i,j}$ претставува јачина на функционалната поврзаност помеѓу јазлите (електроди или региони на интерес) i и j . Конективноста е пресметана со четири метрики:

- Линеарна Пирсонова корелација, што одговара на апсолутната вредност на класичната линеарна корелација помеѓу двете временски серии.

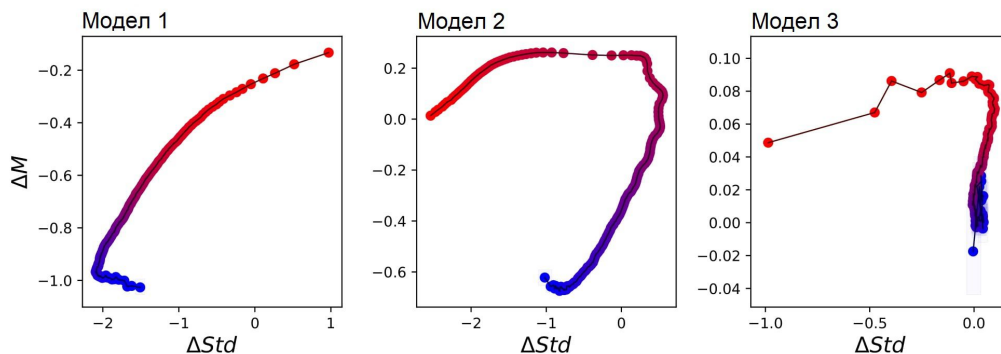
- Granger Causality (GC), линеарна метрика на каузалност заснована врз проценка на подобрувањето во предвидувањата на временската серија Y кога се вклучени информации за втора временска серија X . Ако грешката во предвидувањето се намали, X се вели дека е Y *Granger-предизвикан* [139]. Вредноста на секој елемент $w_{i,j}$ се дефинира како $-\log_{10}$ од p -вредноста на тестот Granger помеѓу временските серии i и j .
- Меѓусебна информација (МИ, Mutual information MI), мерка за взаемната зависност помеѓу двете променливи. Таа е дефинирана како количина на информации добиени за една временска серија преку набљудување на втората [140].
- Трансфер ентропија (TE), метрика за мерење на количината на насочен пренос на информации помеѓу две временски серии. Поконкретно, се дефинира како колку се намалува несигурноста во идните вредности на временската серија Y со познавање на минатото на втората временска серија X [141].

3.2 Резултати и дискусија

Претходно да доведе комплетен опис на методот предложен во овој дел, со графички приказ на процесот на сл. 3-1. Резултатите за синтетички мрежи создадени од генеративни модели се претставени на сл. 3-2. Го примениме предложениот метод на голема колекција на анатомски и функционални податоци за мозокот снимени со разни техники на невровизуелизација од луѓе кои страдаат од голем број невролошки и психијатриски патологии, како и од соодветни контролни субјекти - сл. 3-8, видете Методи и податоци за детали за податоци и обработка.

Треба да се споменат неколку работи. Прво, во повеќето случаи, трансформацијата дава нетривијална слика за целокупната структура на разликите помеѓу дадената патологија и нејзината соодветна контролна група во однос на разликите во рангирањето на јачината на конективноста (поврзаноста). Овој резултат укажува на тоа дека, дури и во отсуство на информации за топографска локализација, кои честопати се сметаат за неопходни во неврологијата на ниво на системот, трансформацијата задржува доволно информации за моделот на мрежна поврзаност за да ги дискриминира патолошките сигнали од здравите контролни.

Второ, нашите резултати сугерираат дека релевантните информации за дискриминација се кодирани во хиерархијата на јачината на мозочната поврзаност, но исто така, што е важно, во нејзината варијабилност. Додека варијабилноста



Слика 3-2: Анализа на резултатите дадени од збир на синтетички модели. Ние го разгледуваме случајот на две групи од по 10.000 мрежи, секоја составена од 20 јазли, каде временска серија со должина $\tau = 8$ е поврзана со секој јазол. Тежините на врските помеѓу паровите јазли се пресметуваат како апсолутна вредност на линеарната корелација помеѓу соодветните временски серии. Во случај на првата група, овие временски серии секогаш се создаваат со мешање на заедничка фиксна шема со случајна компонента. Во случај на јазолот i , неговата временска серија е дадена со $x_t^{(i)} = \pi_t + \alpha^{(i)}\mathcal{U}(-1, 1)$. π е фиксна временска серија од вредности $\tau = 8$, создадена со избирање случајни броеви од униформна распределба $(0, 1)$, врз која се надува случаен сигнал. Амплитудата на ваквиот додатен сигнал е контролирана од $\alpha^{(i)}$, случаен број извлечен од униформна распределба $\mathcal{U}(0, 1)$. Забележете дека секој пар на јазли со мала α ќе има слична динамика, поради доминацијата на π , а со тоа и голема корелација. Мрежите од втората група се конструирани според различни генеративни модели, со временски серии дефинирани со $y_t^{(i)} = \pi_t + \beta^{(i)}\mathcal{U}(-1, 1)$

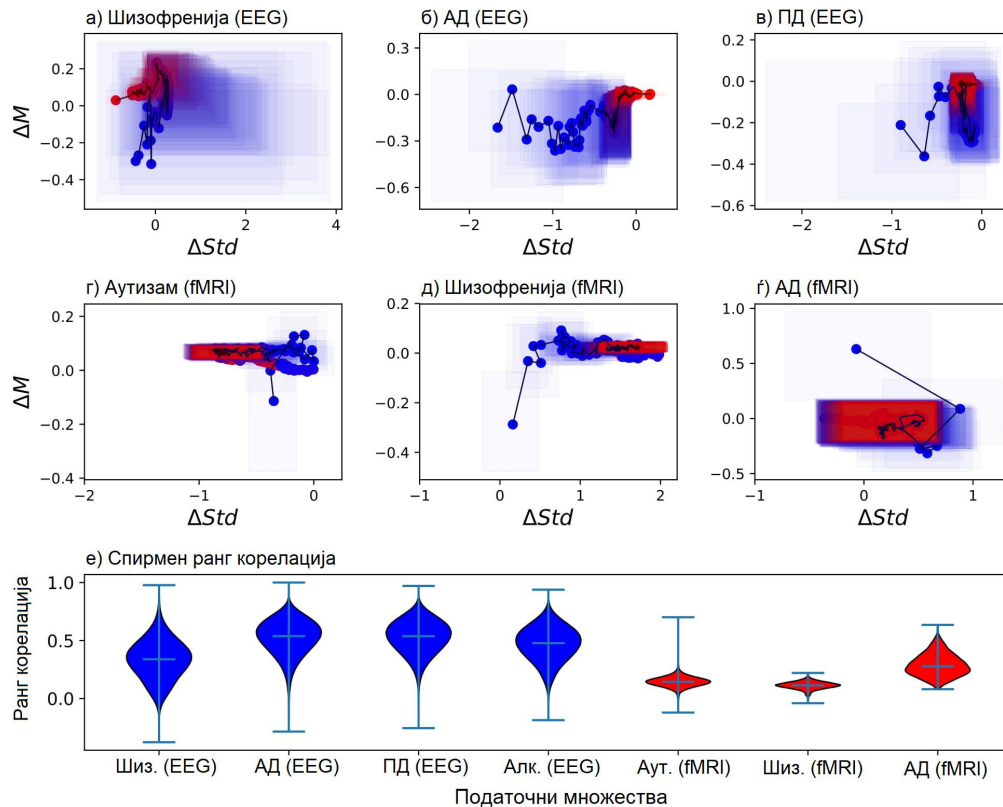
Панел (а), модел 1. $\beta^{(i)}$ е број извлечен од униформна распределба $\mathcal{U}(0, 8)$. Бидејќи β е поголема од α , компонентата на шум е посиљна во втората група, а со тоа и врските имаат помала тежина - видете ја негативната вредност од Δ . На глобално ниво, слабите врски имаат тенденција да имаат повеќе негативни Δ , бидејќи разликите се зголемуваат, во релативна смисла, со помалата тежина; и помала варијабилност, бидејќи поголемата вредност на β ја намалува веројатноста да има силна врска на нискорангираните позиции.

Панел (б), модел 2. $\beta^{(i)}$ е случаен број извлечен од униформна распределба $\mathcal{U}(0, 0, 2)$ за $i \leq 10$; и од експоненцијална распределба со скаларен параметар $1/\lambda = 4$ инаку. Првите десет јазли на тој начин се карактеризираат со високо синхронизирана динамика и хомогени тежини, со што се доведува до позитивно Δ и негативно ΔStd ; синхронизацијата потоа се губи за преостанатите врски, враќајќи ја кривата слична на онаа на моделот 1.

Панел (в), модел 3. Во овој случај, $\beta^{(i)}$ е број извлечен од униформна распределба $\mathcal{U}(0, 0, 3i)$, и е со тоа зависни од јазли: некои јазли (мали i) имаат високо корелирана динамика, додека други (i близу 20) се похетерогени. Додека јачината на глобалната поврзаност е претежно константна (т.е. $\Delta \approx 0$), некои парови на јазли се принудени да се синхронизираат, па оттука и нивната варијабилност е намалена (забележете ја негативната ΔStd за петте најсилни врски).

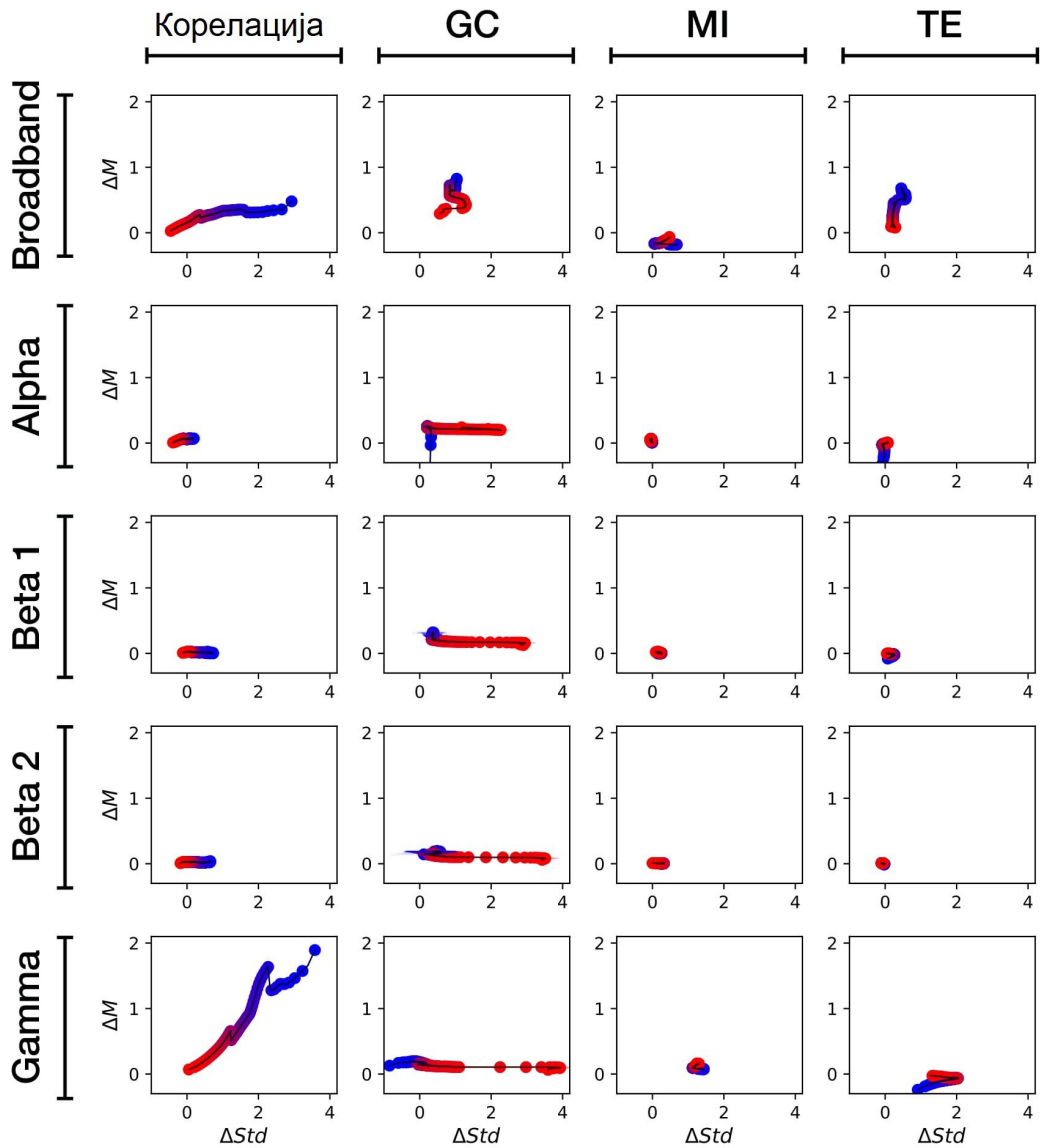
честопати посветува малку внимание при анализата на мрежата, нашите резултати се во согласност со истакната улога во здравите биолошки системи и во нивната патологија [142, 143].

Трето, геометријата на ранг-различната трансформација во медијана-



Слика 3-3: Анализа на резолуцијата на fMRI податоците. Панелите (а), (б) и (в) го прикажуваат графички рангирањето за соодветно податочните множества за шизофренија, АД и ПД EEG, кога соодветните временски серии се намалуваат со земање на една вредност на секои 128 - со што се симулира временска резолуција слична на онаа кај fMRI множествата. Резултатите се fuzzy структури, слични на оние добиени за fMRI множествата. Панелите (г), (д) и (ф) ги прикажуваат графички рангирањето за трите разгледувани fMRI множества, кога добиените вредности се просечни на 4 последователни врски во рангирањето. Добиените графици сугерираат дека fuzzy структурата не се должи на големиот број на јазли (т.е. на превисока просторна резолуција), туку на малата временска резолуција на временските fMRI серии. Панелот (е) конечно ја претставува распределбата на Спирмен ранг-корелацијата помеѓу тежината на врската на сите контролни субјекти во секоја група податоци. Општо, рангирањето на fMRI (црвените виолини) е помалку стабилно од оние со EEG (сини виолини), придонесувајќи за појава на fuzzy структури.

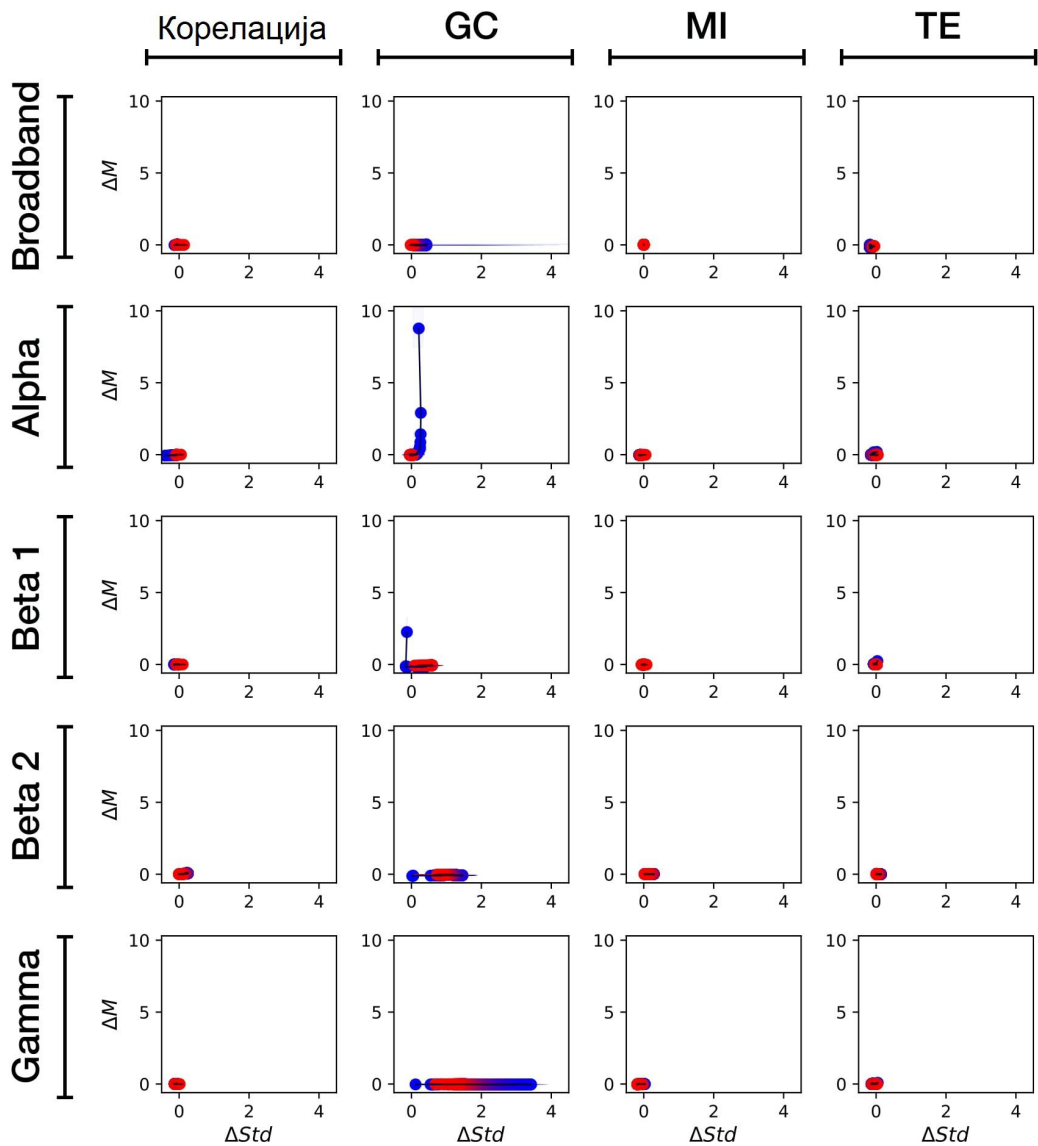
стр.девијација простор дава изразена специфичност на патологијата, која често може да се увиди на прв поглед. На пример, шизофренијата е поврзана со значително зголемување и на просечната тежина и на варијабилноста за слабите врски - забележете дека ΔStd 3 е еквивалентно на осумкратно зголемување на варијабилноста - додека најсилните врски не се разликуваат значително од оние



Слика 3-4: Анализа на резултатите добиени со предложениот метод за множеството за пизофренија ЕЕГ, како функција на конективната метрика за поврзување (колони) и на фреквентниот опсег (редови). Најдобрата диференцијација помеѓу контролните субјекти и пациентите, во однос на Δ , се добива со линеарна корелација за високи фреквенции; од друга страна, поголемите разлики во однос на ΔStd се добиени од Granger Causality.

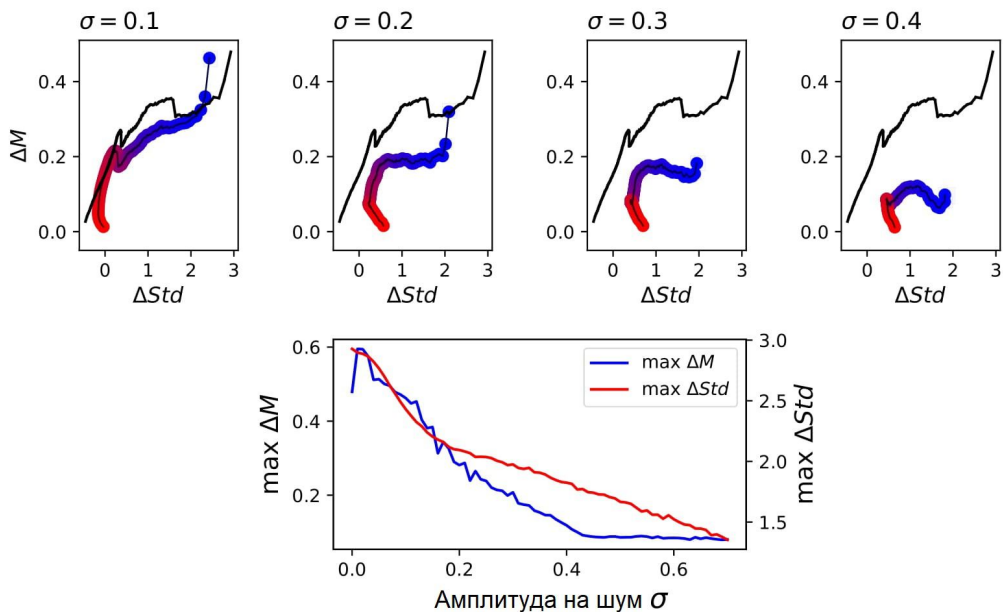
на контролните субјекти (слика 3-8 панел (а)). Спротивната шема е забележана кај Алцхајмеровата болест (АД) (панел (б)), во која зголемувањето на варијабилноста е ограничено на силни врски. Некако на средина од двата претходни примери е оној и на алкохолизирани пациенти, со поголема варијабилност во најсилните и најслабите врски, но не и во средните (панел (г)).

Панелите (ж) и (з) покажуваат како анатомските (за разлика од функционал-



Слика 3-5: Анализа на резултатите добиени со предложениот метод за податочното множество Алцхајмерова болест ЕЕГ, како функција на конективната метриката за поврзување (колони) и на фреквентниот опсег (редови). Најдобрата диференцијација помеѓу контролните субјекти и пациентите ја дава Granger Causality, при пониски фреквенции во однос на ΔM и на повисоки фреквенции во однос на ΔStd .

ните) мрежи за конективност, овде добиени преку DTI слики, исто така може да се анализираат преку предложената методологија. Слабите врски се оние што покажуваат голема разлика во двата случаи, додека најсилните врски се претежно стабилни на $\Delta \approx 0$. Важно е да се каже, со оглед на тоа што повеќето тековни студии за реконструкција на мрежата обично ги филтрираат слабите врски, обично задржувајќи низок процент од најсилните, резултатите и за функ-

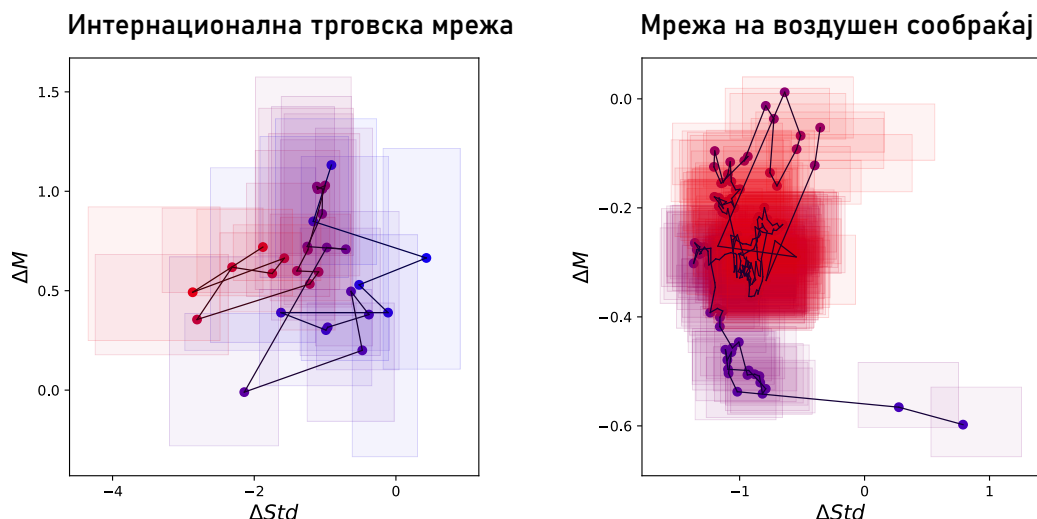


Слика 3-6: Анализа на робусноста на методот во однос на додаден шум. Четирите панели горе го прикажуваат резултатот од методот, кога се додава шум во временската серија во множеството за шизофренија ЕЕГ. Овој шум има нормална распределба, со нула средна вредност и стандардна девијација σ соодветно 0, 1, 0, 2, 0, 3 и 0, 4 од стандардната девијација на оригиналниот сигнал. Црвените / сините точки го претставуваат добиениот резултат, додека црната линија го прикажува резултатот за суровата временска серија. Дополнително, долниот панел ја прикажува еволуцијата на максималните ΔM и ΔStd како функција на амплитудата на шум σ .

ционалните и за анатомските податоци индиректно сугерираат дека способноста за откривање на овие студии може да биде специфична и за патологијата.

Конечно, се чини дека резултатите се постојано појасни за електроенцефалографијата (ЕЕГ) отколку за податоците за функционална магнетна резонанца (fMRI) (споредете ги резултатите од ЕЕГ со резултатите од fMRI, сл. 3-8 (д-е)). Со оглед на просторниот карактер на конективноста и неизмерно супериорната просторна резолуција на fMRI во однос на ЕЕГ, ова може да изгледа најпрвин изненадувачки. Овој резултат во принцип може да произлезе од тешкотија во управувањето со димензионалноста на матриците за конективност кај fMRI. Сепак, понатамошната анализа јасно укажува на значително помалата временска резолуција на fMRI во однос на ЕЕГ како вистинска причина - видете на сл. 3-3.

Ова може да укаже на тоа дека информациите кодирани во мозочната динамика можат да бидат поважни од нејзините просторни аспекти во дискриминацијата меѓу физиолошката и патолошката конективност и дека релативно слабата



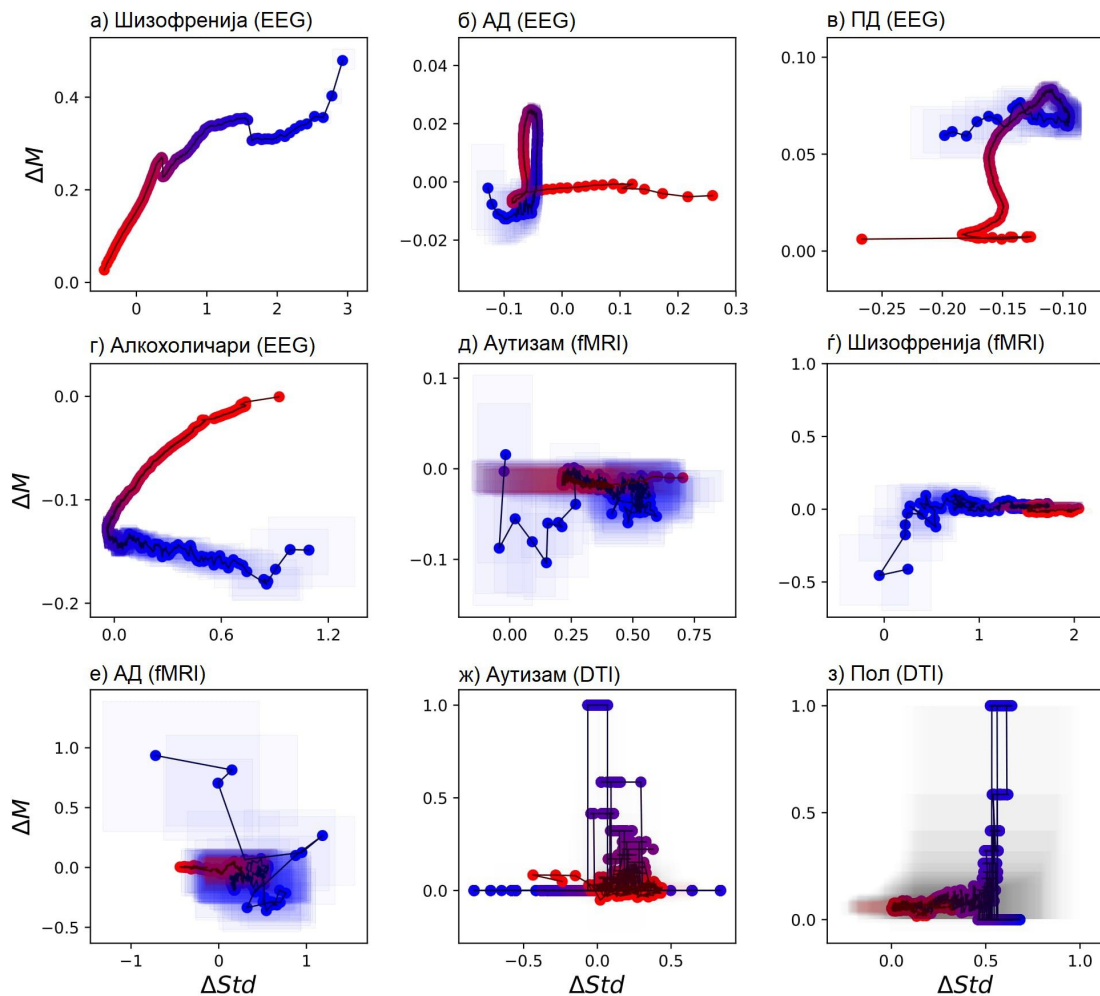
Слика 3-7: Примери за примена на предложената методологија на други сложени системи.

Левиот панел ги прикажува резултатите за меѓународната трговска мрежа помеѓу 6 светски региони, т.е. Африка, Азија, Австралија и Нов Зеланд, Европа, Северна Америка и Јужна и Централна Америка и Карибите. Податоците се добиени од веб-страницата на Светската трговска организација (СТО) (timeseries.wto.org) и одговараат на годишниот извоз на стока за тие региони од 2001-2018 година. 18те мрежи се поделени во две групи, соодветно на годините 2001 - 2011 и 2012-2018. Резултатите покажуваат дека меѓународната трговија е зголемена меѓу сите региони ($0 \leq \Delta \leq 1,2$); и дека се случила значителна стабилизација во најсилните врски ($\Delta Std \approx -3$).

Десниот панел одговара на мрежите за воздушен транспорт, поточно за бројот на патници кои патувале меѓу најголемите градови од 20 најголеми градови во САД. Податоците се добиени од Министерството за транспорт на САД (data.transportation.gov/Aviation) и опфаќаат години од 2009 - 2019. Двете групи тука одговараат на првата и последните пет години, за вкупно 48 мрежи во секоја од нив. Може да се увиди генерализирано намалување на бројот на патници, што особено влијае на послабите врски (веројатно како резултат на прелевање на ефектите во други начини на превоз).

временска резолуција на fMRI може да доведе до премногу поедноставена слика за поврзаност на мозокот и затоа не успеваат да се фатат функционално дискриминаторските аспекти.

Опсегот на споредби каде што предложениот метод може да помогне е далеку поголем од оној помеѓу различни популации. Забележително, методот може да се користи како брз *post hoc* метод за валидација при реконструкција на функционална мрежа. Реконструирањето на конективноста и својствата на мрежата од експерименталните податоци за невро-слика е многу нетривијална задача, со



Слика 3-8: Графичка претстава дадена од предложената методологија за деветте функционални и структурни податочни множества за мозочни мрежи разгледани овде. Погледнете во Методи и податоци за детали за податоци и обработка.

различни дискрециони чекори за кои не постојат прифатени теоретски упатства. На пример, сè уште нема принципиелен критериум за дефинирање на јазли или за избор на најсоодветна метрика за поврзување од многуте достапни, а како што наведовме во претходната глава, реконструкцијата на мрежата на повторен начин е компјутерски интензивен процес [59].

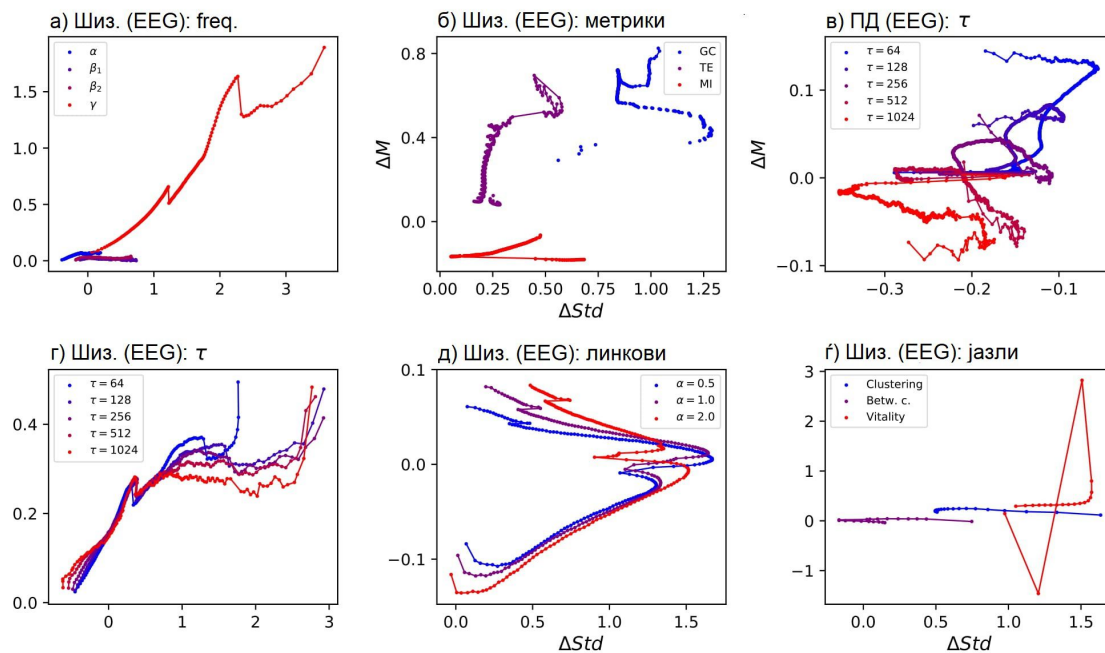
За споредба, трансформацијата може да помогне да се разберат кои аспекти на снимениот сигнал, на пр. кој дел од фреквентниот спектар на сигналот со широк опсег, содржи дискриминаторски информации. Сл. 3-9 илустрира неколку примери на такви апликации. Панел (а) ја известува графичката претстава што одговара на филтрирање на временските серии на различни фреквентни опсези за податочно множество за шизофренија (ЕЕГ); панел (б) употреба на

различни метрики за конективност. Како што беше претходно прикажано, високите фреквенции и линеарните корелации даваат најдобри резултати [144] - видете ги резултатите и за сите комбинации на фреквенции - метрики во сл. 3-4. Спротивното се забележува кај Алцхајмеровата болест (ЕЕГ), видете на сл. 3-5: алфа-опсегот кодира поголем дел од информациите, што е добро познат резултат во литературата [145].

Панелите (в) и (г) на сл. 3-9 исто така ги прикажуваат репрезентациите со различна должина τ од временскиот прозорец што се користи за проценка на корелацијата помеѓу јазлите, соодветно за податочните множества за Паркинсоновата болест (ЕЕГ) и за шизофренија (ЕЕГ).

Конечно, ние покажуваме како методологијата може да се примени и за мрежни тополошки метрики од повисок ред, секогаш кога тие можат да се пресметаат или за врски или за јазли. На начин сличен како за тежините на врските, едноставно може да се рангираат овие метрики за врски или јазли, да се пресметаат меѓу групата Δ и ΔStd и да се претстават резултатите во дијаграм. Сл. 3-9 (д) ги прикажува резултатите за *betweenness* централноста на работ (врската), за податочното множество за шизофренија (ЕЕГ), кога тежините на врските се зголемуваат на експонент α ; и панел сл. 3-9 (ѓ) резултати за три метрики засновани на јазли, т.е. кластерирачки коефициент, *betweenness* централност и виталност.

Како заклучок, во овој дел презентиравме лесно толкувачки и пресметковен графички метод за да ги потенцираме разликите во конективноста помеѓу системите под различни услови. Овој метод може да се искористи за да се согледаат на прв поглед основните аспекти на структурата и динамиката на биолошките, технолошките и економските мрежни системи (види сл. 3-7 за други примери) и итеративно да се рафинира нивниот графички приказ.



Слика 3-9: Примери за алтернативна употреба на предложената методологија. (а, б) Анализа на податочното множество за шизофренија (EEG) како функција на фреквентните опсези и разгледаните метрики за конективност. (в, г) Податочни множества за Паркинсонова болест (EEG) и шизофренија (EEG), за различни вредности на должината τ од временскиот прозорец што се користи за проценка на корелациите. (д, е) Метрики засновани на врски и јазли, за податочното множество за шизофренија (EEG).

Глава 4

Статистички методи и методи базирани на машинско учење за селекција на карактеристики кај мозочни функционални мрежи

Од почетокот на модерната невронаука, главен фокус во истражувањата се става на опишување на разлики помеѓу групи популации, при што едната од нив обично опфаќа луѓе кои страдаат од дадена болест, а другата одговара на здрави контролни субјекти. Целта е да се опише што е значително различно помеѓу здравите субјекти и пациентите, па оттаму што е потенцијален предизвик на таа состојба и, идеално, како може да се ублажи нејзиното влијание. Ова исто така носи една друга придобивка, т.е. валидација; ако не се детектираат разлики кај пациенти кои страдаат од состојба што длабоко ги менува когнитивните способности, како на пр. Алцхајмерова или Паркинсонова болест, може да се заклучи дека податоците користени во споредбата не карактеризираат важни аспекти на мозочната активност за таа состојба.

Во последниве години, невронауката минува низ промена во фокусот на истражување. Историската слика на мозокот е една збирка региони кои работат на доста независен начин, така што може да се воспостави јасна асоцијација помеѓу регионите и когнитивните задачи - на пр. Брока областа и говорот, или Верник областа и разбирањето на јазикот [146]. Сепак, ова е во спротивност со две типови на студии: студии за активации на мозокот, кои покажуваат дека повеќе региони заедно непречено придонесуваат за едноставни задачи; и студии за пластичност на мозокот, процес што овозможува една функција да се префрли на околните

мозочни региони. Се појавува нова парадигма, во која различни области можат да работат и заедно и изолирано, што резултира во соодветна интеграција и поделба на функциите и информациите. Иако ова е одамна претпоставено, само неодамнешното воведување на комплексна мрежна теорија овозможува ригорозни математички студии. Комплексните мрежи се математички објекти составени од јазли, заедно во парови на врски (ребра, линкови) [147, 11, 148]. Јазлите и врските потоа можат да се користат за соодветно да ги претставуваат мозочните региони и нивните интеракции, давајќи структурни и функционални мозочни мрежи [18, 149].

И покрај значајните успеси, употребата на репрезентации на мозочни мрежи воведува и концептуални и пресметковни проблеми, а еден од нив е комплексноста при вршење на гореспоменатата споредба помеѓу две групи на популации. За илустрација, да го разгледаме едноставниот случај на локализиран тумор на мозокот. Кога се анализира преку fMRI, па дури и податоци за ЕЕГ, некој доктор би лоцирал регион со пониска (или скоро нула) активност; следејќи ја оригиналната рамка, тој/таа би заклучил која област е засегната, а со тоа и која функција. Работите сепак стануваат посложени со мрежната рамка. Од една страна, еден регион може да биде (директно или индиректно) вклучен во многу когнитивни задачи. Од друга страна, и од пресметковна перспектива, анализата има поголема димензионалност, бидејќи треба да се утврди не само која област е засегната, туку и кои врски се вклучени. Комбинирајќи ги и двете идеи, можно е некое општетување во регион a да влијае, на пр. преку пластичност, на конективноста помеѓу два други региона b и c , региони што не се поврзани со a .

Кога групната споредба се изведува преку автоматизирани алатки, како на пр. преку модели за машинско учење, мрежните претстави доведуваат до друг проблем во димензионалноста [150]. Накратко, бројот на карактеристики (овде на врски, број што квадратно се мери со бројот на јазли) може да стане поголем од бројот на инстанци, т.е. на субјекти што се достапни во анализата; ова, пак, може да доведе до прекумерен over-fitting на моделите за класификација, а со тоа и до несигурни и негенерализирани резултати. Присуството на голем број потенцијално ирелевантни врски може дополнително да ја намали точноста на алгоритмот за учење и да ги зголеми меморијата и пресметковните барања. Избирањето карактеристики во анализата на мозочната мрежа може да помогне во откривање на значајни биомаркери за одредена мозочна болест.

Ваквото зголемување на комплексноста може делумно да се реши со употреба на процес на *избор*, т.е. почетен чекор посветен на идентификување само на оние врски што се релевантни за дадена студија. Ова е слично на задачата за *избор*

на карактеристики при машинско учење, според која првичното множество на карактеристики што ги опишуваат инстанците на проблемот се филтрира за да ги избрише оние што не се релевантни. Докторот тогаш може да ја игнорира целосната мрежа, само да се фокусира на подмножеството на јазли и врски што се чини дека се поврзани со состојбата што се изучува. Ова не само што резултира во концептуално поедноставен проблем: тој исто така може драстично да ја намали пресметковната комплексност на истиот, со што се овозможува употреба на покомлексни нумерички техники во следните чекори.

Со оглед на оваа состојба, останува едно суштинско прашање: од големиот број на методи предложени во литературата за да се изврши таков избор, како тие се споредуваат едни со други кога се применуваат на мозочните мрежи? Овој дел од оваа дисертација одговара на ова прашање преку голем преглед и евалуација на најчесто користените стратегии за избор, засновани врз принципи за статистика и машинско учење. Евалуацијата се изведува со помош на добро познато и јавно достапно функционално податочно множество за магнетна резонанца (fMRI), кое опфаќа и пациенти со шизофренија и соодветни контролни субјекти; и понатаму се потврдува со две дополнителни податочни множества. Методите се рангираат според количината на информации што ја задржуваат, мерена преку резултатот постигнат во последователна задача за класификација; нивната стабилност, т.е. колку е конзистентен збирот на задржани врски; и нивната пресметковна цена. Резултатите покажуваат дека повеќето методи се однесуваат на квалитативно сличен и ефикасен начин, во смисла дека последователните задачи за класификација не влијаат негативно. Сепак, множествата задржани врски силно се разликуваат во методите; додека е зачувана истата количина на информации, секој метод ја локализира на различен начин, со што нуди комплементарни прегледи на структурата на мозокот. Меѓу најдобрите методи за изведување, вреди да се потенцира NBS (мрежна заснована статистика) за најголемо количество задржана информација, AnovaNet за највисока стабилност и ExT (Extra trees) за најниска пресметковна цена.

Остатокот од текстот е организиран на следниов начин. Сек. 4.1 претставува преглед на методи за избор на врска, засновани врз принципи за статистика и машинско учење. Потоа, сек. 4.2 дава детали за трите податочни множества што се користат во овој дел од дисертацијата и за тоа како мозочните мрежи се реконструирани од нив; и моделите на класификација кои се користени за проценка на ефективноста на методите за избор. Резултатите се презентирани во сек. 4.3, организирани во однос на перформансите (сек. 4.3.1), стабилност (сек. 4.3.2), пресметковна цена (сек.4.3.3) и генерализирање (сек. 4.3.4); На крај,

Статистика

Униваријантни Мултиваријантни

FDR
Bonferroni
KBest

Corr
NBS
aSPU
Hotelling

DDT
AnovaNet
Nevada

Машинско учење

Филтрирачки Обвиткувачки Вградливи

SURF

SVM-RFE
Bouta
LR-RFE
SLDA

LASSO
Elastic-Net
RF
Ext

Слика 4-1: Преглед на методите за избор на карактеристики, класифицирани според нивната природа. Заради јасност, кодот за боја од иста класа ќе се користи во сите следни слики.

сек. 4.4 извлекува некои заклучоци, дискутира за ограничувањата на сегашното истражување и предложува идни линии на истражување.

Оваа глава од дисертацијата е објавена во трудот Ilinka Ivanoska, Kire Trivodaliev, Slobodan Kalajdziski, and Massimiliano Zanin. “Statistical and Machine Learning Link Selection Methods for Brain Functional Networks: Review and Comparison.” *Brain Sciences* 11, no. 6 (2021): 735. IF 3.394.

4.1 Методи за селекција на линкови кај функционални мозочни мрежи

Како што претходно беше спомнато, во овој дел ќе споредуваме методи - видете Сл. 4-1 за избор на врски во мозочни мрежи. Овие можат да бидат поделени во две големи семејства, т.е. оние засновани на статистички методи и оние засновани на модели за машинско учење. Овие две семејства понатаму можат да бидат поделени во различни групи. Поточно, статистичките методи можат да бидат униваријантни, т.е. да се оценува една единствена врска во едно време (види сек. 4.1.1) или мултиваријантни, кога секоја врска се оценува земајќи ги предвид и другите (сек. 4.1.2). Од друга страна, методите за машинско учење може да се класифицираат како филтрирачки, во кои врските се избрани како чекор до класификацијата (сек. 4.1.3); обвиткувачки, т.е. кога изборот е извршен со употреба на резултат на модел на класификација (сек. 4.1.3); и вградливи методи, во кои процесот на селекција се изведува со гледање на внатрешниот процес извршен од модел на класификација (сек. 4.1.3). Имајте на ум дека оваа класификација е нејасна и има за цел само да обезбеди општ водич за главниот основен принцип - за илустрација, секој униваријантен метод коригиран за повеќе споредби технички ги разгледува сите врски во последниот чекор и затоа може да се класифицира

како мултиваријантен.

4.1.1 Статистички, униваријантни (локални) методи

Лажна стапка на откривање (False discovery rate FDR)

Пристапот со лажна стапка на откривање (FDR) работи со еднодимензионални податоци, така што врските помеѓу различни парови на јазли се сметаат како независни случаи. Се споредуваат тежините на соодветните врски на двата примерока, за потоа да се утврди дали разликата е статистички значајна за секоја одделна врска. Овој пристап ја користи матрицата на соседство на мозочните мрежи и ги споредува тежините на врските што припаѓаат на долниот триаголен дел од матрицата (полу-векторизација на матрицата). Таквата споредба може да се направи преку едноставен студентски t -тест. Тука го користиме Welch-овиот тест [151, 152], варијација на t -тестот за случаи во кои двата примерока имаат различна варијанса.

Тестирање на повеќе врски на независен начин резултира во добро познатиот проблем на повеќе споредби. Поточно, треба да се контролира веројатноста за едно или повеќе лажни отфрлања, исто така познати како грешки од тип I или family-wise стапка на грешки (Family-wise error rate FWER). Една од неколкуте постапки за решавање на ова прашање е Лажната стапка на откривање (False discovery rate FDR) [153], која обезбедува помалку строга контрола на грешка од типот I, придонесувајќи за поголема статистичка моќ. Од друга страна, овој метод резултира и со зголемен број на лажни одбивања (грешки од тип I). Оваа процедура за контрола работи со подредување на вредностите p во растечки редослед, за потоа да се најде најголемиот индекс i што го задоволува условот $p_i \leq \alpha_i/N$ (со $i = 1, 2, \dots, N$ и N се бројот на тестови или врски). Со други зборови, тој наоѓа најголем индекс i за кој е отфрлена нултата хипотеза (двата примерока се различни).

Униваријантниот пристап на FDR може да се сумира во следниве чекори: 1) пресметај ја полу-векторизацијата на матриците за мозочните мрежи; 2) изврши го Welch-овиот тест на секој елемент од векторот; 3) примени ја постапката FDR за да се поправат p вредностите; 4) отфрли ја нултата хипотеза за секоја врска ако $pvalue < \alpha$; 5) подредете ги врските според нивната p вредност; и 6) извлечете ги значително различните врски.

Bonferroni

Како што претходно беше дискутирано, постојат неколку процедури [154, 155, 156, 157, 158] кои го разгледуваат прашањето за повеќе споредби користејќи вредност на корекција, со цел да се осигура дека FWER се одржува под одредено ниво. Една од нив е корекција на Bonferroni [159], што е силен метод за контрола на вредностите. Овој метод поверојатно дава повеќе лажно негативни вредности, со што се намалува статистичката моќ на тестовите. Постапката за избор на врската со користење на корекцијата на Bonferroni е слична на претходната, само FDR се заменува во чекор 3 со Bonferroni p вредностите.

Избор на K најдобри (KBest)

Методот KBest се заснова на проценка на функција (Chi-квадрат статистика во нашиот случај) што ја проценува корелацијата помеѓу карактеристиките (т.е. тежините на врската) и целната променлива (ознаките на класите). Конечно се избираат фиксен и предодреден k број на врски со највисоки оценки.

4.1.2 Статистички, мултиваријатни (глобални) методи

Корелација на карактеристики (Corr)

Овој метод ја користи идејата дека добрите карактеристики се во голема корелација со целната променлива, но во исто време и неповрзани меѓу себе. Следствено, ако две карактеристики се во голема корелација, се задржува само онаа што е во поголема корелација со целната променлива. Праг од 0,5 се користи за да се утврди дали паровите карактеристики се линеарно корелирани меѓу себе.

Статистика заснована на мрежа (Network Based Statistics - NBS)

Статистиката заснована на мрежа (Network-based statistics NBS) [160] е метод кој врши mass-univariate тестирање на мрежните врски, со последователна контрола за FWER. Кластерирачката структура на мрежата, т.е. поврзаните подмрежи во тополошкиот простор, се користи за корекција на почетната тест статистика и p вредностите, а резултатот е збир на компоненти што се статистички значајни. Тестот го контролира FWER во слаба смисла и дава корегирани p вредности на секоја подмрежа користејќи пермутирано тестирање, со цел да се заклучи дали секоја компонента има статистички значаен ефект во мрежата.

Подетално, NBS работи на следниот начин. Првиот чекор вклучува пресметување на индивидуална униваријантна тест статистика за секоја врска, како на

пр. t -тест статистика. После тоа, се применува праг дефиниран од корисникот, што резултира во збир на врски над прагот. Овој праг е минимална вредност на веројатносната статистика на врската што може потенцијално да ја отфрли нултата хипотеза. Следниот чекор е да се користи алгоритам за пребарување по ширина на множеството врски, за да се пронајде бројот на добиените поврзани компоненти и нивните големини. Овие компоненти може да се сметаат како под-мрежи што содржат врски што имаат статистички значаен ефект. Големината на најголемата компонента е зачувана; дополнително, и со употреба на повторени пермутационски тестови, членовите на примероците случајно се пермутираат и се чува големината на најголемата компонента. Конечно, врските на крајната компонента се задржуваат за понатамошни анализи. NBS доаѓа со добивка на статистичка моќ, но нејзината слабост е што не може да се локализира ефектот на одделните врски, бидејќи наместо тоа, се задржуваат целосните поврзани компоненти.

Предложени се неколку подобрувања и варијации на постапката за тестирање со NBS. За илустрација, варијанта без праг е опишана во [161]. Предложена е и специфична постапка за прилагодување на големините на кластерите во [162] и имплементирана во рамката за пермутационо тестирање во софтверот Mrtrix connectomestats (<https://mrtrix.readthedocs.io/en/latest/reference/commands/connectomestats.html>). Конечно, други високо прилагодливи тестови на податоците кои избираат неколку параметри за да ја рефлектираат несигурноста со добиената моќ, се предложени во [163].

Тука ја искористивме основната имплементација на NBS достапна на <https://github.com/aestrivex/bctpy/blob/master/bct/nbs.py>, со grid пребарување за да го најдеме оптималниот праг што води до најдобрата класификација.

Адаптивна сума на степенски бодуван тест Adaptive sum of powered score test (aSPU)

Сума на степенски бодуван тест (sum of powered score test SPU), е статистички тест првично предложен во [164] за идентификување на ретки варијанти во студии за асоцијација на широк геном, а последователно се користи на мозочни мрежи [165]. Овој пристап обезбедува глобално тестирање, односно значајниот ефект не може да се поврзе со одделни врски.

Наједноставниот SPU тест пресметува бодуван вектор U како $U = \sum_{i=1}^n X_i(Y_i - \hat{Y}_i)$, каде што X_i се набљудувања (мозочна мрежа полувекторизирана) за индивидуата i , Y е бинарен вектор што ја означува групата на поединци, и \hat{Y} средната вредност на овие. Секој елемент U_j содржи не-нулти

ефект на секоја врска во мрежата. Овој вектор со бодови се користи за пресметување на веројатносната статистика како $T_{SPU} = \sum_{j=1}^k U_j^\gamma$, каде што γ е влезен параметар што треба да се подеси со цел да се добие најголема статистичка моќ. За да се направи статистички заклучок, следниот чекор е да се изврши параметарски bootstrapping или пермутации, за добивање на p корегирани вредности.

Адаптивниот збир на бодуван тест (aSPU) е варијација насочена кон наоѓање на оптимална вредност на параметарот γ [164]. Ако $\Gamma = 1, 2, \dots, \infty$ и $P_{SPU}(\gamma)$ е p -вредност на тестот за SPU со параметар γ , тогаш статистиката за тестот aSPU се пресметува како $T_{aSPU} = \min_{\gamma \in \Gamma} P_{SPU}(\gamma)$. p вредноста на тест статистиката се пресметува преку bootstrap или пермутации. [164] ја демонстрира употребата на бодуваната компонентата U_j како придонес на секоја врска во тест статистиката. Така, може да се подредат врските според нивните U_j , и да се изберат врвните k како најважни k врски. Ние ја искористивме R имплементацијата на основниот тест за aSPU достапна на <https://github.com/ikwak2/aSPU>.

Т-квадрат Хотелинг тест

T^2 Хотелинг статистика [166, 167] е природно мултидимензионално генерализирање на студентската t -статистика. Овој тест ги споредува средните вредности на примероците што содржат податоци со повеќе променливи (карактеристики). Со цел да се изврши избор на врска, се започнува со пресметување на N Лапласови матрици, една по субјект, како $L = D - A$, каде што D и A се соодветните матрици за степен и соседство. После тоа, пресметани се две Fréchet средини, соодветно на Лапласовите матрици на двете групи на субјекти; овие усреднети матрици потоа се трансформираат во вектори преку полу-векторизација и се означуваат понатаму со $vec(\hat{L}_1)$ и $vec(\hat{L}_2)$.

Тогаш ² Хотелинг статистиката се дефинира со

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (vec(\hat{L}_1) - vec(\hat{L}_2))^T \hat{\Sigma}^{-1} (vec(\hat{L}_1) - vec(\hat{L}_2)), \quad (4.1)$$

каде $\hat{\Sigma}^{-1}$ е инверзна на коваријансната матрица $\hat{\Sigma} = \frac{1}{n_1 + n_2 - 2} \sum_{j=1}^2 n_j \hat{\Sigma}_j$, а $\hat{\Sigma}_j$ ја означува коваријансната матрица на посебниот примерок

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (vec(L_i) - vec(\hat{L})) (vec(L_i) - vec(\hat{L}))^T \quad (4.2)$$

и n_1 и n_2 се големините на двете групи примероци.

Според [167], кога големината на примерокот е доволно голема, T^2 е приближно χ^2 -квадрат дистрибуирана со p степени на слобода. Кога важи хипотезата за

нормалност, Хотелинг тестот е оптимален.

Со оваа дефиниција на T^2 статистиката, следниот чекор е да се пресмета матрицата на коваријанса. Како што е наведено во [168], примерокот (емпириска) коваријанса врз основа на набљудуваните податоци е сингуларна кога димензијата е поголема од големината на примерокот ($n \ll p$), т.е. не може да се сврти за да се пресмета прецизната матрица. Исто така, матрицата на коваријантност на примерокот веројатно е нумерички нестабилна. Затоа, треба да се користи метод за проценка за добивање соодветна матрица на коваријанса. Развиени се повеќе проценувачи како решение за проблемот и тие се применуваат во различни полиња [169]. Дел од нив се проценувачи на матриците на коваријанса [169], некои се за ретки матрици на коваријанса [168, 170], а други се користат за проценка на инверзната на матрицата на коваријанса [171]. Ние овде користиме форма на проценка на намалување [172] што ја намалува матрицата на коваријанса на примерокот кон идентитетската матрица и што е еквивалентно на разгледување на асимптоматска оптимална конвексна линеарна комбинација на двете. Имплементацијата одговара на онаа во [173]. Потоа ја пресметуваме матрицата на коваријанса од индивидуалните проценети матрици на коваријанса, T^2 статистиката, степенот на слобода и p вредноста. Користејќи p -value $< \alpha$, утврдуваме дали примероците се различни.

Придонесот на секоја врска во T^2 статистиката може да се одреди преку линеарна декомпозиција на T^2 . Со земање на квадратниот корен на инверзната коваријанса матрица, може да се направи декомпозиција $\frac{n_1+n_2}{n_1n_2}T^2 = A^T A = \sum_{i=1}^p A_i^2$, каде $A = \widehat{\Sigma}^{-\frac{1}{2}}(vec(\widehat{L}_1) - vec(\widehat{L}_2))$, и A_i^2 е придонесот на врската i во вредноста на T^2 . Со ова, инверзната матрица се декомпозира преку спектрална декомпозиција и се проценуваат тежините за врските. Конечно, врските се сортираат според нивната тежина и се задржуваат најгорните k врски.

Тест на разлики во степен (Difference degree test DDT)

Тестот за разлика во степенот (DDT) [174] е метод за споредба на мозочните мрежи предложен за надминување на ниската статистичка моќ на mass-униваријатните пристапи. Тестот ја открива мрежата на разликата во степенот што ги идентификува регионите на мозокот поврзани со значителен број на диференцијални тежински рабови (differentially weighted edge DWE).

Првиот чекор од тестот вклучува изградба на мрежа на разлика $D = d_{ij}$, $i, j \in 1, 2, \dots, N$; каде секој елемент d_{ij} претставува статистичка разлика во врските помеѓу јазлите i и j , и се дефинира како p -вредност на тест за разликата помеѓу групата. После тоа, првиот и вториот момент на разликата во мрежата

D се добиваат како $\bar{e} = E(\bar{d}_{ij})$ и $\sigma^2 = Var(\bar{d}_{ij})$, според кој се генерираат одреден број на случајни мрежи со употреба на ефикасен алгоритам предложен во [175]. Така, случајните мрежи се совпаѓаат со набљудуваната мрежа за разлика, за да се долови нејзината вистинска топологија. После тоа, се применува адаптивен метод на избор на праг за ограничување на D . Во нашата анализа, го искористивме тестот за емпириска разлика на степенот (empirical difference degree test eDDT), кој ја користи емпириската критична вредност за праг. Избраниот праг се применува на D и се добива матрицата на соседството на разликата $A = a_{ij}$, каде a_{ij} ги претставува разликите во врската помеѓу јазлите i и j . Врз основа на A , се мери мерка за степен на разлика d_i за секој јазол i . d_i е моделиран со биномна дистрибуција $B(N - 1, p_i^{null})$, каде што p_i^{null} е очекуваната веројатност за секоја врска на јазолот i за да се демонстрираат разликите во групата и се пресметува врз основа на генерираните случајни мрежи. Конечно, се проценува статистичкото значење на DWE за секој јазол. Ние овде предлагаме да ја креираме листата на релевантни врски со избирање на оние што се поврзани со статистички различен јазол.

Анова тест за мрежи (AnovaNet)

Друг ново развиен статистички тест за идентификување на разликите во популациите на мрежите е тестот за анализа на варијанса (ANOVA) (AnovaNet). [176]. Тој има за цел идентификување на подмрежите каде се појавуваат различни видови на модификации помеѓу групите: модификации локализирани токму во одредени врски; нелокализирани модификации, во кои некои врски се менуваат, но тие се разликуваат меѓу поединците; и глобални модификации, во кои има глобални измени меѓу групите. Како резултат, алгоритмот дава подмножество на врски или подмрежа, каде што е присутна најголемата мрежна разлика помеѓу групите.

За даденото множество на мозочни мрежи G_1, G_2, \dots, G_n , тестот AnovaNet споредува две групи со пресметување на статистиката $T = \frac{\sqrt{2}}{a} \sum_{i=1}^2 \sqrt{n_i} (\frac{n_i}{n_i-1} \bar{d}_{G_i}(M_i) - \frac{n}{n-1} \bar{d}_G(M_i))$, каде n_i ($i \in 1, 2$) е бројот на мрежи во секоја група, $n = n_1 + n_2$, a е константа на нормализација, $\bar{d}_G(M) = \frac{1}{l} \sum_{k=1}^l d(G_k, M)$ е просечното растојание околу мрежа (пресметано со користење на растојанието помеѓу графовите [177]) и M_i ($i \in 1, 2$) е просечната матрица на соседството за групата i . Со други зборови, оваа статистика ја споредува мрежната варијабилност за секоја група од една страна, и просечното растојание помеѓу мрежите на едната група и нивниот соодветен просек од друга страна.

Како последен чекор, дефинирана е дополнителна статистика $T_{\bar{E}}$, како T статистика добиена за подмрежата составена од сите јазли и подмножество \bar{E} од сите врски. Врските \bar{E} кои го минимизираат $T_{\bar{E}}$ се оние кои најмногу придонесуваат за разликата меѓу групите и затоа се задржуваат за последователни анализи.

Анализа на мрежни податоци **NEtwork-VALued Data Analysis (Nevada)**

NEtwork-VALued Data Analysis (Nevada) тестот [178] е пермутациски тест за правење заклучоци за дистрибуција на мрежи. Предложената рамка е многу флексибилна, бидејќи овозможува избор на различни матрични претстави на мрежите, различни метрички растојанија и различна тест статистика.

Nevada може да се прилагоди за да изврши избор на врска како што следува во продолжение. Првиот чекор вклучува избор на репрезентација на матрицата, меѓу неколкуте достапни, што дава најдобри резултати: матрици за соседство, Лапласови или модуларност. Второ, мора да се избере статистика; голем број од нив се разгледуваат, предлагаат и користат во [178]. За дадени две групи $i = 1, 2$ и соодветните мрежи $G_i = \{G_{i1}, G_{i2}, \dots, G_{in_i}\}$, ние ја користевме T статистиката за откривање на повисоки разлики помеѓу мрежите дефинирана како

$$T = \frac{\frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} d^2(G_{1i}, G_{2j}) - (\hat{\sigma}_1^2 + \hat{\sigma}_2^2)}{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}. \text{ Тука, } \hat{\sigma}_1^2 \text{ и } \hat{\sigma}_2^2 \text{ се проценувачи на варијанс}$$

сите во рамките на примерокот во двете групи и се дефинирани како $\hat{\sigma}_1^2 = \frac{1}{n_1(n_1-1)} \sum_{i=1}^{n_1} \sum_{j>i}^{n_1} d^2(G_{1i}, G_{1j})$ и $\hat{\sigma}_2^2 = \frac{1}{n_2(n_2-1)} \sum_{i=1}^{n_2} \sum_{j>i}^{n_2} d^2(G_{2i}, G_{2j})$. Забележете дека растојанието $d(G_k, G_l)$ помеѓу две мрежи G_k и G_l е овде дефинирано како соодветно Frobenius растојание. Како последен чекор, пермутациски тест е дизајниран со непараметарски комбиниран пристап кој ги собира разликите во средната и варијансата помеѓу мрежните групи со висока статистичка моќ.

Слично на пристапот на AnovaNet, се избира подмножеството на врски што ја дава оптималната вредност на статистиката, а со тоа и најголемите разлики помеѓу двете групи. Во нашата имплементација го искористивме пакетот Nevada (NEtwork-VALued Data Analysis) R преземен од <https://github.com/astamm/nevada>.

Вреди да се напомене дека се предложени и други статистички методи за откривање на групни разлики во мрежните структури - заинтересираниот читател може да се повика на: [179, 180, 181, 182, 183]. Овие методи не се вклучени овде заради практични и имплементациони проблеми, како на пр. премногу големи пресметковна цена.

4.1.3 Машинско учење

Изборот на карактеристики е активно поле на истражување во рамките на заедницата за машинско учење и како такви се предложени голем број алгоритми и методи. Иако во принцип сите може да се применат на функционални мозочни мрежи, целосен преглед го надминува опсегот на овој придонес - преглед што заинтересираниот читател може да го најде, на пример, во [184, 185]. Тука, само дискутираме и оценуваме некои од најрепрезентативните и најпознатите методи. Заради јасност, методите понатаму се поделени во три семејства, следејќи ја вообичаената класификација во машинското учење: филтрирачки, обвиткувачки и вградливи методи [184].

Филтрирачки методи

Филтрирачките методи за избор на карактеристики користат одредени метрики за рангирање на карактеристиките и за формирање на најдоброто множество на атрибути [186]. Тие се независни од алгоритмот за учење и затоа најчесто се користат како техника за претходна обработка. Филтрирачките методите ја одредуваат релевантноста на карактеристиките со нивно корелирање со зависната варијабла, користејќи метрика како што се информациска добивка, Gini индекс и други [187, 188, 189].

Spatially Uniform Relief (SURF)

Relief [190] е метод за рангирање на карактеристики заснован на инстанци за проблем со бинарна класификација. ReliefF [191] е продолжение на Relief за повеќекласни проблеми. Relief ги рангира карактеристиките во зависност од тоа колку добро тие прават разлика помеѓу инстанците блиски едни до други.

Алгоритмот случајно избира инстанца G_i од податоците и го наоѓа најблискиот сосед H од истата класа и најблискиот сосед M од друга класа. Потоа, се ажурира ранг-резултатот на карактеристиките споредувајќи ги вредностите на карактеристиките на G_i со H и M . Ако G_i и H имаат различни вредности за карактеристиката f , т.е. двете инстанци од иста класа се одделени со f (што не е пожелно), ранг-резултатот е f е намален. Од друга страна, ако карактеристиката f ги одвојува G_i и M (што е пожелно), тогаш резултатот на f е зголемен. Процесот се повторува за случајно избрани m инстанци. ReliefF е исто така поробуствен од Relief користејќи k најблиски соседи. ReliefF ги рангира сите карактеристики и му треба предефиниран праг t за избор на најдобри N карактеристики.

Алгоритмот SURF [192] се надоврзува на ReliefF, но, за разлика од овој, го елиминира корисничкиот параметар k и наместо тоа, усвојува праг на растојание T за да утврди кои случаи ќе се сметаат за соседи. Ние ја искористивме имплементацијата на SURF вклучена во scikit-learn пакетот во Python достапна на <https://github.com/EpistasisLab/scikit-rebate>.

Обвиткувачки методи

Обвиткувачките методи користат збир на карактеристики за да научат модел, за потоа да ги користат параметрите на моделот за да го оценат изборот на карактеристики. Со други зборови, тие одлучуваат дали да додадат или отстранат карактеристики од подмножеството со проценка на влијанието што тие карактеристики го имаат во резултатот на класификација. Проблемот во суштина се сведува на проблем за пребарување, што го прави обично пресметковно многу скап. Дополнително, постои зголемен ризик од overfitting, особено кога бројот на набљудувања е мал. Добро познати примери на обвиткувачки методи се избор на карактеристики нанапред, елиминација на карактеристиките наназад и елиминација на рекурзивните карактеристики. Ние користиме пристап на рекурзивна елиминација користејќи го моделот за учење SVM (Support Vector Machine - Recursive Feature Elimination SVM-RFE) и логистичката регресија (LR-RFE), како и Boruta.

Support Vector Machine-Recursive Feature Elimination (SVM-RFE)

Овој метод, наречен векторска машина за поддршка-рекурзивна елиминација на карактеристики (SVM-RFE), се заснова на изведување на рекурзивна елиминација на карактеристика (во која карактеристиките се бришат едновремено) под водство на линеарен класификатор SVM [193]. SVM алгоритмот го дели даденото множество на инстанци со користење на хиперрамнина со максимална маргина, т.е. хиперрамнина што е максимално одделена од двете класи на инстанци. Целта е да се најде функција за поделба што точно ќе ги класифицира новите примероци. Параметрите на линеарен SVM дефинираат хиперрамнина на одлука во мултидимензионалниот простор на карактеристиките, т.е. $g(x) = w^T x + b = 0$, каде што x е вектор на карактеристики, w е вектор со тежини за карактеристиките, и b е bias параметар.

Според алгоритмот SVM-RFE, релевантноста на елементот n во векторот на карактеристиките се одредува според вредноста w_n од тежинскиот вектор. Практично, ако w_n е близу до 0, соодветната карактеристика не придонесува значител-

но во вредноста на $g(x)$. Важноста на секоја карактеристика потоа се одредува со сортирање на апсолутните вредности на векторот w и, при секое повторување на алгоритмот, одреден број карактеристики се отфрлаат.

Logistic Regression with Recursive Feature Elimination (LR-RFE)

Моделот на логистичка регресија (Logistic regression LR) ги предвидува параметрите на логистичкиот модел со цел да се моделираат веројатностите на класата. Секоја карактеристика на податоците за обука е поврзана со број помеѓу 0 и 1, што претставува веројатност да припаѓа на дадена класа. Основната идеја за логистичка регресија е користење на веќе развиен механизам за линеарна регресија со моделирање на веројатноста p_i со употреба на функција за линеарно предвидување, односно линеарна комбинација на дадени атрибути и збир на коефициенти на регресија. Функцијата за линеарно предвидување f е дадена со $f(x) = w * x + b = \log(p(x)/(1 - p(x)))$, каде w е вектор со коефициенти на регресија, x е вектор на предефинирани атрибути на податоци, а b се однесува на bias променлива. На начин сличен на пристапот SVM-RFE, користиме алгоритам за елиминација на рекурзивната карактеристика за да ја одредиме важноста на секоја од карактеристиките, да ги сортираме и земаме специфичен број на карактеристики (тука, врски) како значајни.

Boruta

Алгоритмот Boruta [194] е обвиткувачки метод за избор на карактеристика околу алгоритмот за класификација на случајна шума (Random forest). Алгоритмот прво го удвојува податочното множество и ги менува вредностите на карактеристиките, кои се нарекуваат карактеристики на сенка, во новосоздадената. Следниот чекор вклучува обука на алгоритам Random forest за спојување на оригиналните и модифицираните множества на карактеристики. Вредностите за значење на карактеристиките се земени од обучениот модел. Алгоритмот потоа проверува кои од вистинските карактеристики биле поважни од модифицираните карактеристики. Случајната промена на карактеристиките, учењето на алгоритмот, како и проверката на значењето на реалните карактеристики се повторуваат N пати и за секоја карактеристика се пресметува бројот на погодоци (број на повторувања во кои таа карактеристика има значење поголемо од максималното значење на изменетите карактеристики).

Потоа, се изведува статистички тест во кој карактеристиката е одбиена ако бројот на погодоци е помал од очекуваната вредност, а се задржува во спротивен

случај. При имплементацијата, се користи двостепен метод за корекција за да се поправи грешката што се јавува при повеќекратно тестирање. Прво, се користи корекција на FDR, а потоа и корекција на Bonferroni за да се забележи дека истите карактеристики се тестираат одново и одново во секоја итерација со истиот тест.

Тука ја користиме имплементацијата на овој алгоритам во Python преземена од https://github.com/scikit-learn-contrib/boruta_py, што има модификации за да ги олабави строго дефинираните правила во оригиналниот код во R, конкретно, тој воведува перцентил за замена на фиксниот праг на одбивање.

Sparse linear discriminant analysis (SLDA)

Процесот на избор на карактеристики за ретка линеарна дискриминаторска анализа (Sparse linear discriminant analysis SLDA) за мозочните мрежи е предложен во [195], и е заснован на две фази со употреба на техника за машинско учење, која ја вклучува реткоста на тежинските мрежи. Во првата фаза од изборот на карактеристиките, се прави ретка линеарна анализа за дискриминација за да се изберат карактеристики на дискриминација. Различни подмножества на податочното множество на мозочните мрежи се итерираат со мултиваријантен bootstrap пристап. Покрај тоа, исто така, во првиот чекор е конструиран ансамбл на ретки линеарни дискриминаторски модели на анализа (Linear discriminant analysis LDA), користејќи ја идејата за реткост од [196] за да се најдат карактеристиките (мрежни врски) што ги дискриминираат двете групи на мозочни мрежи. Подмножеството на избраните карактеристики е минимизирано, а разликата помеѓу двете групи е максимизирана. Вториот чекор е чекор за избор на стабилност, каде што се чуваат само карактеристиките што често се избираат низ итерациите.

Дефиницијата на SLDA проблемот користи l_1 и l_2 регуларизација како кај проблем со еластична мрежа Elastic-Net (види подолу). Регуларизацијата l_1 обезбедува намалување на overfitting, принудувајќи на избор на мал број карактеристики. Понатаму, регулацијата l_2 ја минимизира грешката во предвидување и помага при управување со димензионалноста на податоците. Двете комбинирани даваат оптимален класификатор и минимален број на карактеристики на дискриминација.

За наша имплементација користевме R алтернатива на оригиналниот код на Matlab (<https://github.com/alecrimi/multi-link>) преземен од <https://github.com/topepo/sparselda>.

Вградливи методи

Вградливите методи ги комбинираат квалитетите на филтрирачките и обвиткувачките методи. Тие се имплементираат со алгоритми кои имаат свои вградливи методи за избор на карактеристики. Алгоритмот за учење користи сопствен процес на избор на карактеристики и истовремено врши избор на модел за класификација/учење.

Lasso

Оператор на најмало апсолутно намалување и селекција (Least Absolute Shrinkage and Selection Operator Lasso) е метод на анализа на регресија кој вклучува избор и регуларизација на карактеристиките со цел да се подобри точноста и толкувањето на добиениот статистички модел [197]. Lasso е во состојба да ги постигне овие две цели со поставување на фиксна вредност за која збирот на апсолутните вредности на коефициентите на регресија мора да се минимизира, принудувајќи одредени коефициенти да бидат поставени на нула, и ефикасно избирајќи поедноставен модел кој не вклучува неважни коефициенти. Користената имплементација се заснова на l_1 регуларизација во логистичка регресија во Python.

Elastic-Net

Друг добро познат начин за додавање на регуларизација е методот Ridge (l_2 регуларизација), кој користи квадрат наместо апсолутна вредност како во Lasso. Методот познат како Elastic-Net [198] истовремено ги користи методите Ridge и Lasso: $L + \lambda((1 - \alpha) \sum w^2 + \alpha \sum |w|)$, каде што α ја контролира мешавината на Ridge и Lasso регуларизации, со вредност од $\alpha = 0$ што означува “чист” Ridge и $\alpha = 1$ “чист” Lasso. Нашата имплементација го користи класификаторот SGDClassifier од пакетот scikit-learn [173] во Python, со Elastic-Net регуларизација со $\alpha = 0,15$.

Random forest (RF)

Алгоритмот за класификација на случајни шуми (random forest RF) има вградена пресметка за важноста на карактеристиките, опфаќајќи го придонесот на карактеристиките во намалувањето на тежинската нечистост на поделбата, додека се обучуваат дрвјата што го составуваат. Карактеристиките се избираат како внатрешен јазол во дрвото според нивното намалување на нечистоста, што,

во случај на задачи за класификација, може да се пресмета преку Gini нечистост или информациска добивка. Просекот на ова намалување на секоја карактеристика кај сите дрвја во шумата е мерка за важноста на карактеристиката. Ги избираме највисоките k карактеристики (врски) со најголемо значење користејќи имплементација на RF класификаторот од пакетот scikit-learn за учење во Python.

Extra Trees (ExT)

На начин сличен на она што е направено во претходниот RF метод, го користиме вграденото рангирање на функциите вклучено во класификаторите за екстремно рандомизирани дрвја (Extra Trees ExT). ExT е ансамбл метод на учење многу сличен на RF, различен само во начинот на кој се градат дрвјата за одлука. За извршување на избор на карактеристика, за време на изградбата на шумата, за секоја од карактеристиките, се пресметува нормалното вкупно намалување во критериумите што се користат при одлуката за поделба (обично со Gini индекс). Карактеристиките се сортираат по опаѓачки редослед според нивната важност и се задржуваат врвните k карактеристики. Слично на претходните методи, користена е Python scikit-learn имплементација со класификаторот Extra Trees.

4.2 Податочни множества и класификациски алгоритми

4.2.1 Аквизиција на мозочните слики и реконструкција на мозочните мрежи

COBRE

Првото податочно множество што се користат во овој дел од дисертацијата е податочното множество за шизофренија COBRE [72], кое вклучува пациенти со шизофренија и контролни субјекти. Опис на ова податочно множество беше даден во претходната глава од оваа дисертација.

PRURIM

Второто податочно множество што се користи тука е преземено од базата на податоци за мултимодална конективност на USC [199] <http://umcd.humanconnectomeproject.org>, што е веб-базирана мрежна платформа и база на

мозочни мрежни податоци. Вклучува 29 функционални мозочни мрежи 116x116 кои произлегуваат од студија со 14 пациенти со хронична псоријаза и 15 здрави контроли. fMRI податоците доаѓаат од Универзитетската болница Брест (CHU де Брест), Франција, стекната со скенер Philips Achieva dStream 3T (fMRI податоци со $TR = 3,15s$, прво време на ехо = 17ms, второ време на ехо = 46ms, големина на воксел = $2,5 \times 2,5 \times 2,6 \text{ mm}^3$, $FA = 90^\circ$, $FOV = 250 \times 250 \times 143$, големина на матрица = 128×128 , број на делови = 45). Извршени се стандардни чекори за претходна обработка (корекција на деловите на време, преуредување, основна регистрација, нормализација, измазнување со 10 mm FWHM, парцелација со AAL-атлас, повлекување на шум фактори и band-pass филтрирање) и се поставени 29 функционални матрици за конективност во базата на податоци за мултимодална поврзаност на USC.

UCSF Progressive Supranuclear Palsy (PSP)

Третото податочно множество што се користи тука е исто така преземено од базата на податоци за мултимодална поврзаност на USC [199]. Вклучува 64 матрици за конективност со големина 27x27 од task-free студија за функционална магнетна резонанца (tf-fMRI) на 12 пациенти со Прогресивна супрануклеарна парализа (Progressive Supranuclear Palsy PSP) и 20 здрави контроли скенирани два пати со 6 месеци разделеност. Конструирани се од податоци за слики од UCSF Neuroscience Imaging Center, Сан Франциско, стекнати со Siemens Trio 3T скенер (tf-fMRI податоци со време на аквизиција = 8min 06sec, аксијална ориентација со interleaved ordering, $FOV = 230 \times 230 \times 129$, големина на матрицата = 92×92 , големина на воксел = $2,5 \times 2,5 \times 3,0 \text{ mm}^3$, $TR = 2000 \text{ ms}$, $TE = 27 \text{ ms}$). Дополнителни информации во врска со чекорите за пред-обработка и реконструкцијата на матриците за конективност се достапни во [200].

4.2.2 Алгоритми за класификација

Откако ќе се добијат повеќе групи на карактеристики, т.е. врски, преку различните методи опишани во сек. 4.1, овие се оценуваат преку проблем на класификација. Моделот на класификација е обучен да прави дискриминација помеѓу двете групи луѓе застапени во податочните множества, само со користење на избраниите врски, а резултатот се оценува во однос на метриците Area under curve (AUC) и F1. Ова се метрики што ги проценуваат перформансите на класификаторот, каде AUC претставува степен за раздвојливост на класи, а F1 резултатот е мерка што балансира помеѓу прецизноста и recall на моделот. Интуитивно, колку е

поголема добиената метрика за класификација, толку повеќе информации релевантни за проблемот со класификацијата се вклучени во користеното множество на врски; следствено, резултатот од класификацијата може да се користи како полномошник за ефикасноста на секој метод за избор на врска [14].

Бидејќи различните модели се засноваат на различни претпоставки за карактеристиките и нивните врски, секоја класификација е извршена со употреба на повеќе алгоритми со повторена stratified 10-fold cross-validation за валидација, со цел да се избегнат пристрасностите специфични за моделот. Користените модели се:

- *Баес*. Исто така познат како класификатор на Баес, овој модел претпоставува целосна независност помеѓу влезните карактеристики на проблемот, за потоа директно да се примени теоремата на Баес и да се процени веројатноста поврзана со секоја класа [201]. Иако има ниска пресметковна цена, неговото главно ограничување е претпоставка за независност, која обично не може да се гарантира.
- *Gaussian Naive Bayes (GNB)*. Тип на класификатор на Баес во кој влезните карактеристики се сметаат за континуирани (не дискретни) и е моделиран со нормална (или Гаусова) дистрибуција.
- *Дрва на одлука Decision Trees (DT)*. Класификациски модели составени од дрвовидна структура на јазли. Секој јазол претставува атрибут (карактеристика) во запис што треба да се класифицира, што се тестира според прашање дефинирано во фазата на обука; секоја гранка тогаш претставува вредност што може да ја земе атрибутот, т.е. одговор на прашањето на јазолот. Главната предност на DT лежи во неговата едноставност и мала пресметковна цена [202].
- *Random Forest (RF)*. Комбинација на предвидувачи за дрво на одлуки, во кои секое дрво е обучено за случајно подмножество на карактеристики и записи; конечната прогноза за класификација се пресметува преку правило за мнозинство. Случајните шуми (Random forests) се особено ценети поради нивната прецизност и малата склоност кон overfitting на класификацијата [203].
- *ExtraTrees (ExT)*. Ансамбл брз метод заснован на комбинација на голем број рандомизирани дрва за одлука на различни подмножества на базата на податоци, на начин сличен на класификаторот RF. Подобрување на предвидливоста на точноста и контролата на overfitting се добива со додавање на рандомизација при избор на оптимални точки за разделување [204].

- *Stochastic Gradient Descent* (SGD). Мета-алгоритам во кој се комбинираат и оптимизираат повеќе линеарни Huber loss функции [205].
- *Ada*. Мета-алгоритам чијшто излез е тежинска сума на големо множество на мали дрвја со одлука, каде што тежините се динамички прилагодени за да се минимизира грешката во класификацијата [206].
- *AdaBoost*. Итеративен ансамбл метод [207] кој комбинира повеќе слаби класификатори за да добие конечен модел со поголема точност. Итеративно обучува копии од основниот класификатор со прилагодени тежини, така што следните класификатори се фокусираат на прилагодени набљудувања.
- *Multi-Layer Perceptron* (MLP). Засновано врз структурните аспекти на биолошките невронски мрежи, MLP е составен од збир на поврзани јазли организирани во слоеви. Секоја врска има тежина поврзана со неа, што се мести низ фазата на учење [208]. Кога се вклучени повеќе од два слоја во моделот, може да се докаже дека MLP можат да ги класифицираат податоците што не се линеарно одделени, и воопшто приближно секоја нелинеарна функција.
- *Support Vector Machine* (SVM). Модел кој учи класификација со наоѓање на најдобрата крива, во хипер-просторот создаден од карактеристиките, за поделба на инстанците според нивната класа, така што инстанците поставени во еден од двата дела на хипер-просторот се претежно од една класа [209]. Разгледуваме две алтернативи: линеарна крива, со што се дели хипер-просторот со хипер-рамнина (LinSVM); и јадро на функцијата за радијална основа, засновано на употреба на Евклидови растојанија и со тоа одделување на примери со користење на хипер-сфери (Radial basis function SVM - RBFSVM).
- *Ridge*. Модел кој учи правило за класификација со пресметување на ridge регресија помеѓу влезните карактеристики и познатите класи. Ridge регресијата е слична на стандардната линеарна, но вклучува казна за големината на коефициентите, давајќи поцврсто решение во присуство на колинеарности [210].
- *Gradient Boost* (GradBoost). Модел на класификација составен од големо множество на слаби класификатори, во овој случај дрва на одлука. Главната разлика во однос на RF е што тука дрвјата се додаваат едновремено, користејќи постапка налик на gradient-descent, а претходно додадените дрвја се замрзнуваат и остануваат непроменети [211].
- *Logistic Regression* (LR). Модел на логистичка регресија кој се користи како (генерално ефективен) класификатор. Зема линеарна равенка како влез и

користи логистичка функција за да изврши задача за бинарна класификација [212].

- *K-Nearest Neighbours* (KNN). Едноставен модел на класификација во кој класата на нова инстанца е поставена како класа најчеста меѓу нејзините k најблиски соседи. Иако е еден од најефикасните алгоритми и не бара почетна обука, тој е исто така чувствителен на локалната структура на податоците и не е погоден за сите проблеми [213].
- *Voting*. Модел базиран на комбинирање на повеќе класификатори за да се направат предвидувања врз основа на најчестиот според постапката за гласање [214]. Особено е добро прилагоден во ситуации кога не е познат најдобриот метод на класификација.
- *Bagging*. Ансамбл метод кој гради неколку случајни инстанци на основен проценувач (овде, дрва на одлука) на случајни подмножества на оригиналното множество за обука. Индивидуалните предвидувања се собираат во последното предвидување. На овој начин се намалува варијансата на основниот проценувач, а стабилноста и точноста се подобруваат [215].
- *XGBoost*. Варијанта на алгоритмот Gradient Boost оптимизиран за обработка на големи податочни множества, со додавање и на паралелна и на дистрибуирана пресметка [216].

Некои од класификационите модели имаат еден или повеќе параметри за кои е потребно првично подесување. Затоа, разгледајте наивно решение, оставајќи ги тие параметри на нивните стандардни вредности и додадовме grid процедура за пребарување, која ги одредува оптималните хиперпараметри на моделот со максимизирање на конечниот резултат на класификација.

Конечно разгледуваме два дополнителни модели, засновани на концептот на длабоко учење (Deep learning DL). DL опфаќа алгоритми за машинско учење кои постепено извлекуваат карактеристики на повисоко ниво од влезот, обично со цел да се изврши надгледувана класификација [217, 218]. Тие ја претставуваат предноста во постигнувањето висока прецизност, иако по цена да се бараат големи множества за обука, високи пресметковни цени и да бидат т.н. црни кутии, т.е. тие не го олеснуваат објаснувањето што стои зад резултатот. Првиот модел е основен пристап со целосно поврзана длабока невронска мрежа (DNN) со 3 слоја, 2 ReLU и 1 Sigmoid. Вториот модел е длабока граф конвулутивна невронска мрежа (DGCNN) [219, 220] со значајните врски извлечени од графот од процесот на избор на карактеристики како влез.

4.3 Резултати

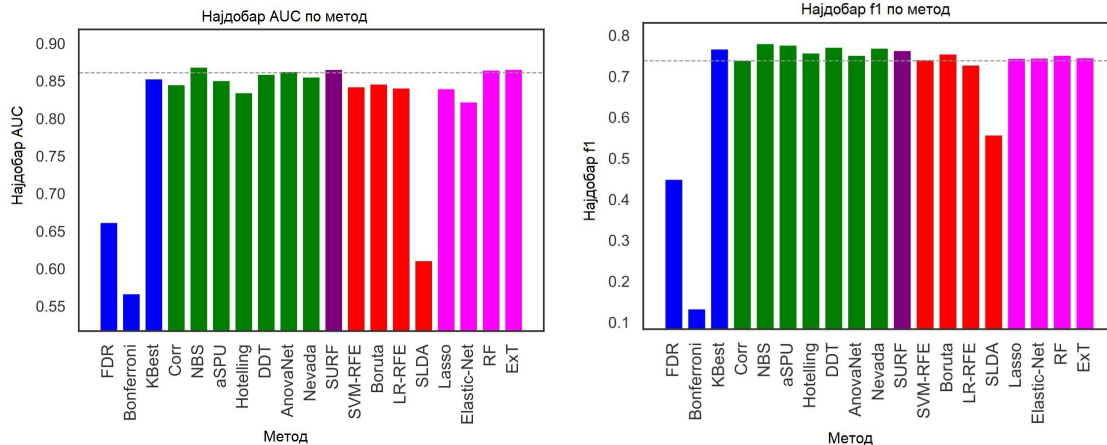
Како што првично беше спомнато, методите за избор на врската опишани во сек. 4.1 овде се оценуваат според податочното множество COBRE според три главни критериуми: нивните перформанси, во однос на количината на задржаната информација (сек. 4.3.1); нивната стабилност, т.е. колку конзистентно го даваат истото множество на врски (сек. 4.3.2); и нивната пресметковна цена (сек. 4-8). Овие резултати потоа се споредуваат во сек. 4.3.4 со оние добиени за двете преостанати податочни множества PRURIM и PSP, за да се процени нивната генерализација.

4.3.1 Споредба на перформанси

Како прва анализа, сл. 4-2 ја прикажува најдобрата класификација добиена со секој метод на избор, измерена во однос на AUC (лев панел) и F1-мерка (десен панел). За секој модел на избор, максимумот е пресметан во сите модели на класификација. Со други зборови, она што е прикажано на сл. 4-2 е колку е добра класификацијата, под услов да се знае и да се користи најдобриот алгоритам за класификација во секој случај. Крајниот резултат на класификација потоа се користи како мерка за количината на информации што ги задржува секој метод за избор на врска.

Резултатите генерално се прилично хомогени. Со исклучок на три методи (SLDA, FDR и Bonferroni), преостанатите методи се некаде помеѓу NBS (AUC од 0,869) и Elastic-Net (AUC 0,822). При споредување на методите со користење на мерката F1, се појавува слична слика, со методите кои се поставуваат помеѓу 0,780 (NBS) и 0,728 (LR-RFE). Ова сугерира дека изборот на специфичен метод на избор има мало влијание во крајната класификација; и, соодветно, дека методите се ефикасни во задржување на релевантни информации за проучената состојба. Ова секако не важи во случајот на SLDA, FDR и Bonferroni, кои очигледно работат полошо од другите; и, во случајот на Bonferroni, дури пристапува кон бариерата на она што се очекува во случајна класификација.

Важно е да се нагласи дека резултатите за класификација овде пријавени се просечни во однос на 10-fold вкрстената валидација; кога ќе се земе предвид стандардната девијација на овие, соодветно 0,094 за NBS и 0,085 за Elastic-Net (и за AUC), станува јасно дека разликата помеѓу најдобриот и најлошиот алгоритам е занемарлива. За понатамошно поставување на овие бројки во контекст, поставени се хоризонталните испрекинати линии на сл. 4-2 добиени со RF класификација. Овие резултати, соодветно 0,861 за AUC и 0,738 за F1-мерката, покажуваат дека

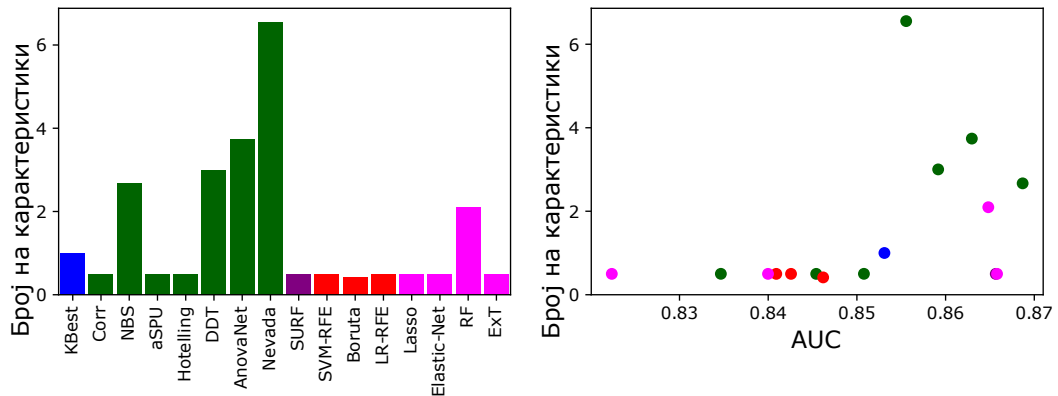


Слика 4-2: Најдобар резултат за класификација, во смисла на AUC (лев панел) и F1 (десен панел), добиен со методите за избор на врски, овде разгледани. Боите го означуваат семејството на секој метод, користејќи го истиот код како на Сл. 4-1. Хоризонталната испрекината линија претставува резултат на класификација добиен со RF модел на целата матрица за конективност, т.е. без никаков избор на врска.

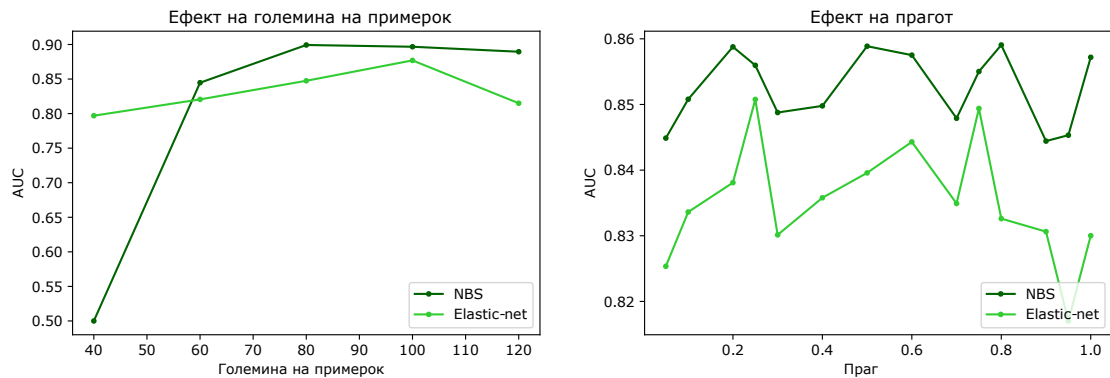
отфрлањето на неинформативни врски не само што ги задржува сите релевантни информации, туку исто така може да биде корисно за проблем со класификација - ефект кој е добро познат во машинското учење.

Следно, анализираме како се однесуваат методите во однос на бројот на врски што ги препорачуваат за анализата. Сл. 4-3, лев панел, го отсликува таквиот број; и десниот панел претставува запис на овој број како функција од резултатот на класификација (AUC). Да се забележи дека три методи се со недоволни перформанси, според сл. 4-2, па заради јасност се изоставени од сл. 4-3. Може да се оцени дека, како што се очекуваше, методите што избираат неколку карактеристики се генерално оние кои работат положно. Додека општото е генерално точно, вреди да се напомене дека методот што дава најголем број на карактеристики (т.е. Nevada) не е најдобриот (NBS, како што беше и претходно).

Сл. 4-4 ја прикажува споредбата на однесувањето на најдобрите и најлошите методи, т.е. соодветно на NBS и Elastic-Net, кога синтетички се менуваат две важни карактеристики на податочното множество: бројот на субјектите во неа (лев панел); и начинот на предобработка на мрежите, конкретно кога одреден дел од најсилните врски се задржуваат и сите други се бришат (процес познат како пропорционален праг [87], десен панел). И во двата случаи, задачата за класификација е извршена со моделот за машинско учење да даде најдобар резултат, т.е. моделот со резултат на сл. 4-2. Промената на големината на примерокот има



Слика 4-3: (Лево) Број на врски задржани со секој метод (со исклучок на трите со најмалку перформанси). (Десно) Број на задржани врски како функција на резултатот (AUC) постигнат со секој метод. Бојата на секој метод одговара на онаа на сл. 4-1.



Слика 4-4: Еволуција на резултатот на класификација, во смисла на AUC, како функција на големината на примерокот (број на субјекти во податочното множество, лев панел) и на пропорционалниот праг што се применува на мрежата (десен панел). Црните и зелените линии соодветствуваат на методите за избор на врски на NBS и Elastic-net.

мал ефект и во двата случаи, иако се чини дека NBS не работи со множества на податоци што содржат помалку од 60 субјекти. Од друга страна, почетниот чекор за бришење на врската има занемарливо влијание, со разлики во AUC $< 0,02$ - т.е. само двапати од стандардната девијација забележана во процесот на вкрстена валидација.

4.3.2 Intra- and inter-стабилност на метод

Кога се занимаваме со различни методи за избор на карактеристики, два важни аспекти се intra- и inter- стабилноста на методот. Подетално, некој сака да процени дали дадениот метод секогаш дава ист збир на карактеристики за фиксен проблем (овде наречено *intra-стабилност на методот*); и дали два различни методи даваат групи на карактеристики што се компатибилни или се преклопуваат (*inter-стабилност на методот*). Забележете дека изборот на нестабилна карактеристика може да доведе до инфериорна класификација; но, уште поважно, може да го поткопа секој обид за разбирање на механизмите вклучени во некоја состојба или патологија.

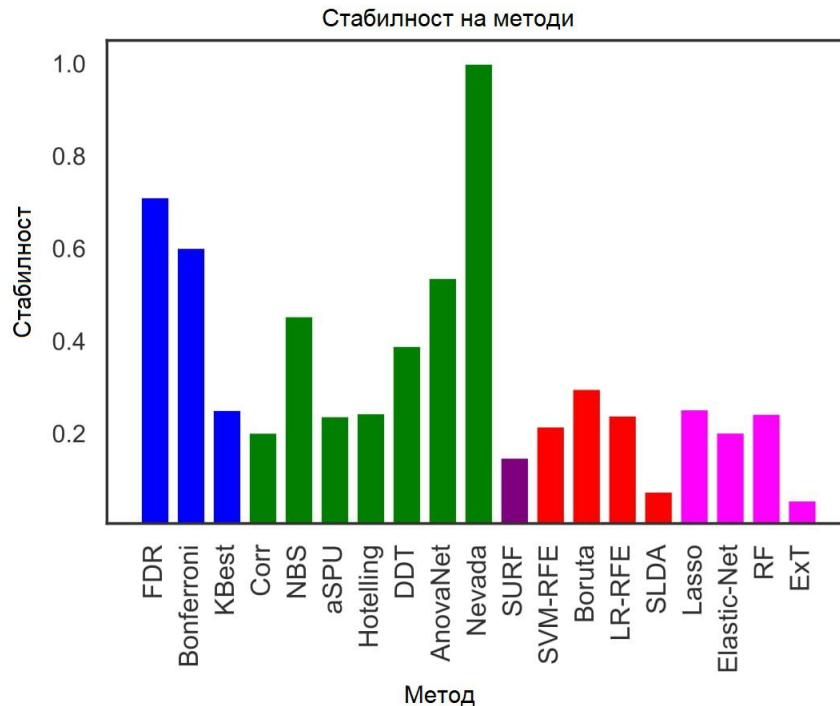
Со цел да се оценат двете стабилности, ние овде користиме метод на сличност предложен во [221]. Сличноста на две избрани множества s и s' се пресметува како еден минус растојанието на Танимото, кое го мери преклопувањето на две множества со произволна кардиналност:

$$D(s, s') = 1 - \frac{|s| + |s'| - 2|s \cap s'|}{|s| + |s'| - |s \cap s'|}. \quad (4.3)$$

Оваа метрика има вредност во опсег $[0, 1]$, каде што 0 значи дека нема преклопување или сличност помеѓу две рангирања на карактеристиките, а 1 значи дека рангирањата на карактеристиките се идентични.

Сл. 4-5 ја прикажува интра-стабилноста на секој метод на избор. За секој од нив се изградени 100 подмножества што ги содржат сите карактеристики, но само 100 субјекти, избрани од оригиналното множество со користење на случајно ресемплирање со bootstrap; прикажаната вредност одговара на просекот во однос на сите вредности во парови од D . Nevada се издвојува како метод со поголема интра-стабилност; ова не е изненадувачки, бидејќи ова беше и методот што дава поголем број на елементи - видете сл. 4-3. Потоа, двата најстабилни методи се FDR и Bonferroni; сепак, ова се две од слабите перформанси идентификувани на сл. 4-2, што сугерира дека тие постојано избираат погрешни интеракции. Конечно, двата следни методи се AnovaNet и NBS, што одговараат на двата најдобри методи.

Сл. 4-6 понатаму ја прикажува интер-стабилноста, т.е. колку се преклопуваат множествата задржани врски дадени со парови на методи, повторно измерени со метриката D . На сличен начин на сл. 4-5, може да се процени дека вредностите се генерално ниски; на пример, двата најдобри методи (NBS и AnovaNet) имаат $D \approx 0,54$, т.е. се согласуваат за приближно 50% од врските. Ваквата хетерогеност понатаму е прикажана на 4-7 каде што е прикажано дека заедничките

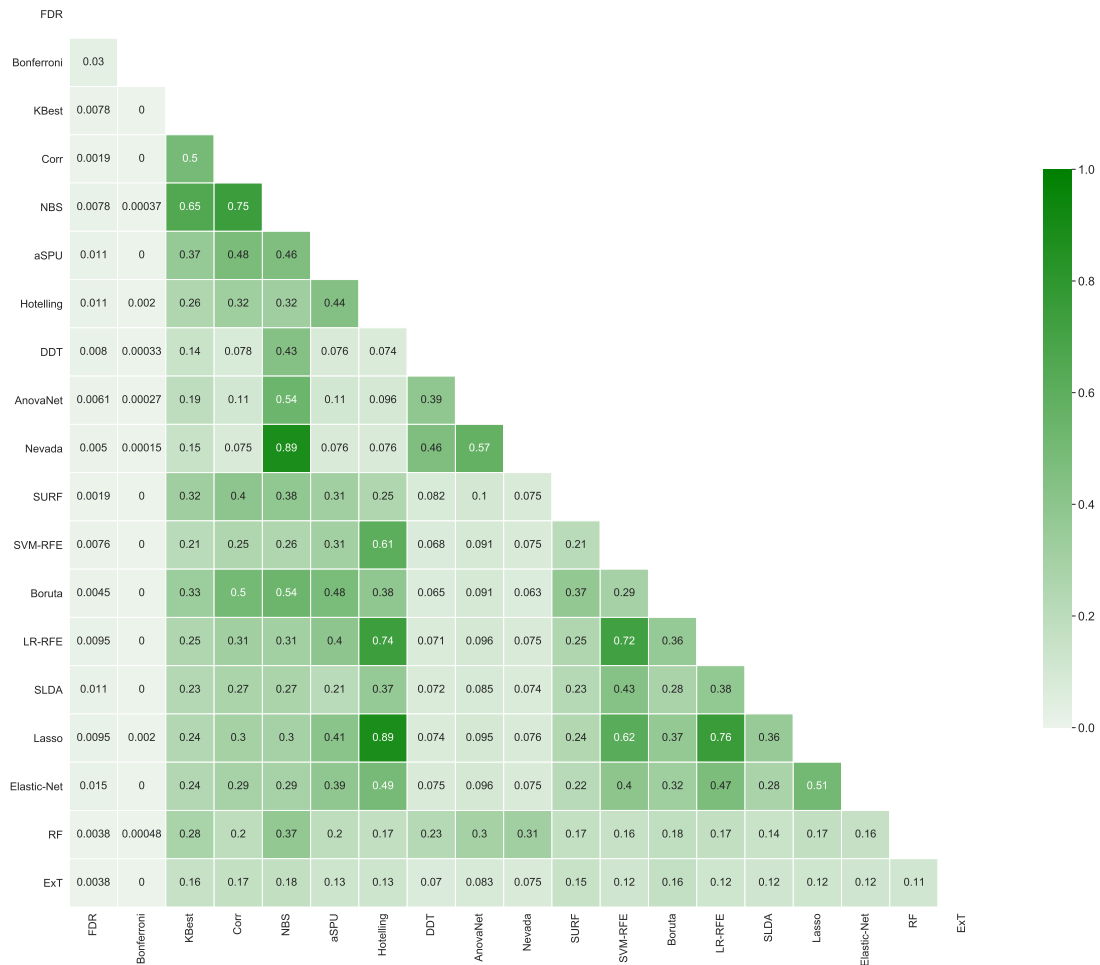


Слика 4-5: Интра-стабилност D за секој метод, што претставува колку конзистентно секој метод дава ист избор на врска преку случајни под-примери на оригиналното податочно множество.

врски се дистрибуираат меѓу сите области на мозокот, со единствен исклучок во таламусот. Овој последен резултат е во согласност со признаената важност на таламичното коло во шизофренијата [222, 223, 224]. Од друга страна, бројот на врски за секој метод варира помеѓу неколку десетици (за SURF) до неколку стотици (NBS). Импликациите од овие резултати понатаму ќе се дискутираат во заклучоците.

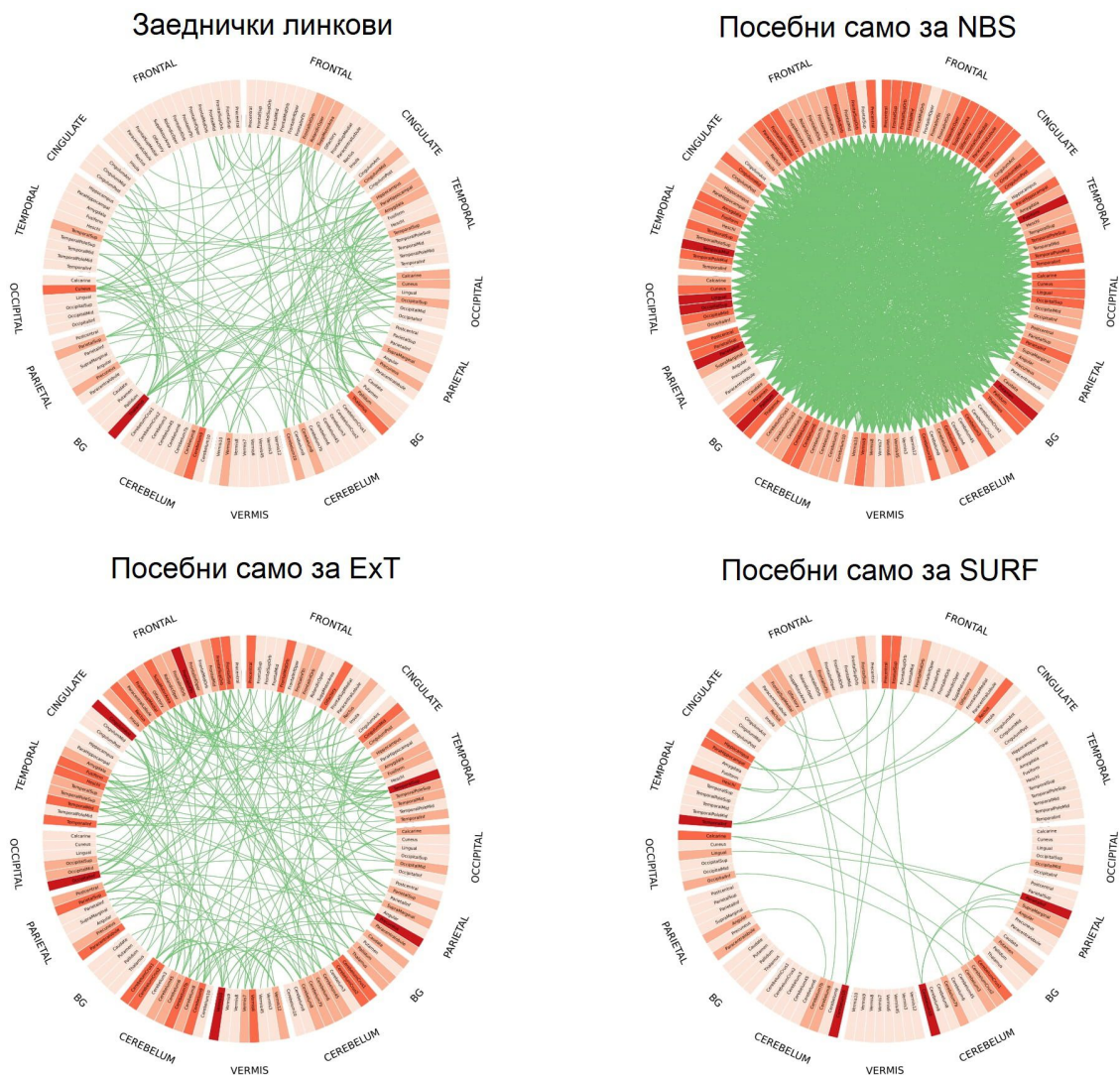
4.3.3 Перформанси

Како последно прашање, од големо значење во реалниот свет каде што целта може да биде да се анализираат големи групи на податоци, овде ја споредуваме пресметковната цена на методите со најдобри перформанси. Таквата цена е проценета како време потребно за извлекување на најдоброто множество врски, со исклучок на времето потребно за последователна класификација и валидација, со употреба на процесор со четири јадра Intel (®) i7-8565U на 1,8GHz, 8 GB RAM меморија. Резултатите се презентирани на сл. 4-8 и независно (лев панел) и како функција од резултатот на класификација добиен со секој метод (десен панел).



Слика 4-6: Интер-стабилност на секој пар методи, т.е. дел од заедничките врски што ги даваат двата, измерени според равенка 4.3.

Помеѓу методите се издвојуваат големи разлики, опфаќајќи скоро пет редови на големина - забележете дека резултатите се прикажани со употреба на логаритамска скала. На некои методи, како KBest, им треба дел од секундата за да ги дадат своите резултати, додека други, на пр. Nevada, десетици минути. Интересно е да се напомене дека статистичките методи обично не се најбрзите; дури и ако тие можат да бидат концептуално поедноставни, тие обично бараат создавање на случајни ансамбли за да се добијат p -вредности, што резултира со значителна пресметковна цена. Од друга страна, спротивното важи за вградените методи на машинско учење. Ова може да се должи на две комплементарни причини: прво, дека тие не бараат дополнителни напори за пресметување надвор од обуката на самиот модел; и второ, дека нивната имплементација, вклучена во стандардни и добро имплементирани библиотеки, како што е пакетот scikit-learn во Python, е

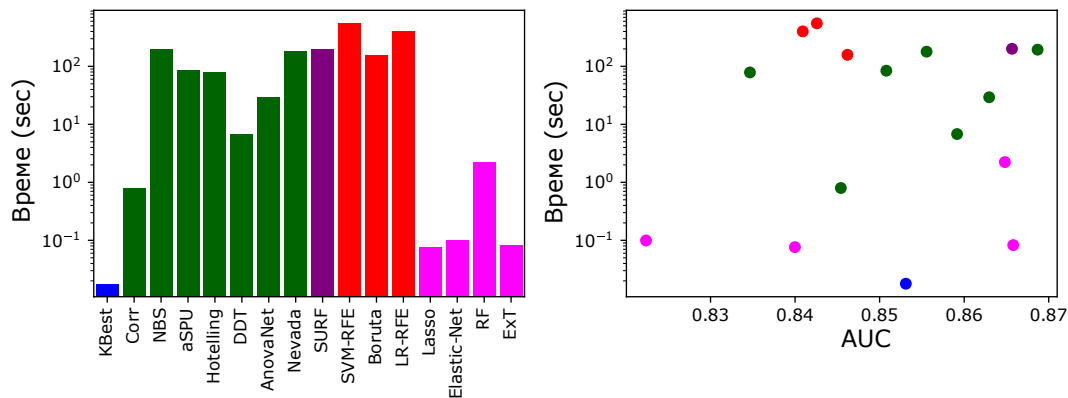


Слика 4-7: Графички приказ на врските кои се заеднички и ексклузивни за трите најдобри методи според AUC. Поточно, горниот лев панел ја дава кружната репрезентација на врските што се заеднички за NBS, ExT и SURF; додека другите три, множеството врски кои се исклучиво избрани од соодветно NBS, ExT и SURF. Јазлите (т.е. региони од интерес, видете сек. 4.2.1, се означени според соодветното име во AAL атласот; боите го означуваат нивниот степен, од посветли (без или неколку врски) до темни нијанси (максимален степен). Четири панели се изработени со помош на Circos софтверот [225].

особено нагодена за ефикасност.

4.3.4 Генерализирање на резултатите

Еден последен важен аспект што треба да се разгледа е колку се општи резултатите овде прикажани; со други зборови, дали некој може да биде сигурен дека,



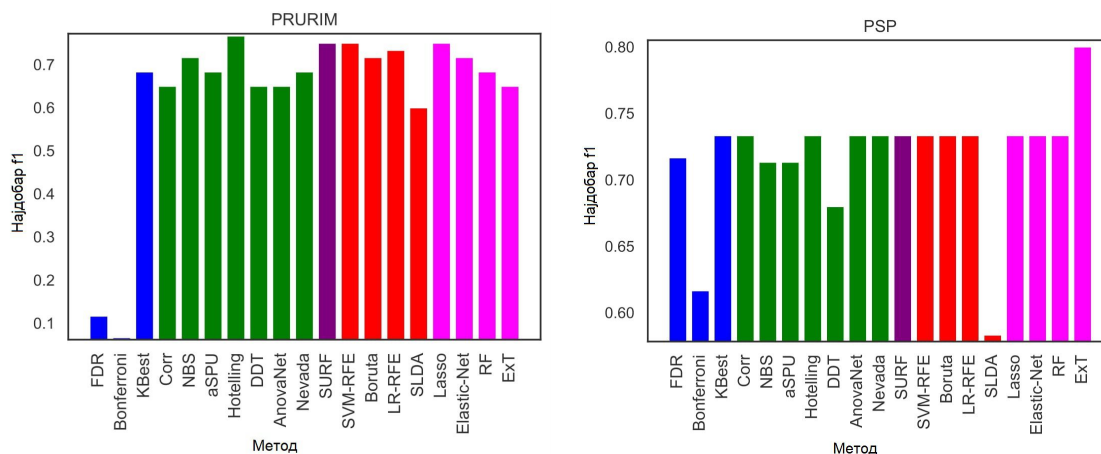
Слика 4-8: Пресметковна анализа. (Лево) Време на пресметка за секој метод за избор на врска (со исклучок на трите со недоволни перформанси), измерено во секунди. (Десно) Времето на пресметување како функција на перформансите (во смисла на AUC) на секој метод. Код на боја според сл. 4-1.

на пример, најдобриот модел за избор на врски откриен на сл. 4-2 исто така ќе биде најдобар со изведба кога ќе се примени на друго податочно множество? Забележете дека генерализацијата на резултатите од невронауката, а особено на оние добиени преку мрежни претстави, е тековно поле на истражување; и дека обично се препознава недостаток на генерализација, главно заради многу произволни чекори вклучени во предобработка на податоците што може да влијае на резултатите на непредвидлив начин [226, 227, 228].

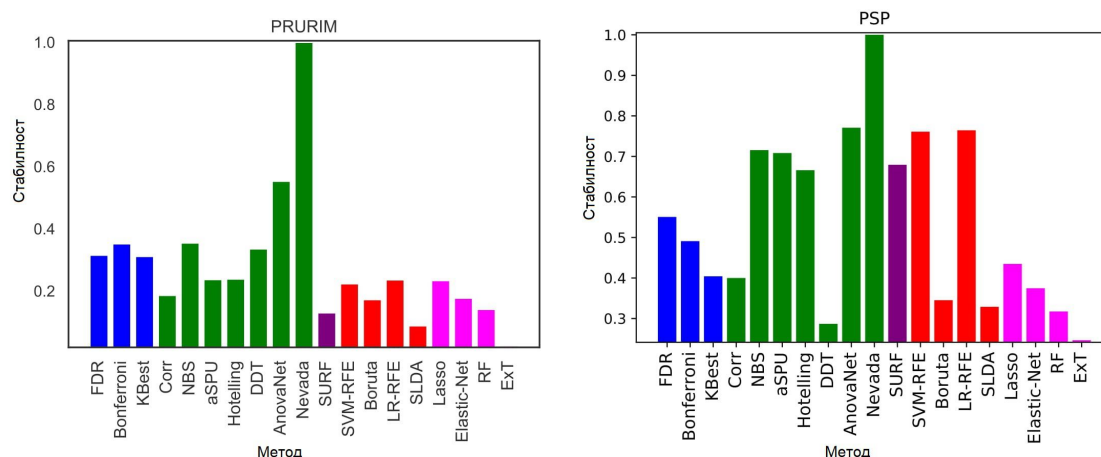
Ние овде извршуваме основен тест со извршување на истата анализа на двете преостанати податочни множества на мозочни мрежи, т.е. PRURIM (видете сек. 4.2.1) и PSP (сек. 4.2.1)). Конкретно, сл. 4-9 ја дава цената на класификацијата, со F1 мерка за двете податочни множества и сл.4-10 соодветната интра-стабилност на методот. Резултатите се генерално компатибилни со оние прикажани на сл. 4-2 и 4-5. Повеќето методи постигнуваат споредливи оценки за класификација, додека FDR, Bonferroni и SLDA имаат слаба изведба. Исто така, може да се оцени дека резултатите се порамнети во случај на податочното множество PSP, т.е. F1 дава многу слични вредности во сите методи; ова се должи на ограничената големина на оваа база на податоци, која опфаќа само 64 мрежи од 27 јазли.

4.4 Дискусија

При анализа на мрежите што ја претставуваат динамиката на мозокот во различни состојби или патологии, употребата на некаков вид *метод на избор* е



Слика 4-9: Најдобри оценки за класификација, во однос на метриката F1, за податочните множества PRURIM (лев панел) и PSP (десен панел).



Слика 4-10: Интра-стабилност D од секој метод, за податочните множества PRURIM (лев панел) и PSP (десен панел). Погледнете ја равенка. 4.3 за дефиниција на D .

привлечен концепт. Користењето на целата мрежа може да биде пресметковно и концептуално предизвикувачки, особено во случај на fMRI, кои во својата сурова форма може да вклучуваат $> 10^4$ јазли и $> 10^8$ врски. Обработката на тие големи мрежи не бара само значителни пресметковни ресурси; но исто така го отежнува прецизирањето кои се релевантните елементи кои карактеризираат патологија или состојба. На начин сличен на изборот на карактеристики при машинско учење, изборот и задржувањето на само подмножество од тие јазли и врски дава полесно толкување на сликата, додека не се отфрлаат релевантни

информации. Понатаму помага во зголемување на статистичкото значење на резултатите и може да ја подобри последователната обработка, како на пр. задачи за класификација. Во овој придонес споредивме голем множество на различни методи за избор на врски, извлечени од статистиката и машинското учење и ги споредивме нивните перформанси во однос на перформансната цена (т.е. колку информации се задржуваат на крајот, измерени преку задача за класификација), стабилноста и пресметковна цена. Рангирањето на најдобрите 5 методи според секој од овие критериуми е дадено во табела 4.1.

Табела 4.1: Рангирање на најдобрите 5 методи, според нивната AUC, стабилност D и пресметковна цена за COBRE податочното множество.

Нај. AUC		Нај. стабилност (D)		Нај. пресм. цена (sec.)	
NBS	(0.869)	Nevada	(1.000)	KBest	(0.018)
ExT	(0.866)	FDR	(0.711)	Lasso	(0.076)
SURF	(0.866)	Bonferroni	(0.602)	ExT	(0.083)
RF	(0.865)	AnovaNet	(0.536)	Elastic-Net	(0.099)
DDT	(0.859)	NBS	(0.426)	Corr	(0.798)

Најважниот и донекаде изненадувачки резултат е тоа што, освен неколку методи (т.е. SLDA, FDR и Bonferroni), сите други даваат сличен резултат на класификација; со други зборови, и покрај нивните внатрешни разлики и основните хипотези, информациите што ги задржуваат се скоро исти - видете сл. 4-2 и првата колона на табела 4.1. Во рамките на оваа група, разликата помеѓу оние со најдобри (NBS) и најлоши (Elastic-Net) перформанси е всушност со иста големина на стандардната девијација на резултатите добиени во различни повторувања на нивната вкрстена валидација; разликата е слична на онаа што се очекува при обработка на различни податочни множества, а со тоа и не е значајна.

Второ, вреди да се напомене дека резултатот на класификација добиен со користење на филтрирани множества врски е еднаков, а во некои случаи дури и поголем, од оној што одговара на употребата на целосното податочно множество. Затоа, бришењето на неинформативни врски не само што го поедноставува проблемот, туку дури и го зголемува нашиот капацитет за извлекување заклучоци од него. Иако ова е добро позната идеја во машинското учење, нејзината примена во невронауката досега е ограничена.

Разгледувајќи ја интра-стабилноста на секој метод на избор, т.е. неговиот капацитет за постојано давање на истото множество врски, резултатите се генерално ниски - видете 4-5. Nevada се издвојува како исклучок; сепак, ова се должи

на големиот број врски што ги задржува и затоа ова не е значајно. Претпоставуваме дека ова може да се должи на две различни причини: или методите не се стабилни, т.е. се дизајнирани на начин што дава нестабилни резултати кога се применуваат на мозочните мрежи; или самата мрежа е нестабилна. Оваа втора можност имплицира дека варијабилноста помеѓу луѓето е голема, таква што состојба може да се манифестира во различни делови на мозочните мрежи кај различни луѓе. Ова всушност се усогласува со претходните истражувања за општа способност на мозочната мрежа [229, 230, 231], но вреди да се напомене дека податоците и анализите овде презентирани не можат да дадат убедлив одговор на овој проблем.

Стабилноста помеѓу методите за избор на врски, т.е. интер-стабилноста, е исто така ниска - видете сл. 4-6. Со други зборови, иако различни методи можат да ги задржат сите релевантни информации за задачата што е предмет, таквите информации не се кодирани од истите врски. Ова сугерира интересна слика, во која релевантните информации се кодираат на редуван начин во различни делови на мрежата на мозокот, делови што се избираат според методите според различни критериуми. Со други зборови, патологијата може да влијае на две групи врски s и s' , со мало преклопување помеѓу нив; изборот на кој било од нив е еквивалентен и не се добива дополнителна вредност со разгледување на двете групи заедно. Иако ова наликува на концептот на мрежи за стандарден режим на мозокот [232, 233], според нашите сознанија, присуството на комплементарни под-мрежи поврзани со патологија претходно не е дискутирано во литературата.

Како овие резултати придонесуваат за одговарање на главното прашање зафатено со овој придонес, т.е. кој е најдобриот метод за избор на врски во мозочните мрежи? Во принцип, заклучокот е дека, со неколку исклучоци, кој било метод е добар, и навистина чекорот за избор дури е неопходен за да се подобри нашето разбирање за мозокот. Во резултатите во врска со стабилноста и пресметковните цени, може да се дефинираат два случаи. Кога приоритет е да се добие најдобра оценка за класификација, NBS е најдоброто решение; тој дава маргинални резултати подобри од останатите методи и добра стабилност, иако по цена на висока пресметковна комплексност. Од друга страна, во случај на големи податочни множества или проблеми во кои е важно времето потребно за да се добие одговор, може да се употребат ExT или KBest: нивната пресметковна цена е помала за три редови и тие даваат помал излез, цената е малку со пониска прецизност во следната класификација, но има многу намалена стабилност во избраните врски.

Како последна забелешка, важно е да се дискутираат некои ограничувања на

оваа анализа. Како прво, се заснова на податоци од три податочни множества, вклучувајќи информации за три патологии (и соодветни контролни субјекти). Генерализацијата на заклучоците овде презентирани на други групи на податоци и патологии не е загарантирана, иако се преземени релевантни активности кон тоа - на пр. употреба на вкрстено валидирање и ресемплирање на субјектите, како и употреба на различни големини и прагови за примероците. Од друга страна, COBRE е добро познато податочно множество, кое се користи како референца во многу студии за невронаука, на пр. [234, 235, 236, 237], па презентираниите резултати се во секој случај релевантни за невронаучната заедница.

Второ, реконструкцијата на функционалните мозочни мрежи е процес кој опфаќа многу чекори, во кои треба да се преземат неколку одлуки (на пример, употреба на различни параметри или методи), и обично склони кон голем степен на субјективност [238]. Ова резултира со различни заклучоци извлечени од исто податочно множество [226, 227, 228]; и, следствено, можноста резултатите овде презентирани да зависат од изборот направен во процесот на реконструкција на мрежата. Истражувачот што прави каква било анализа на податоците за невро-слика, секогаш треба да практикува претпазливост, исто така, при изборот на методот за избор, и на пример, споредувајќи неколку од нив.

Трето, постои голема хетерогеност во кодовите што се користат за проценка на секој метод: некои од нив се вклучени во признати и добро одржувани библиотеки, други се објавени од авторите на методите, а други се развиени специјално за овој придонес. Како таква, споредбата на пресметковните цена (сл. 4-8) мора да се земе со соодветна претпазливост, бидејќи некои методи можат да бидат подложни на значителни оптимизации. Сепак, презентираниите резултати се репрезентативни на она со што може да очекува да се соочи истражувач со просечни вештини за кодирање.

Конечно, овој придонес во дисертацијата се фокусира на две широки семејства на методи за избор на врски, т.е. статистички и машинско учење. Заинтереси-раниот читател мора да биде свесен дека може да се најдат и други алтернативи. За илустрација, предложени се некои методи засновани на длабоко учење, видете на пример [239, 240, 241]; додека DL методите обично даваат одлични резултати, тие се исто така *црни кутии*, т.е. тие не олеснуваат проверка или објаснување на нивната внатрешна логика. Второ, машинското учење е зрело, активно поле на истражување, со многу апликации за проблеми со невронауката [242, 243]; па не е изненадувачки што се објавени многу студии, обете насочени кон специфични патологии [244, 245, 246, 247, 248] и фокусирање на алтернативни методи за избор на врски [249, 250, 251]. Читателот треба да биде свесен за

литературата поврзана со анализата на мозочните мрежи [18, 149]; Додека поголемиот дел од времето целата мрежа се разгледува и се анализира со користење на тополошки метрики, некои студии исто така идентификуваат специфични подмрежи или подмножества на врски користејќи принципи на мрежната наука [252, 253, 246, 254, 255, 256, 257].

Глава 5

Веб базиран систем за евалуација на техники базирани на мрежни вгнездувања за обработка на мозочни мрежи

Развојот на ефективни пристапи и алгоритми за разбирање на графовите е од голем интерес. Општо земено, овие алгоритми спаѓаат во една од следниве четири категории: класификација на јазли, предвидување на рабови, кластерирање и визуелизација на графови [258]. Методите за класификација на јазли се поделени во два различни пристапа - методи кои користат случајни прошетки низ графови за пропација на лабелите на класите [259] и методи кои извлекуваат карактеристики на јазли и применуваат класификатори на овие карактеристики [260]. Методите за предвидување на врски се состојат од методи базирани на јазли, методи со максимална веројатност [261] и веројатносни модели [262]. Методите на кластерирање вклучуваат методи базирани на модел [263] и модели кои се обидуваат да го максимизираат растојанието на различни групи [264].

Методите за анализа на графови најчесто користат матрица за соседство или изведен векторски простор [258]. Јазлите од графот можат да бидат претставени како вектори од одреден векторски простор, и ова се нарекува векторска претстава на графот (векторизација) или вгнездување на графот.

Векторските простори се поприфатливи во податочното рударење отколку графовите. Графовите содржат врски и јазли, а тие мрежни врски можат да користат само специфично подмножество на математички, статистички и методи на машинско учење. Векторските простори имаат побогат пакет алатки во

тие домени, а исто така, векторските операции често се поедноставни и побрзи од еквивалентните операции на графови. На многу графови од реалниот свет, по неколку чекори (пример во наоѓање на најблиски соседи), има малку значајни информации (на пр. препораки од пријатели на пријатели на пријатели). Меѓутоа, во векторските простори, може да се користат метрички растојанија за да се добијат квантитативни резултати (на пример, евклидово растојание или косинусна сличност). Метриките за квантитативно растојание во значаен векторски простор може да го направи проблемот со наоѓање најблиски соседи да биде многу појасен.

Добивањето векторска репрезентација од даден граф е тешка задача што наметнува повеќе предизвици. Еден од главните проблеми е да се избере својството на графот што векторската репрезентација треба да ја зачува. При трансформација на графот, структурата и врските на јазлите во графот треба да се задржат. Постојат многу метрики и својства на графови што можат да бидат избрани за зачувување, па затоа изборот обично зависи од примената на векторската претстава, како и од посакуваната изведба. Добиените вектори се користат како влезови во моделите за анализа на мрежи и соодветните параметри се изучуваат со помош на обучувачко множество. На овој начин се зголемува ефикасноста и бројот на модели што можат да се користат во анализата на графот.

Во денешно време, може да се сретнеме со графови кои достигнуваат до милиони јазли, така што, векторските претстави треба да бидат доволно прилагодливи за да работат со големи графови. Ова е исклучително тешко ако истовремено е потребно да се задржат некои глобални својства на графот. Во случајот со мозочните мрежи, графовите и не се така големи, но треба да се внимава при работата со нив, за да се зачуваат карактеристиките од мрежата кои се значајни. Димензијата на векторските репрезентации е исто така важен избор што треба да се направи при креирање модели за анализа. Зголемувањето на бројот на димензии може да доведе до поголема точност, но времето на извршување станува побавно, додека мемориските барања се зголемуваат.

Многу од моделите за граф трансформација имаат свои имплементации во различни програмски јазици кои се со отворен извор. Така, потребното познавање на програмскиот јазик во кој е имплементиран соодветниот модел може да се најде како пречка за некои од корисниците. Овие модели често применуваат неколку различни параметри кои треба да се прилагодат на графот од интерес, како и планираната употреба на векторската претстава. Тестирање на различни вредности на параметрите бара рачно поставување на секој параметар поединечно или пишување скрипти на програмата што би го направиле процесот

автоматски. Дополнително, споредбата на перформансите на различните модели на трансформација применети на еден граф честопати е проблематична поради нееднакви типови на влез, како и поради различни типови на излези. Сите горенаведени причини, го отежнуваат процесот на избор на најсоодветен модел на трансформација за даден граф и апликацијата од интерес.

Во овој дел, ние презентираме веб-базиран систем со едноставен кориснички интерфејс, користен за споредба на перформансите на различни модели за трансформација на даден граф од интерес. Системот овозможува повеќекратно споредување на параметрите на модели при задача на реконструкција на графовите. Прво, корисниците можат да ги споредат перформансите на моделите за различни димензии на векторските претстави. Покрај тоа, системот исто така овозможува споредба на перформансите на одделни алгоритми за различни вредности на нивните параметри.

Во продолжение во сек. 5.1 ќе ги објасниме искористените методи за вгнездување на графови кои се применуваат во изработениот систем, заедно со употребените метрики за евалуација и методите за проценка (реконструкција на граф). Потоа ќе презентираме некои детали за имплементација на веб системот во сек. 5.2, и на крај ќе дадеме пример на употреба на овој систем со мозочни мрежи.

Дел од оваа глава од дисертацијата е објавена во трудот Pinka Ivanoska, Martin Milenkoski, Slobodan Kalajdziski, Kire Trivodaliev, “Web tool for graph embeddings representation techniques evaluation”, MIPRO 2019: 983-988.

5.1 Методи

5.1.1 Методи за граф трансформација во векторски простор

Локално линеарно вгнездување - **Locally Linear Embedding(LLE)**

Локалното линеарно вгнездување - **Locally Linear Embedding(LLE)** е техника што претпоставува дека секој јазол на графот е линеарна комбинација на соседите [265]. LLE претпоставува дека тежината на врска помеѓу два јазли W_{ij} ја претставува тежината на јазолот j во векторската претстава на јазолот i . Математички, векторската претстава на јазолот i е дефинирана со: $X_i \approx \sum_j W_{ij} X_j$ за $\forall i \in V$.

Векторска претстава на јазлите се добива на овој начин, но димензијата на оваа претстава е еднаква на бројот на јазлите во графот ($D = V$). LLE претпоста-

вува дека јазлите во овој векторски простор лежат на мазна нелинеарен manifold со многу помала димензија $d \ll D$. Во овој случај, постои линеарна трансформација која се состои од транслации, ротации и скалирање. Ги трансформира високодимензионалните координати на секое соседство на јазли во глобалните внатрешни координати на тој manifold. Бидејќи тежините на графот можат да ја карактеризираат локалната геометрија во високодимензионален простор, тие можат да ја карактеризираат геометријата на локалните делови на manifold-от. Особено, истите тежини W_{ij} може да се користат за претставување на јазлите во d -димензионалниот векторски простор. Врз основа на оваа идеја, LLE конструира векторска претстава што ги зачувува соседствата помеѓу јазлите. Секој високо-димензионален вектор X_i е мапиран во ниско-димензионален вектор Y_i . Ова се прави со избор на d -димензионални вектори Y_i кои ја минимизираат функцијата $\varphi(Y) = \sum_i |Y_i - \sum_j W_{ij} X_j|^2$.

Laplacian Eigenmaps (LE)

Laplacian Eigenmaps (LE) е метод за трансформација на графови, кој исто така се обидува да ги зачува соседствата помеѓу јазлите [266]. На ист начин како и LLE, овој метод претпоставува дека точките лежат на нелинеарен manifold со помала димензионалност од онаа на оригиналниот векторски простор. LE се обидува да ги задржи векторските трансформации на два јазли, слични едни на други, ако тежината на врската помеѓу нив W_{ij} е доволно голема.

Граф факторизација - Graph Factorization (GF)

Граф факторизација - Graph Factorization (GF) е првиот алгоритам што успева да постигне време на извршување што е линеарно зависно од бројот на врските на графот, $O(|E|)$. Ова е значително забрзано бидејќи повеќето графови од реалниот живот имаат ретки матрици на соседство; затоа, бројот на врските на графот е со димензија $(|V|)$. На овој начин, времето на извршување се подобрува од квадратно во линеарно во однос на бројот на јазли на графот. Главната идеја во GF е да се добие матрица на векторска претстава на графот Y , така што $Y^T Y$ е што е можно поблиску до неговата матрица на соседство W [267].

High-Order Proximity preserved Embedding (HOPE)

High-Order Proximity preserved Embedding (HOPE) [268] е уште еден алгоритам заснован на факторизација. Ја одржува блискоста меѓу јазлите во графов од висок ред со минимизирање на матрицата на сличност. При овој процес може да

се користат неколку метрики за сличност, како што се Katz индекс, коренов Page Rank, Заеднички соседи, или Adamic-Adar score. HOPE користи декомпозиција на единечна вредност (Singular Value Decomposition SVD) за ефикасно да добие векторската репрезентација на графот.

Структурно длабоко мрежно вгнездување - Structural Deep Network Embedding (SDNE)

Structural Deep Network Embedding SDNE [269] е метод на вгнездување на граф кој се обидува да ја зачува нелинеарната структура на мрежата. Дополнително, SDNE се обидува да ја зачува локалната структура на граф со зачувување на соседностите од прв ред, но исто така и глобалната структура на графот со зачувување на соседностите од втор ред. Ова го решава проблемот со ретки графови каде сите врски не носат доволно информации за глобалната структура. Моделот SDNE се состои од длабока конволуциона мрежа и користи полунадгледуван пристап. Системот се состои од две компоненти, дел без надзор и дел со надзор. За делот без надзор, се користи изменета верзија на длабок авто-кодер.

Node2vec

Node2vec [270] е еден од алгоритмите за вгнездување на граф што користи случајни прошетки. Овие алгоритми се корисни кога графот е преголем за да може да се анализира во целост или кога е можно да се набљудува само дел од графот. Node2vec одржува близина на јазли од висок ред со тоа што ја максимизира веројатноста за последователни јазли со случајни прошетки со фиксна должина. Користи biased-случајни прошетки кои обезбедуваат рамнотежа помеѓу пребарувањето по широчина (BFS) и пребарување по длабочина (DFS) и на тој начин произведува векторски репрезентации со висок квалитет.

5.1.2 Метод за евалуација - реконструкција на граф

Бидејќи основната задача на методите за вгнездување на графови е одржување на графовидната структура во векторски простор, основниот метод за проценка на овие методи е да се процени нивната способност да ги реконструираат постојните врски во одреден граф. Вистинските врски на графовите се претходно познати, така што тие лесно може да се споредат со реконструираниите добиени со векторските репрезентации. Секој метод на вгнездување користи различен начин за добивање на векторски репрезентации, така што методот за реконструкција

на тежините на врските помеѓу два јазли ќе зависи од применетиот метод на вгнездување.

5.2 Спецификации на системот

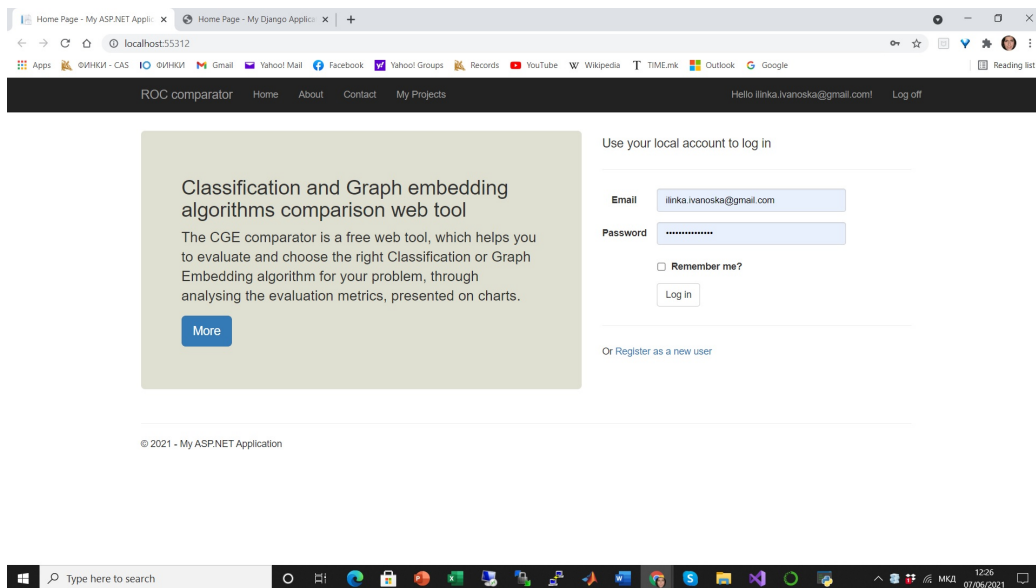
За развојот на овој веб-базиран систем, ги искористивме имплементациите на техниките за трансформација на граф вгнездувања од библиотеката GEM [258]. Некои од техниките работат добро со големи графови, додека други немаат можност да завршат во реално време со графови поголеми од неколку стотици јазли. Со оглед на тоа што тука работиме со мозочни мрежи, кои не се големи графови и податочните множества се релативно мали, различни имплементации беа прилагодени со цел да работат на графови со иста големина. На пример, SDNE архитектурата беше намалена со цел да има можност да генерализира помали податочни множества. Користејќи податочно множество на мозочни мрежи беа направени експерименти со цел да се утврдат оптималните вредности на некои од хиперпараметрите во алгоритмите имплементирани во системот.

5.2.1 Функционалности на системот

Системот обезбедува неколку основни функционалности на корисникот. Почетниот поглед на апликацијата е прикажан на сл. 5-1. На почеток корисникот треба да се регистрира и логира во системот. Потоа тој креира проект на кој може да додаде колаборатори, а потоа во проектот може да врши мрежни вгнездувања на графовите што ќе ги внесува.

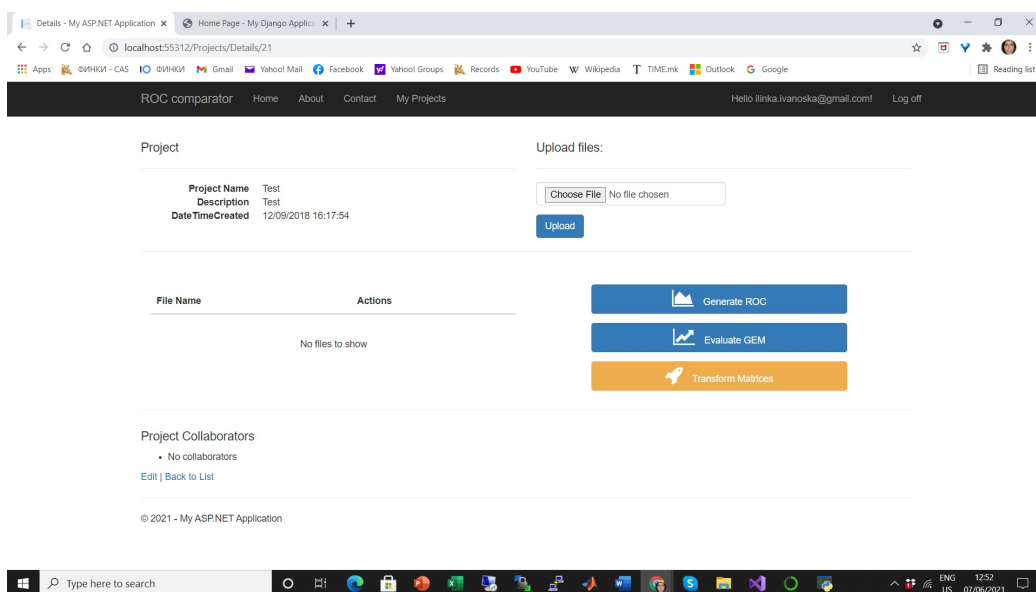
По влезот во некој проект, добива поглед со повеќе можности како на сл. 5-2. Во делот за трансформација на матрици корисникот може да избере одреден алгоритам за вгнездување на граф и да добие векторска репрезентација за графот од интерес како на сл. 5-3. Корисникот може да го тестира ефектот на димензионалноста на векторската репрезентација врз перформансите на секој од алгоритмите.

Покрај тоа, корисникот може да го тестира влијанието на одредени хиперпараметри на перформансите на алгоритмите за вгнездување на графовите. Поточно, системот нуди можност за тестирање на бета параметар во алгоритмот NOPE, бета параметар во алгоритмот SDNE и ламбда параметар во алгоритмот GF. Дополнително, корисникот може да ги одреди минималните и максималните вредности на хиперпараметарот, како и големината на чекорот помеѓу различните вредности на параметрите. Формуларот исто така нуди опција за избор на



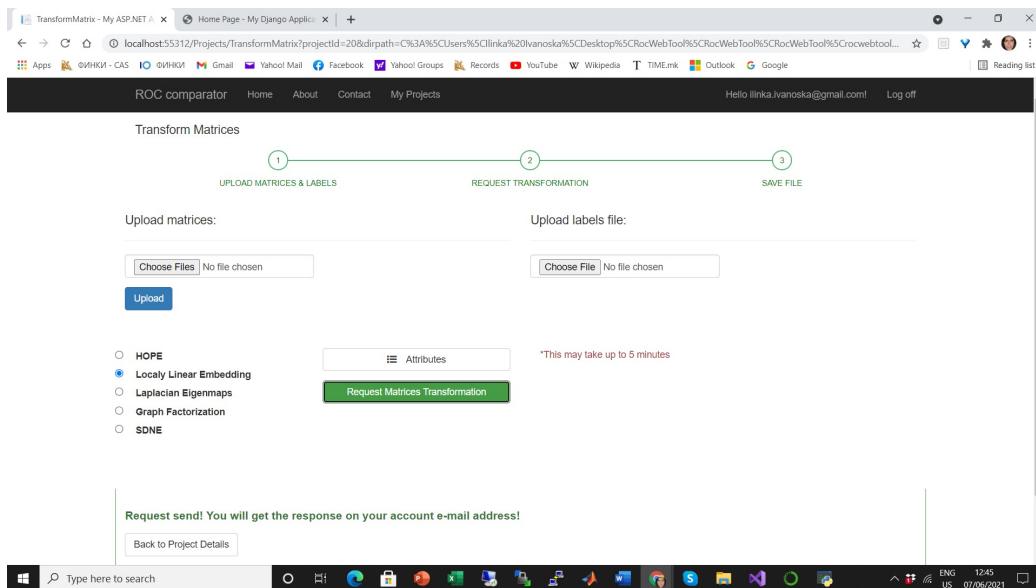
Слика 5-1: Почетен поглед на веб-системот за мрежни вгнездувања за логирање.

димензија на векторска претстава.

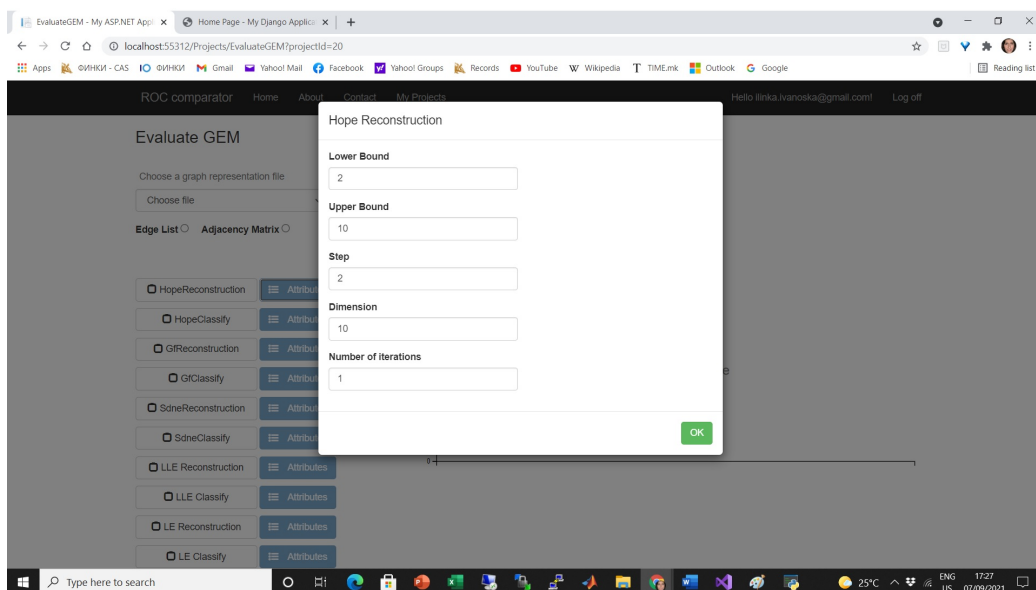


Слика 5-2: Поглед на веб-системот за мрежни вгнездувања за граф трансформации со сите опции откако ќе се логира корисникот.

Корисниците имаат можност да внесат граф и да го реконструираат со помош на некој од алгоритмите за вгнездување како на сл. 5-5. Како пример, тука е избрана мозочна мрежа за реконструкција од податочното множество COBRE

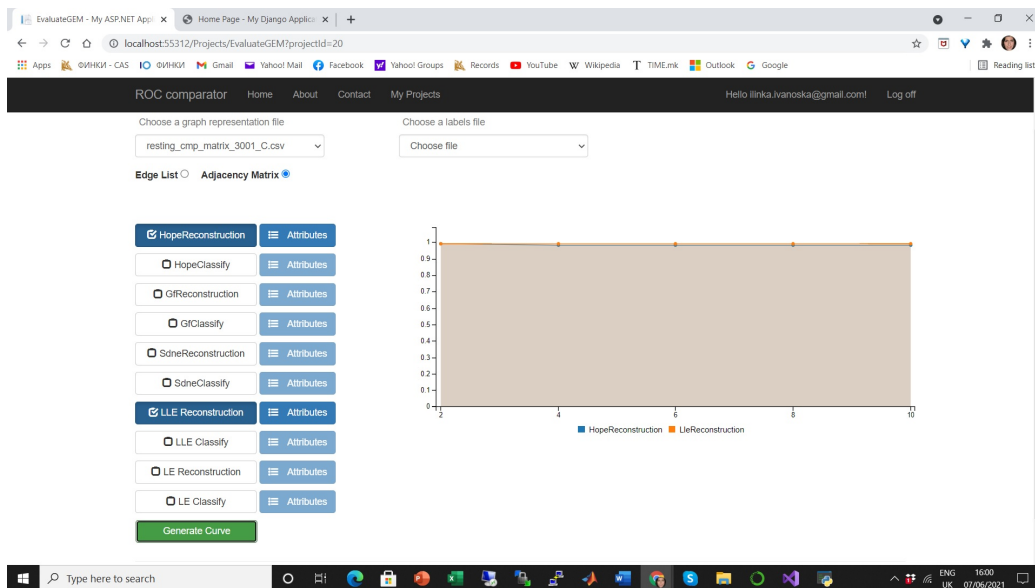


Слика 5-3: Поглед на веб-системот за мрежни вгнездувања за граф трансформации за преземање на векторската репрезентација.



Слика 5-4: Поглед на веб-системот за мрежни вгнездувања за реконструкција на граф за избор на хиперпараметрите на алгоритмот HOPE.

(шизофренија), два алгоритми за мрежно вгнездување HOPE и LLE и соодветно се избрани хиперпараметрите на секој алгоритам. Изборот на хиперпараметрите на алгоритмот HOPE може да се види на сл. 5-4. И двата алгоритма даваат добра реконструкција на мозочната мрежа при различни димензии на векторската



Слика 5-5: Поглед на веб-системот за мрежни вгнездувања за реконструкција на граф со внесена мозочна мрежа.

репрезентација на графот.

Тековната верзија на системот овозможува проценка на ефектот на димензијата на векторска претстава врз перформансите на алгоритмите, како и проценка на ефектот на некои хиперпараметри на соодветните алгоритми. Дополнително, системот овозможува преземање на векторска претстава со избор на алгоритам и вредноста на неговите хиперпараметри. На овој начин, им овозможува на корисниците брзо и лесно да ја пронајдат најдобрата техника за нивниот проблем и да ја преземат векторската претстава за понатамошна анализа.

Исто така, тековната верзија на системот работи со релативно мали графови со неколку стотици јазли. Ова се должи на фактот дека сегашните имплементации на алгоритми од библиотеката GEM не се скалабилни за големи графови во реално време, што иако е проблем што треба да се реши во иднина, не е голем проблем од аспект на малите мозочни мрежи кои се предмет на оваа дисертација.

Во системот изборот на модели и тест параметрите се врши преку веб-форма и не бара никакво техничко знаење. Ова ја надминува потребата од специфично познавање на програмскиот јазик. Резултатите од споредбите се прикажуваат на интерактивни графикони. Откако ќе се споредат перформансите на различни модели со различни параметри, корисниците можат да донесат поинформирана одлука за моделот и параметрите што ќе ги користат за нивната примена. Системот нуди алатка во која корисниците можат да го специфицираат избраниот

модел и неговите параметри и да ја добијат векторската претстава генерирана од овие параметри за идна употреба во нивните апликации. На овој начин, алгоритмите за вгнездување на графови се достапни за поголем број на корисници, а процесот на избор на најсоодветна техника станува многу полесен. Системот нуди можност за едноставно проширување на бројот на модели за трансформација, тест параметрите и задачи за споредба на моделите. Сето ова им дава можност на невронаучниците како корисници кои често се со лимитирано техничко познавање на одреден програмски јазик едноставно да добијат резултат за анализа на некоја мозочна мрежа преку алгоритми за вгнездување на графови.

Почетна јавно достапна верзија на системот е дадена на <https://github.com/ilinkaivanoska/rocwebtool/tree/master>. Веб системот е сеуште во доработка, и е планирано како идна работа да се вметнат и функционалности со внес на сите алгоритми обработени во оваа дисертација за анализа на мозочни мрежи.

Глава 6

Заклучок

Како заклучок од оваа докторска дисертација може да се извлече дека се развиени и поставени методи кои опфаќаат неколку аспекти за обработка на мозочните мрежи. Овие аспекти вклучуваат:

- (i) развој на методи (предложен Бесов метод) за реконструкција и ренормализација на функционални мозочни мрежи;
- (ii) развој на пресметковно брз метод за да се измерат разликите во мозочната поврзаност засновани на рангирање на тежината на врските;
- (iii) преглед на статистички методи и методи базирани на машинско учење за селекција на карактеристики кај мозочни функционални мрежи; и
- (iv) развој на веб базиран систем за евалуација на техники базирани на мрежни вгнездувања за обработка на мозочни мрежи.

Во однос на првиот аспект, нашите откритија укажуваат на тоа дека: (i) некои чекори во реконструкцијата на функционалните мрежни репрезентации на fMRI активноста, и конкретно ренормализацијата на врските на ROI, имаат важно влијание во набљудуваните тополошки структури; (ii) големината на овие тополошки разлики зависи од изборот на стратегијата за праг и од употребата на фреквенциски пристапи наспроти баесови реконструкции; (iii) ваквите разлики не влијаат на целокупната моќ на дискриминација на реконструираниите мрежи. Затоа треба да се внимава на изборот на достапни методи и во толкување на резултатите што ги даваат овие.

Во однос на вториот аспект, треба да се споменат неколку работи. Прво, во повеќето случаи, трансформацијата дава нетривијална слика за целокупната структура на разликите помеѓу дадената патологија и нејзината соодветна контролна група во однос на разликите во рангирањето на јачината на конективноста (поврзаноста). Овој резултат укажува на тоа дека, дури и во отсуство на информации за топографска локализација, кои честопати се сметаат за неопход-

ни во неврологијата на ниво на системот, трансформацијата задржува доволно информации за моделот на мрежна поврзаност за да ги дискриминира патолошките сигнали од здравите контролни. Второ, нашите резултати сугерираат дека релевантните информации за дискриминација се кодирани во хиерархијата на јачината на мозочната поврзаност, но исто така, што е важно, во нејзината варијабилност. Трето, геометријата на ранг-различната трансформација во медијана-стр.девијација простор претставува изразена специфичност на патологијата, која често може да се увиди на прв поглед. Конечно, се чини дека резултатите се постојано појасни за електроенцефалографијата (ЕЕГ) отколку за податоците за функционална магнетна резонанца (fMRI).

Во однос на третиот аспект, овој придонес се фокусира на две широки семејства на методи за избор на врски, т.е. статистички и машинско учење. Споредивме големо множество на различни методи за избор на врски, извлечени од статистиката и машинското учење и ги споредивме нивните перформанси во однос на перформансната цена (т.е. колку информации се задржуваат на крајот, измерени преку задача за класификација), стабилноста и пресметковна цена. Изненадувачки, со некои исклучоци, резултатите покажуваат, сите методи даваат сличен резултат на класификација; со други зборови, и покрај нивните внатрешни разлики и основните хипотези, информациите што ги задржуваат се скоро исти. Сепак, бришењето на неинформативни врски не само што го поедноставува проблемот, туку дури и го зголемува нашиот капацитет за извлекување заклучоци од него. Во однос на интра-стабилноста на секој метод на избор, т.е. неговиот капацитет за постојано давање на истото множество врски, резултатите се генерално ниски. Истото се случува и со стабилноста помеѓу методите за избор на врски, т.е. интер-стабилноста, која е исто така ниска. Како заклучок, било метод е добар, и навистина чекорот за избор дури е неопходен за да се подобри нашето разбирање за мозокот.

За крај, презентиравме развој на веб системот за проценка на техниките за векторско претставување на мозочни мрежи. Овие техники се една од најмоќните алатки за анализа на графикони денес. Сепак, изборот на најсоодветната техника за даден проблем се меша со нееднаквоста на нивните имплементации и потребата од техничко знаење за нивно модифицирање и тестирање. Веб-системот претставен им овозможува на корисниците да оценат различни алгоритми на лесен и интуитивен начин. Во моментов, системот поддржува неколку техники за векторско претставување на мозочни мрежи. Покрај тоа, системот овозможува проценка на техниките со употреба на метод на реконструкција на граф. Тековната верзија на системот овозможува проценка на ефектот на димензија-

та на векторска претстава врз перформансите на алгоритмите, како и проценка на ефектот на некои хиперпараметри на соодветните алгоритми. Веб системот е сеуште во доработка, и е планирано како идна работа да се вметнат и функционалности со внес на сите алгоритми обработени во оваа дисертација за анализа на мозочни мрежи.

Додаток А

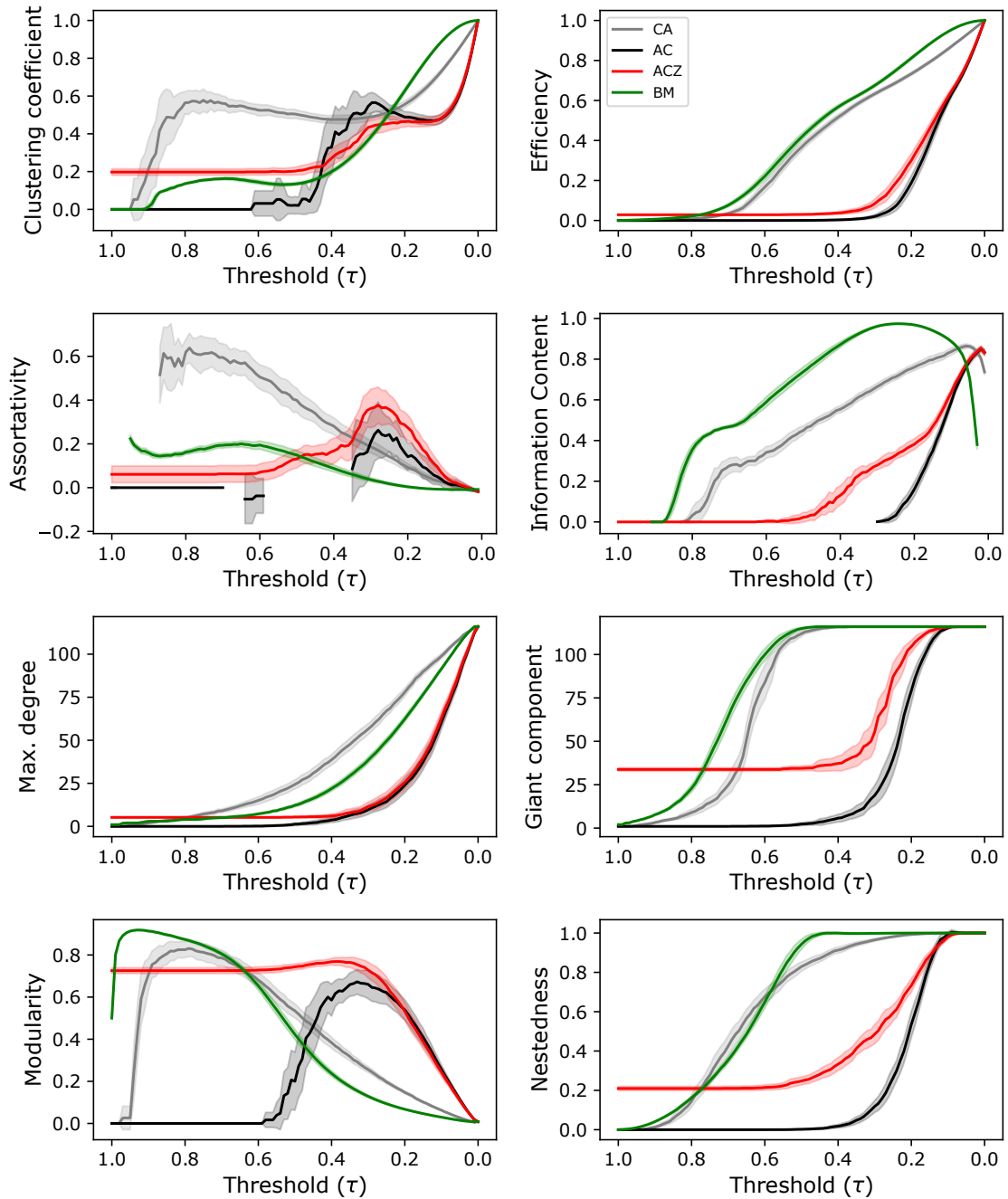
Дополнителни слики и табели

Табела А.1: Синтеза на сите дополнителни слики вклучени овде.

Под. мно.	Атлас	Праг	Субјекти	Слика
ABIDE	AAL 1mm	Апсолутен	Контролни	Доп. сл. А-1
ABIDE	AAL 1mm	Пропорционален	Контролни	Доп. сл. А-2
ABIDE	AAL 1mm	Апсолутен	Пациенти	Доп. сл. А-3
ABIDE	AAL 1mm	Пропорционален	Пациенти	Доп. сл. А-4
ABIDE	AAL 2mm	Апсолутен	Контролни	Доп. сл. А-5
ABIDE	AAL 2mm	Пропорционален	Контролни	Доп. сл. А-6
ABIDE	AAL 2mm	Апсолутен	Пациенти	Доп. сл. А-7
ABIDE	AAL 2mm	Пропорционален	Пациенти	Доп. сл. А-8
ABIDE	Brainnetome	Апсолутен	Контролни	Доп. сл. А-9
ABIDE	Brainnetome	Пропорционален	Контролни	Доп. сл. А-10
ABIDE	Brainnetome	Апсолутен	Пациенти	Доп. сл. А-11
ABIDE	Brainnetome	Пропорционален	Пациенти	Доп. сл. А-12
ABIDE	НО	Апсолутен	Контролни	Доп. сл. А-13
ABIDE	НО	Пропорционален	Контролни	Доп. сл. А-14
ABIDE	НО	Апсолутен	Пациенти	Доп. сл. А-15
ABIDE	НО	Пропорционален	Пациенти	Доп. сл. А-16
ABIDE	AAL 1mm	Апсолутен	Контролни <i>vs.</i> Пациенти	Доп. сл. А-17
ABIDE	AAL 1mm	Пропорционален	Контролни <i>vs.</i> Пациенти	Доп. сл. А-18
ABIDE	AAL 2mm	Апсолутен	Контролни <i>vs.</i> Пациенти	Доп. сл. А-19
ABIDE	AAL 2mm	Пропорционален	Контролни <i>vs.</i> Пациенти	Доп. сл. А-20
ABIDE	Brainnetome	Апсолутен	Контролни <i>vs.</i> Пациенти	Доп. сл. А-21
ABIDE	Brainnetome	Пропорционален	Контролни <i>vs.</i> Пациенти	Доп. сл. А-22
ABIDE	НО	Апсолутен	Контролни <i>vs.</i> Пациенти	Доп. сл. А-23
ABIDE	НО	Пропорционален	Контролни <i>vs.</i> Пациенти	Доп. сл. А-24
ABIDE	Сите атласи	Апсолутен	Контролни	Доп. сл. А-25
ABIDE	Сите атласи	Пропорционален	Контролни	Доп. сл. А-26
ABIDE	Сите атласи	Апсолутен	Пациенти	Доп. сл. А-27
ABIDE	Сите атласи	Пропорционален	Пациенти	Доп. сл. А-28
COBRE	AAL 2mm	Апсолутен	Контролни	Доп. сл. А-29
COBRE	AAL 2mm	Пропорционален	Контролни	Доп. сл. А-30
COBRE	AAL 2mm	Апсолутен	Пациенти	Доп. сл. А-31
COBRE	AAL 2mm	Пропорционален	Пациенти	Доп. сл. А-32
COBRE	AAL 2mm	Апсолутен	Контролни <i>vs.</i> Пациенти	Доп. сл. А-33
COBRE	AAL 2mm	Пропорционален	Контролни <i>vs.</i> Пациенти	Доп. сл. А-34

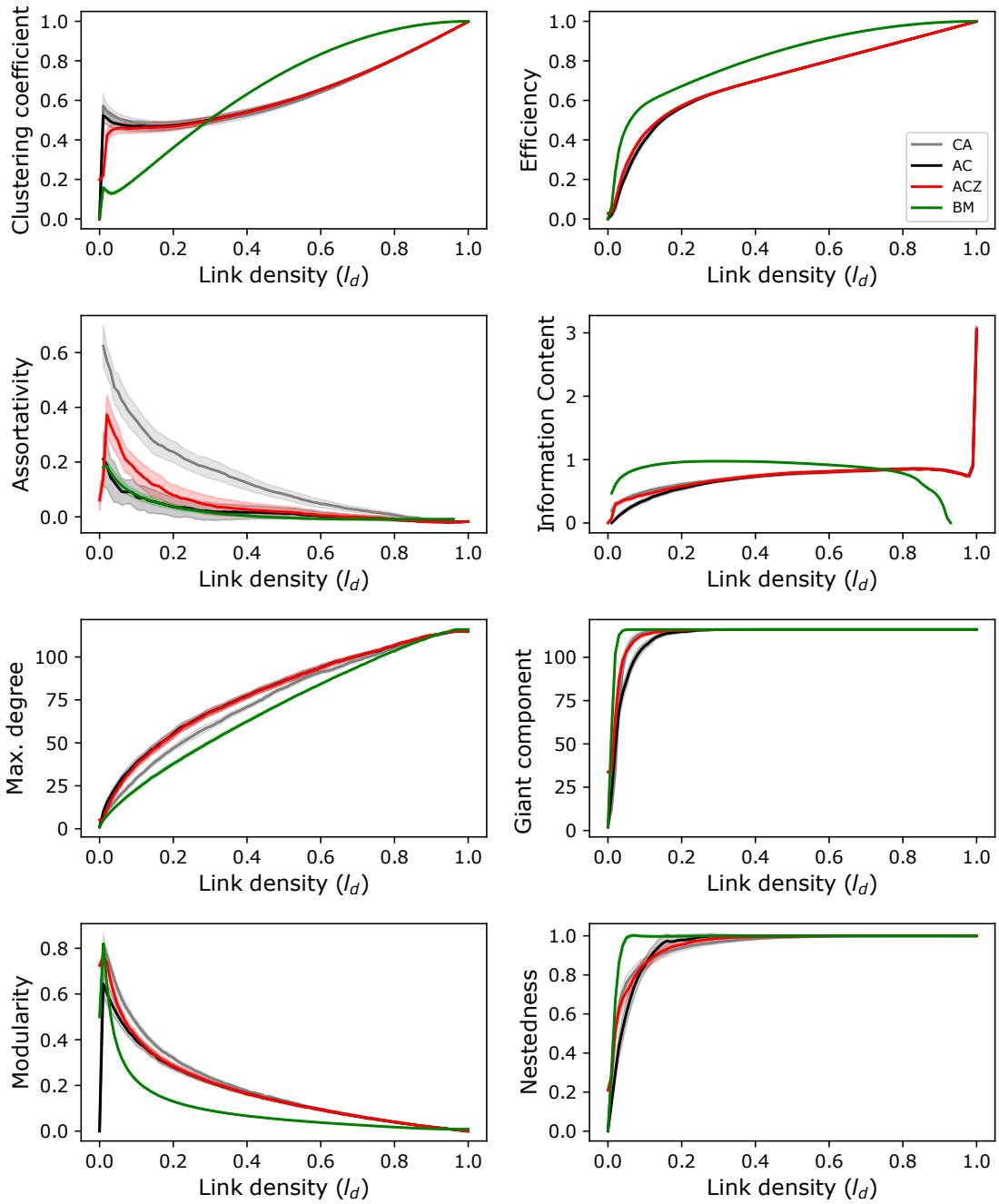
A.1 ABIDE: Сите резултати

ABIDE: AAL 1mm, τ , контролни субјекти



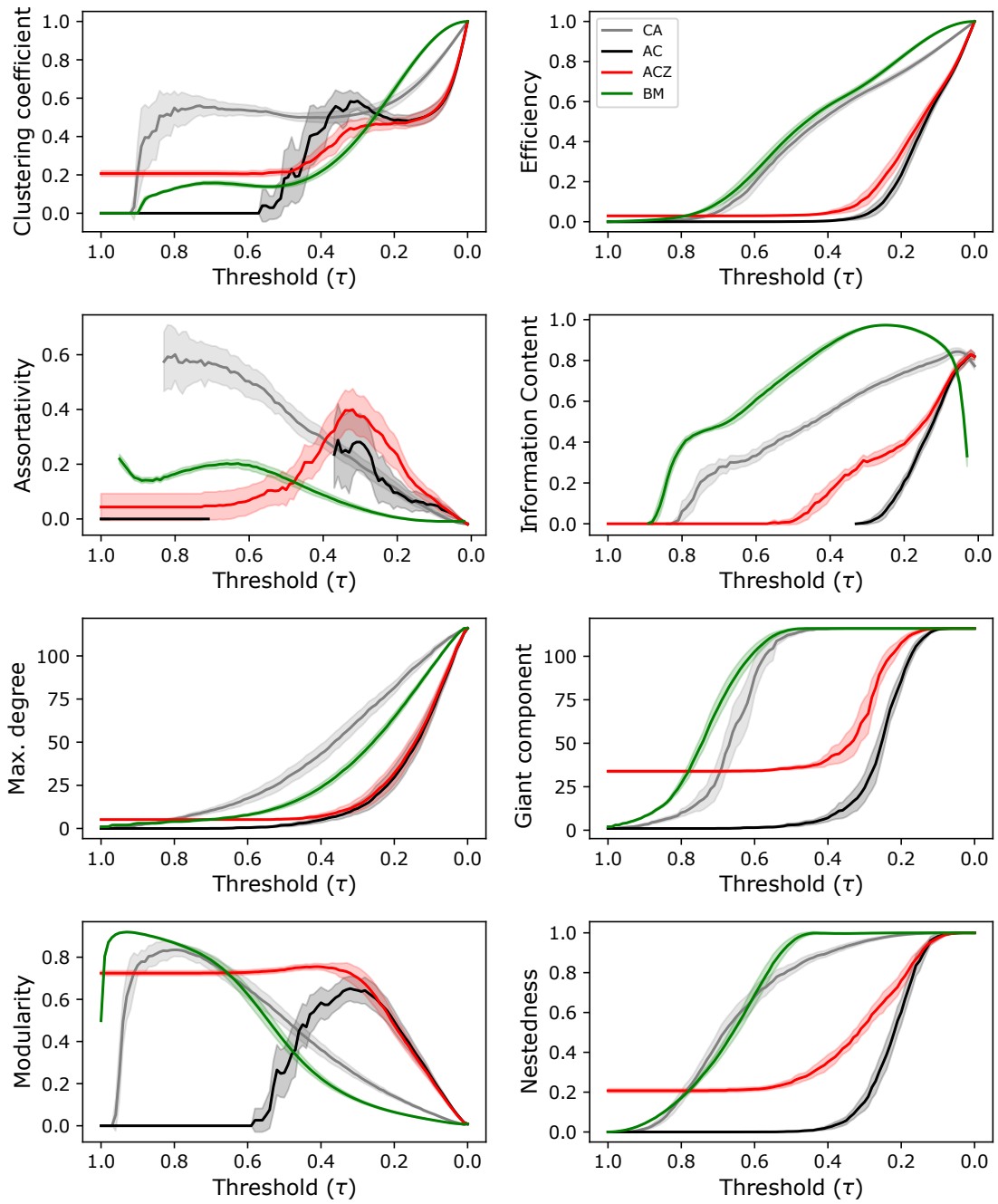
Слика А-1: ABIDE, AAL 1mm, τ , контролни субјекти.

ABIDE: AAL 1mm, l_d контролни субјекти



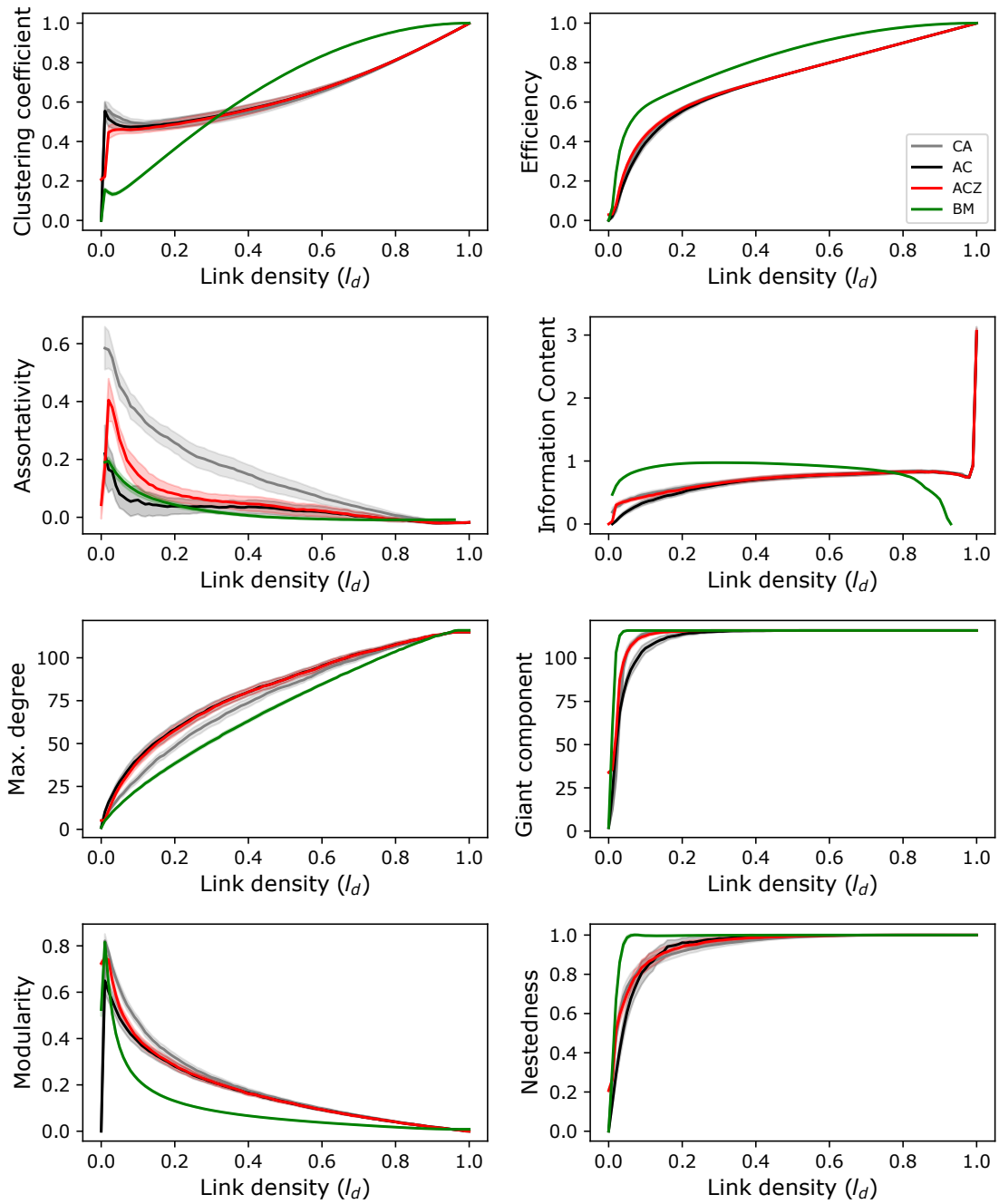
Слика А-2: ABIDE, AAL 1mm, l_d , контролни субјекти.

ABIDE: AAL 1mm, τ , пациенти



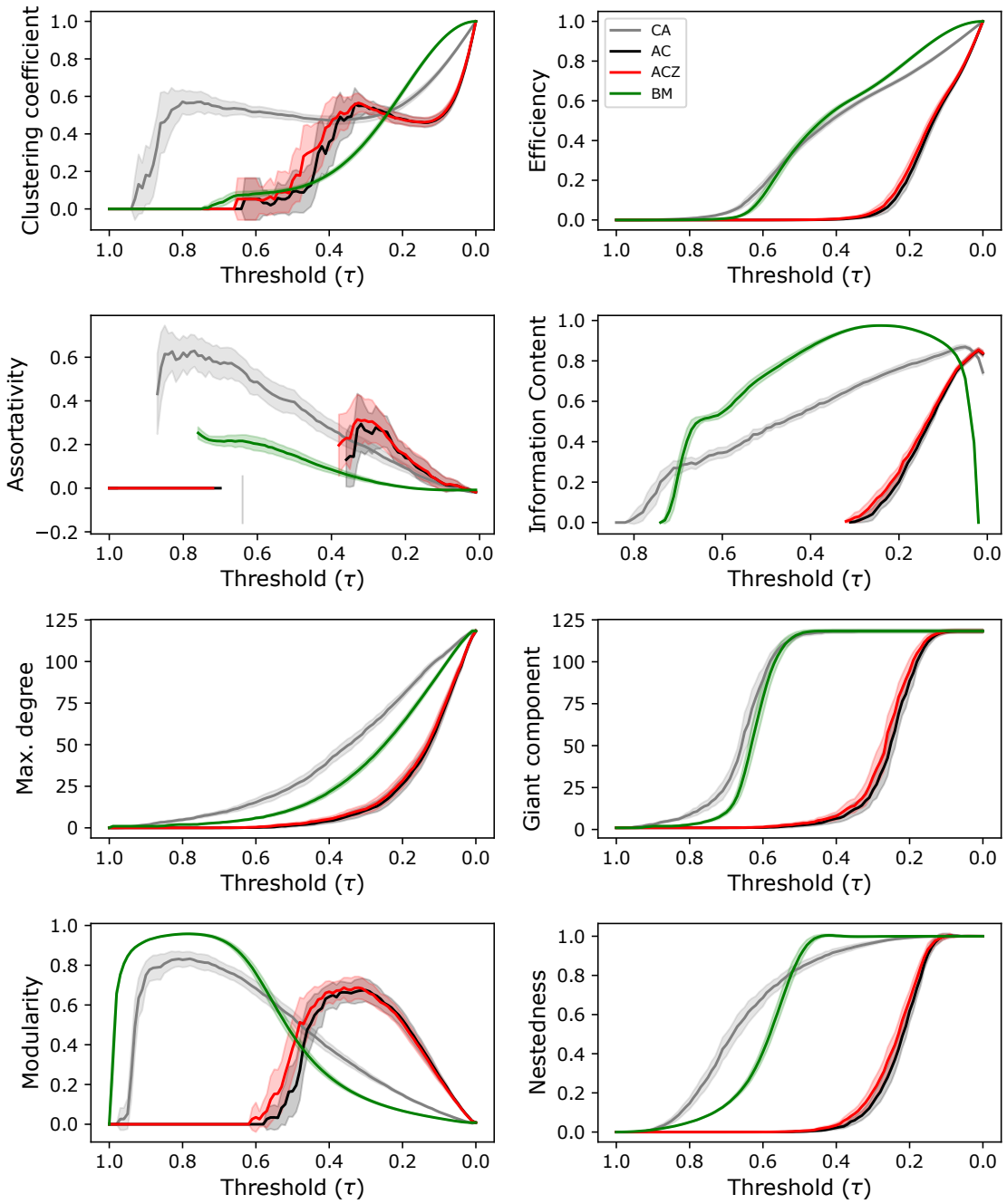
Слика А-3: ABIDE, AAL 1mm, τ , пациенти.

ABIDE: AAL 1mm, l_d , пациенти



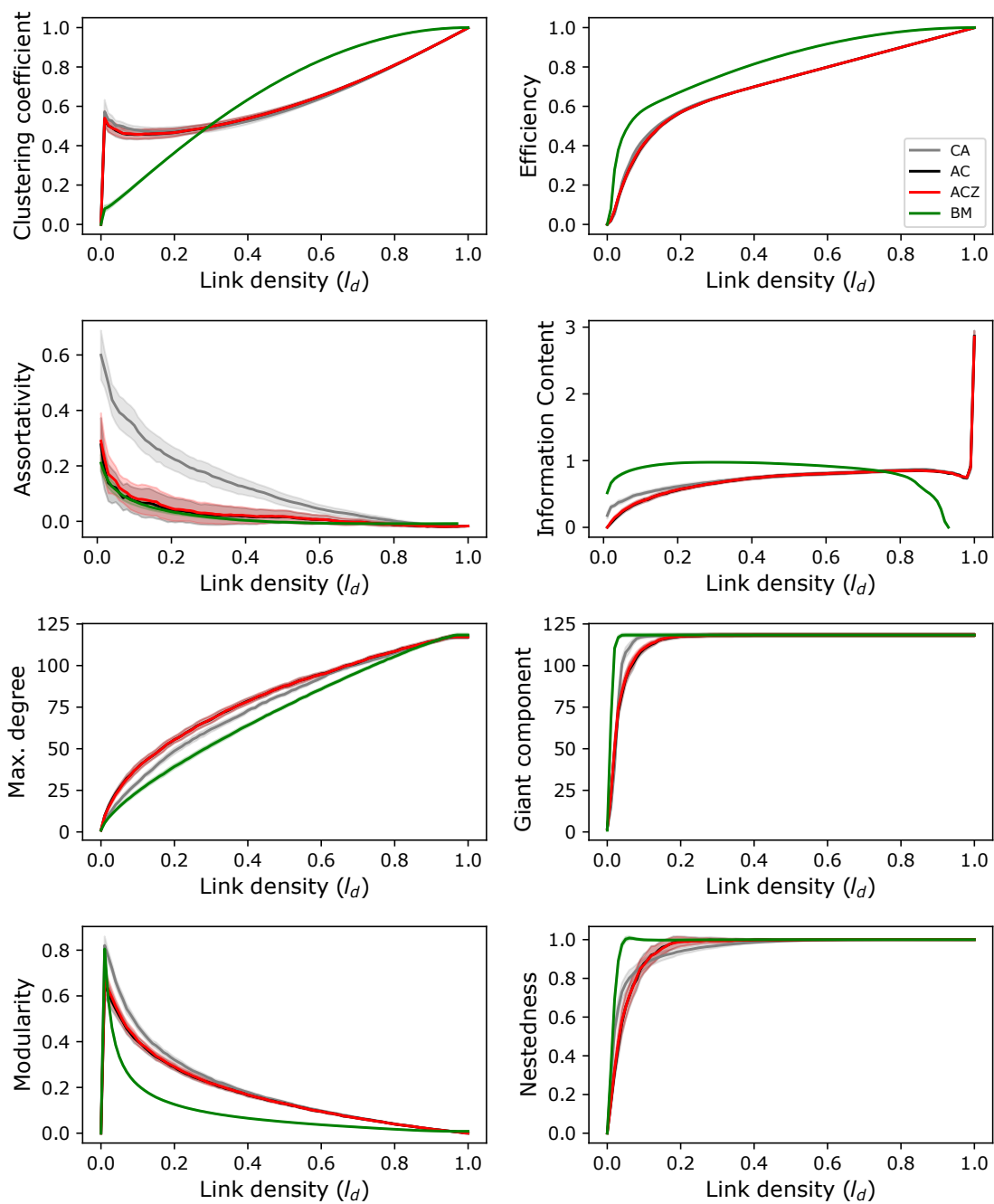
Слика А-4: ABIDE, AAL 1mm, l_d , пациенти.

ABIDE: AAL 2mm, τ , контролни субјекти



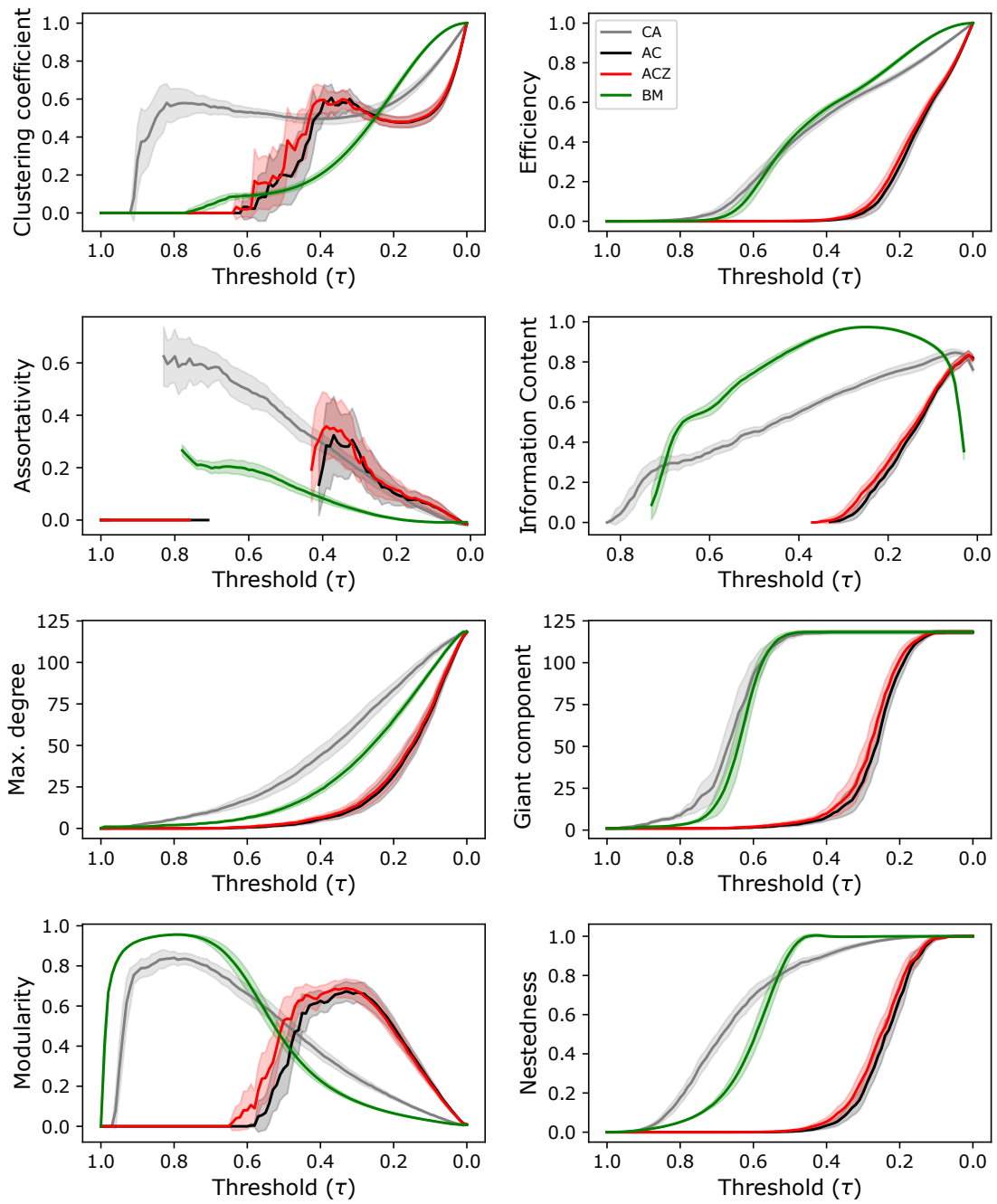
Слика А-5: ABIDE, AAL 2mm, τ , контролни субјекти.

ABIDE: AAL 2mm, l_d контролни субјекти



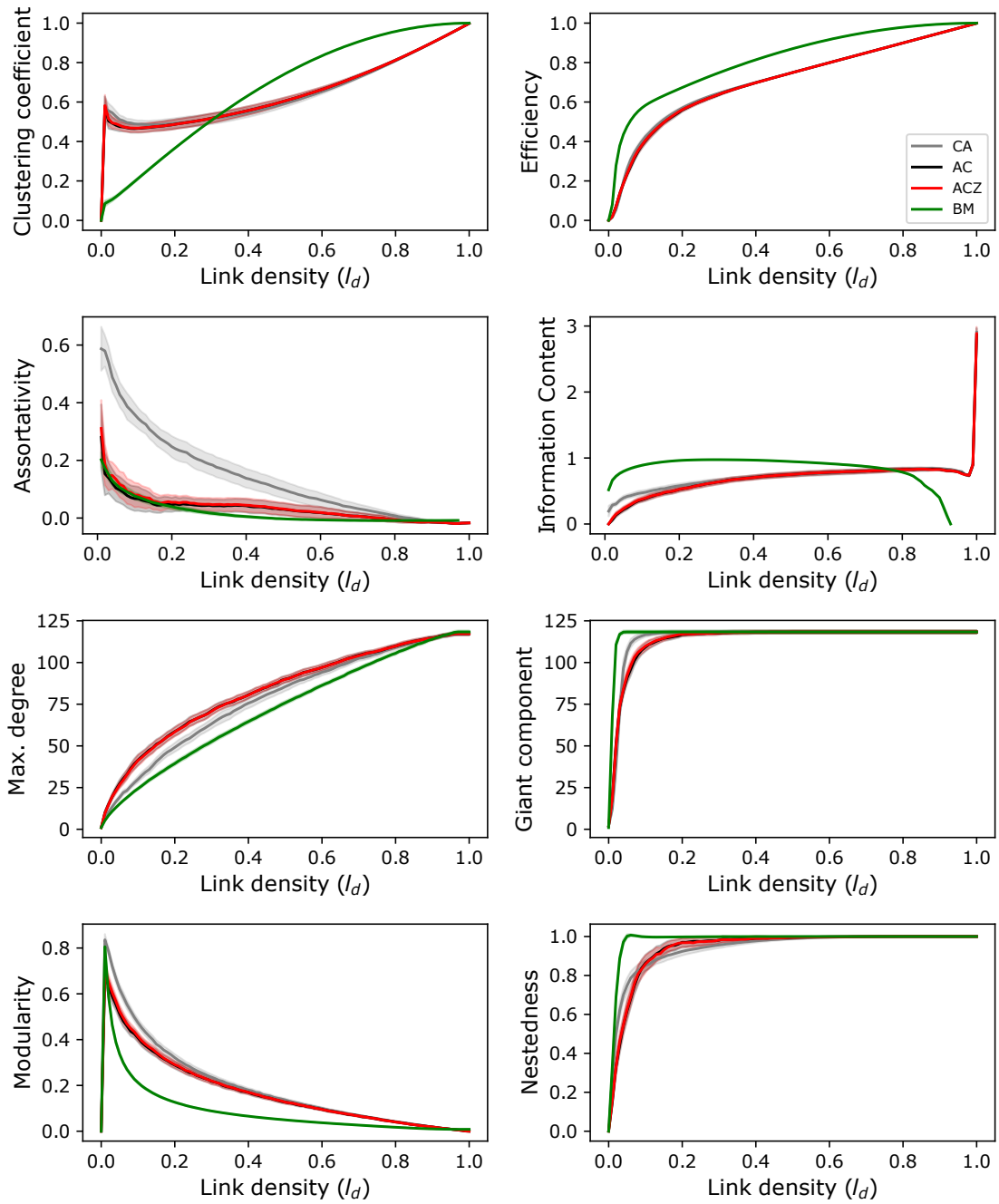
Слика А-6: ABIDE, AAL 2mm, l_d , контролни субјекти.

ABIDE: AAL 2mm, τ , пациенти



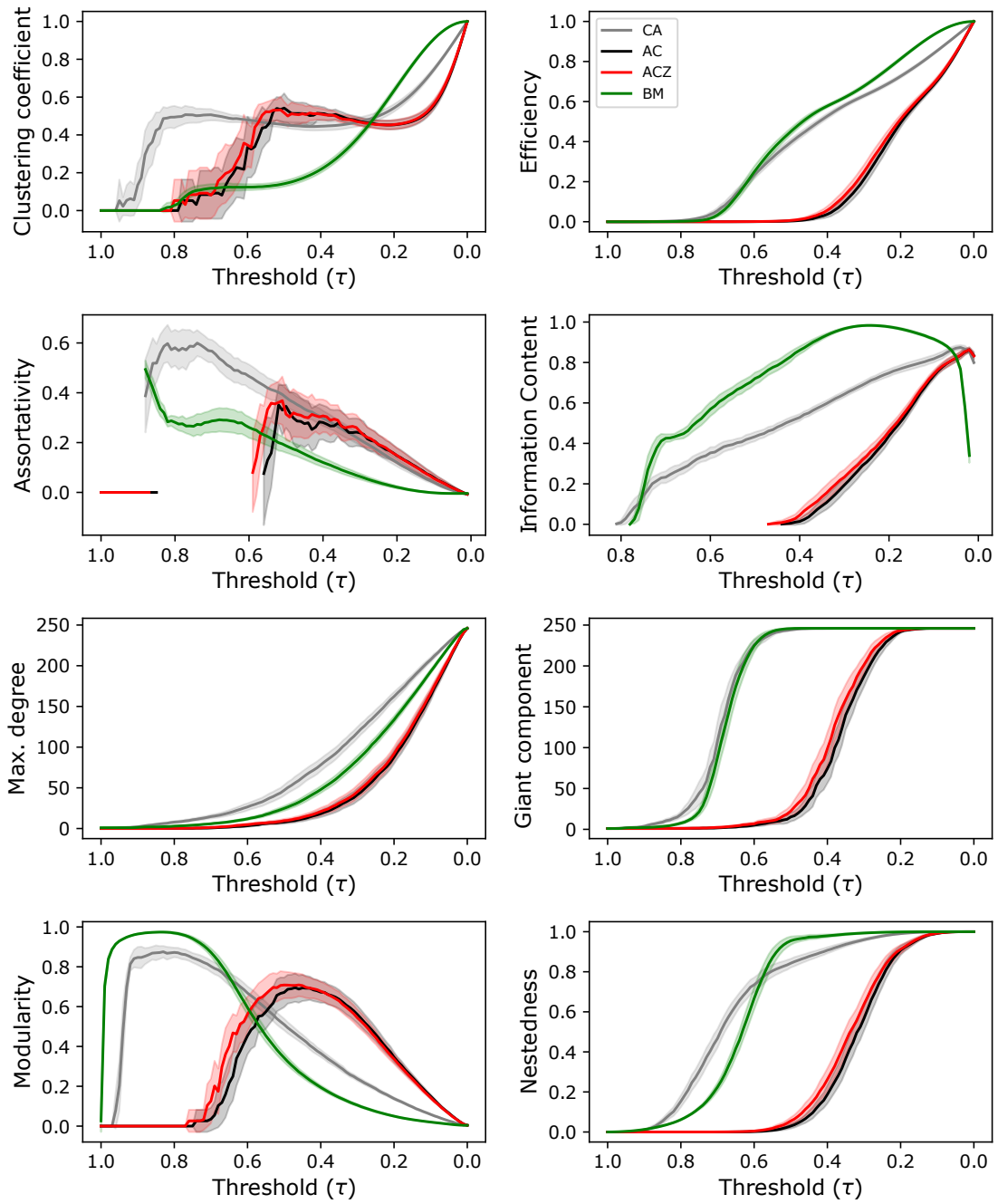
Слика А-7: ABIDE, AAL 2mm, τ , пациенти.

ABIDE: AAL 2mm, l_d , пациенти



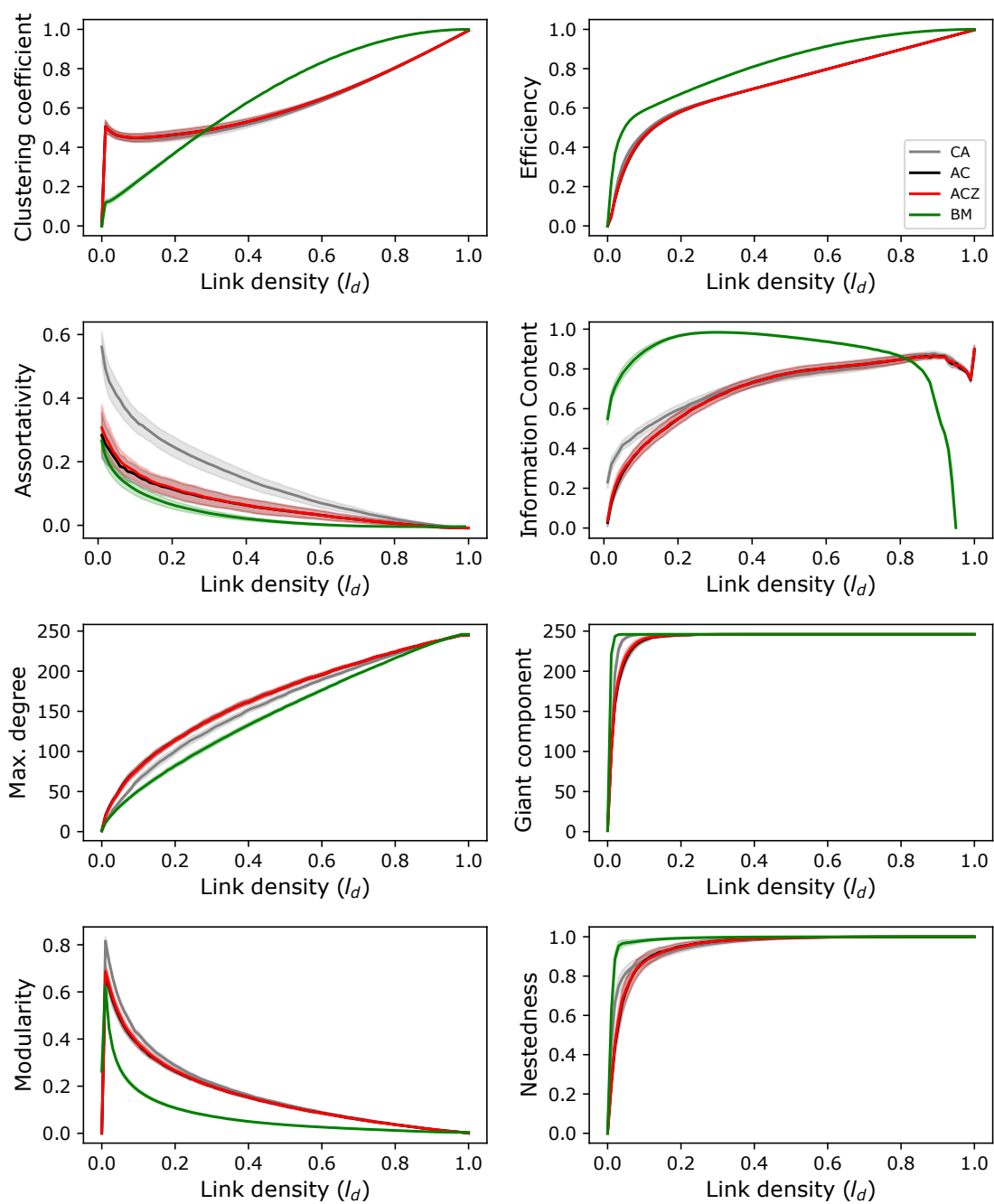
Слика А-8: ABIDE, AAL 2mm, l_d , пациенти.

ABIDE: Brainnetome, τ , контролни субјекти



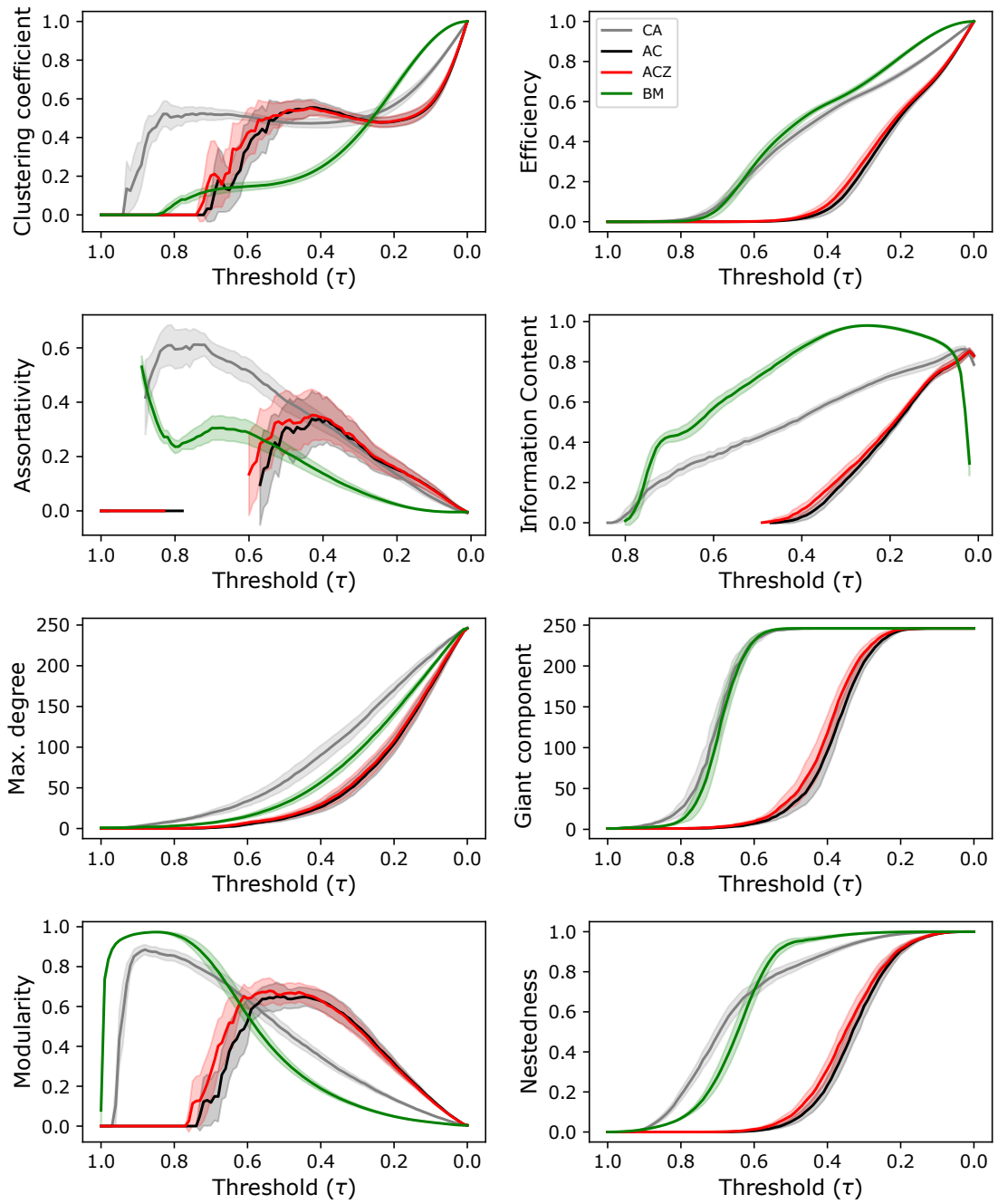
Слика А-9: ABIDE, Brainnetome, τ , контролни субјекти.

ABIDE: Brainnetome, l_d контролни субјекти



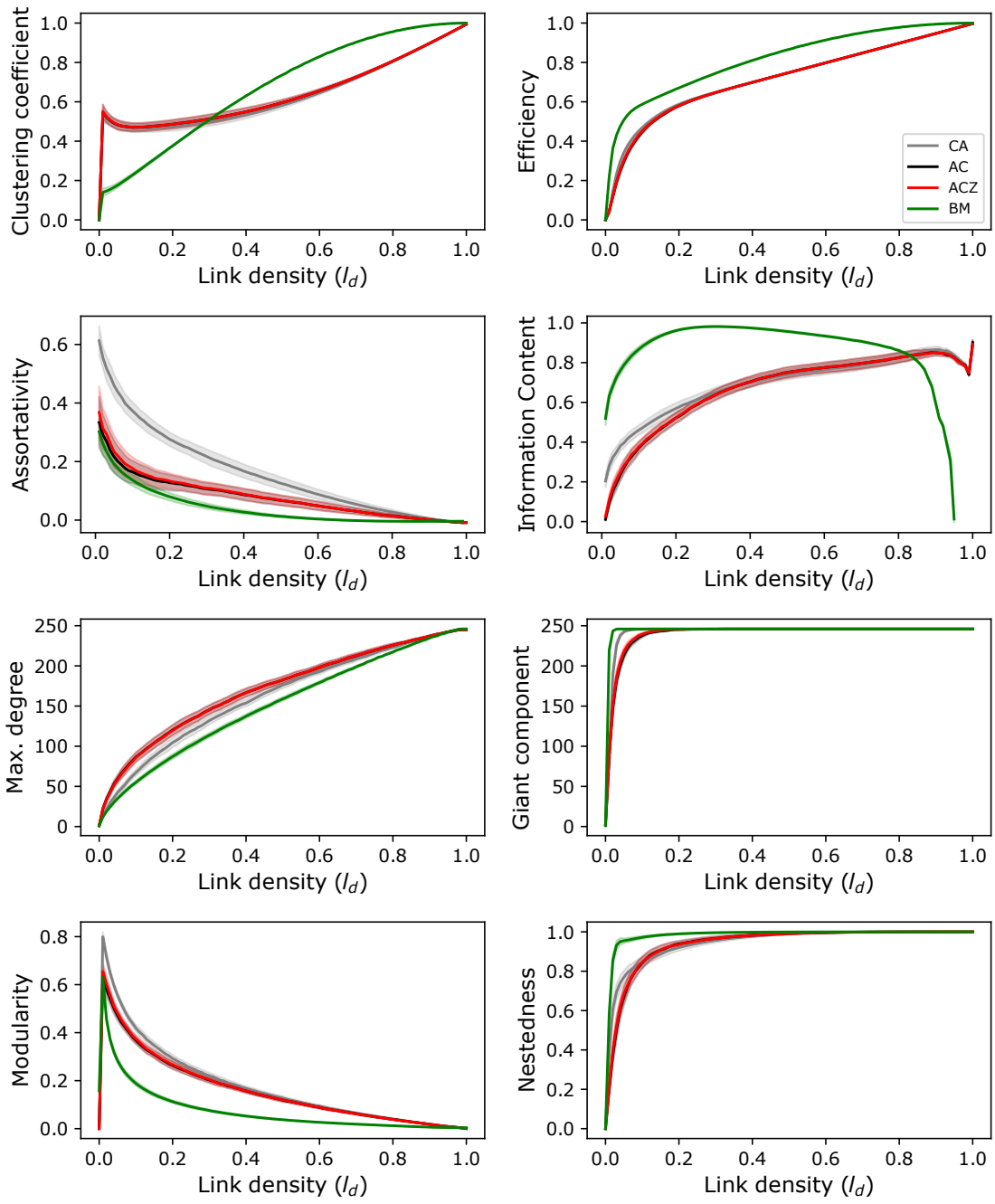
Слика А-10: ABIDE, Brainnetome, l_d , контролни субјекти.

ABIDE: Brainnetome, τ , пациенти



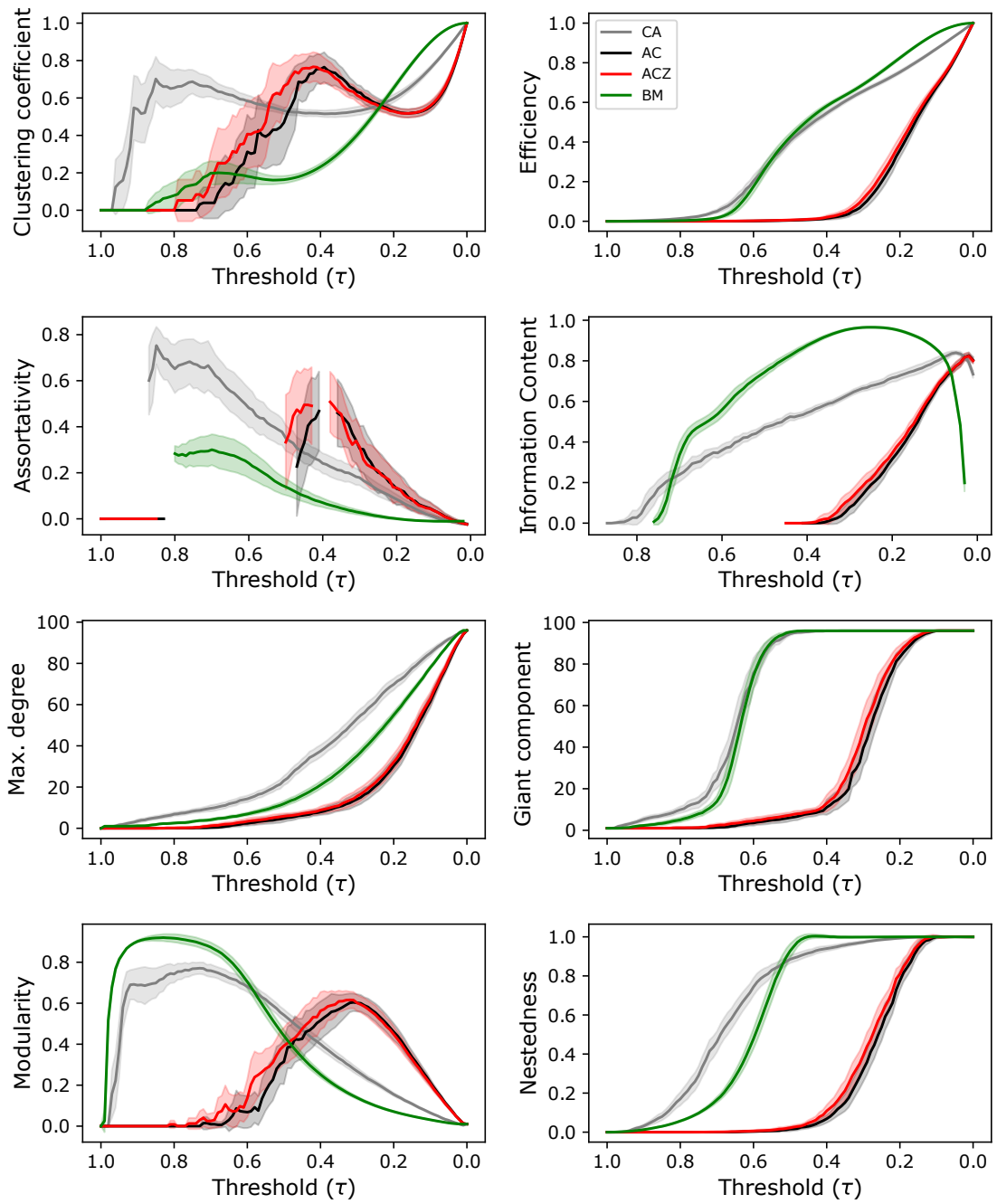
Слика А-11: ABIDE, Brainnetome, τ , пациенти.

ABIDE: Brainnetome, l_d пациенти



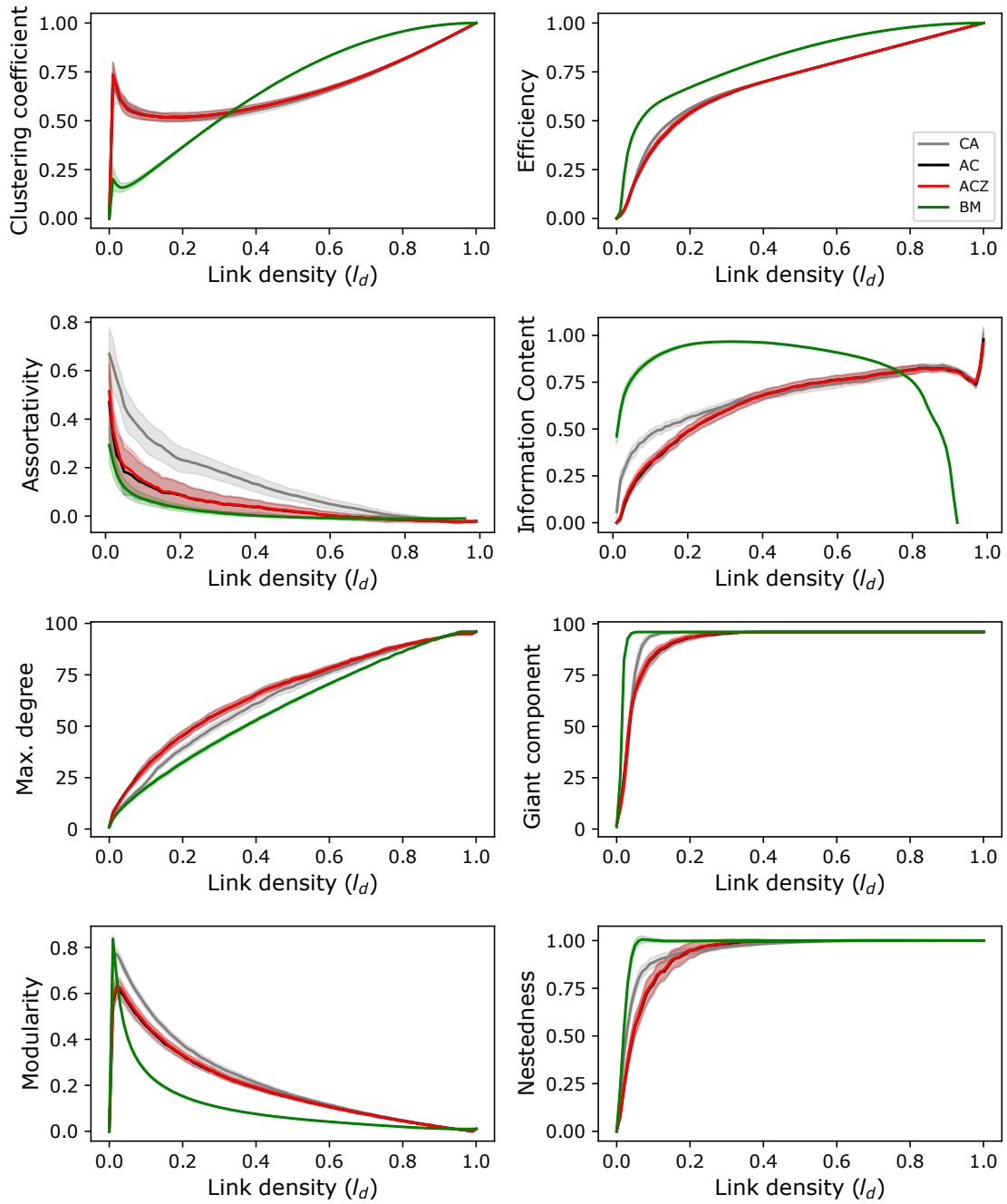
Слика А-12: ABIDE, Brainnetome, l_d пациенти.

ABIDE: HO, τ , контролни субјекти



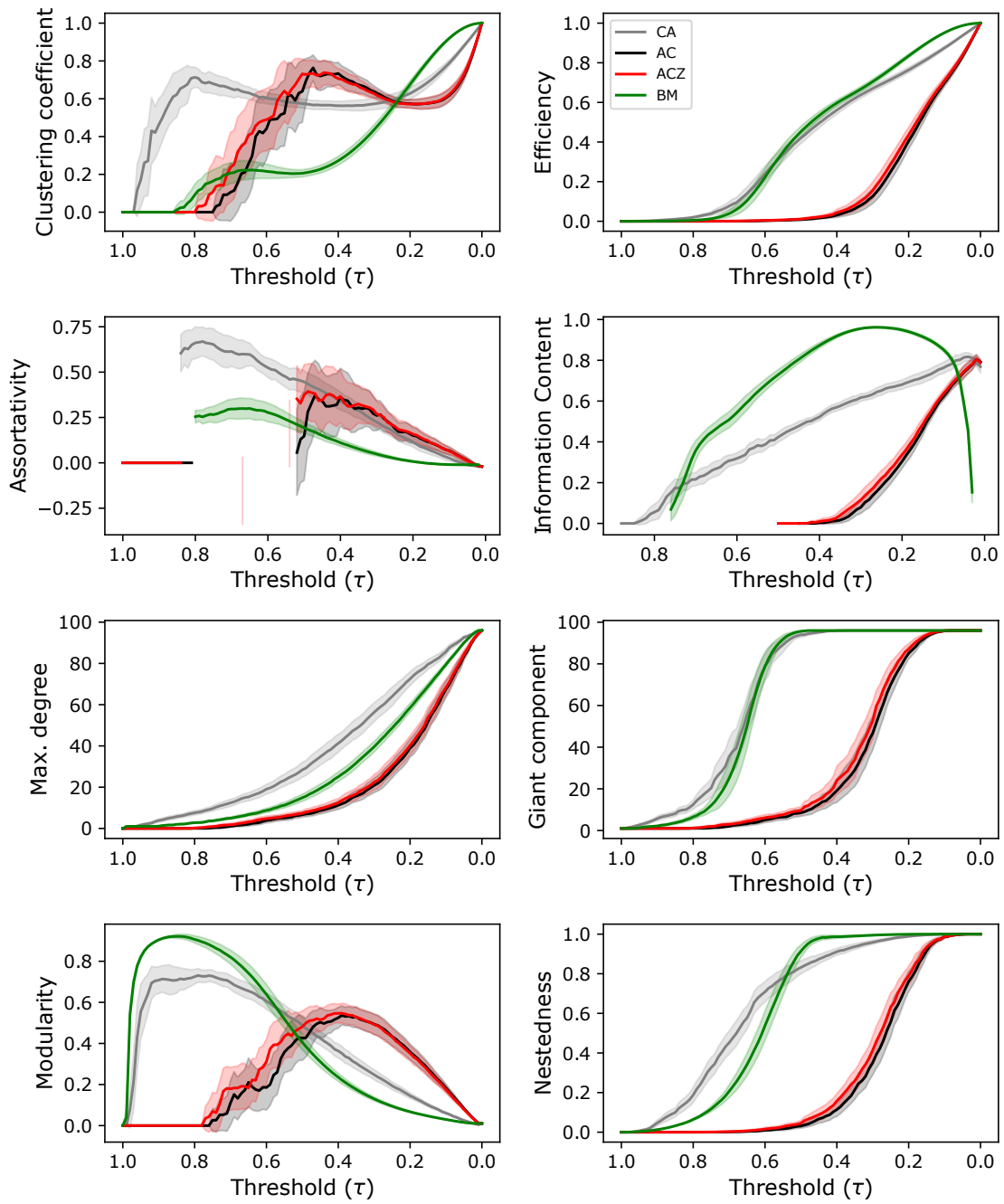
Слика А-13: ABIDE, HO, τ , контролни субјекти.

ABIDE: HO, l_d контролни субјекти



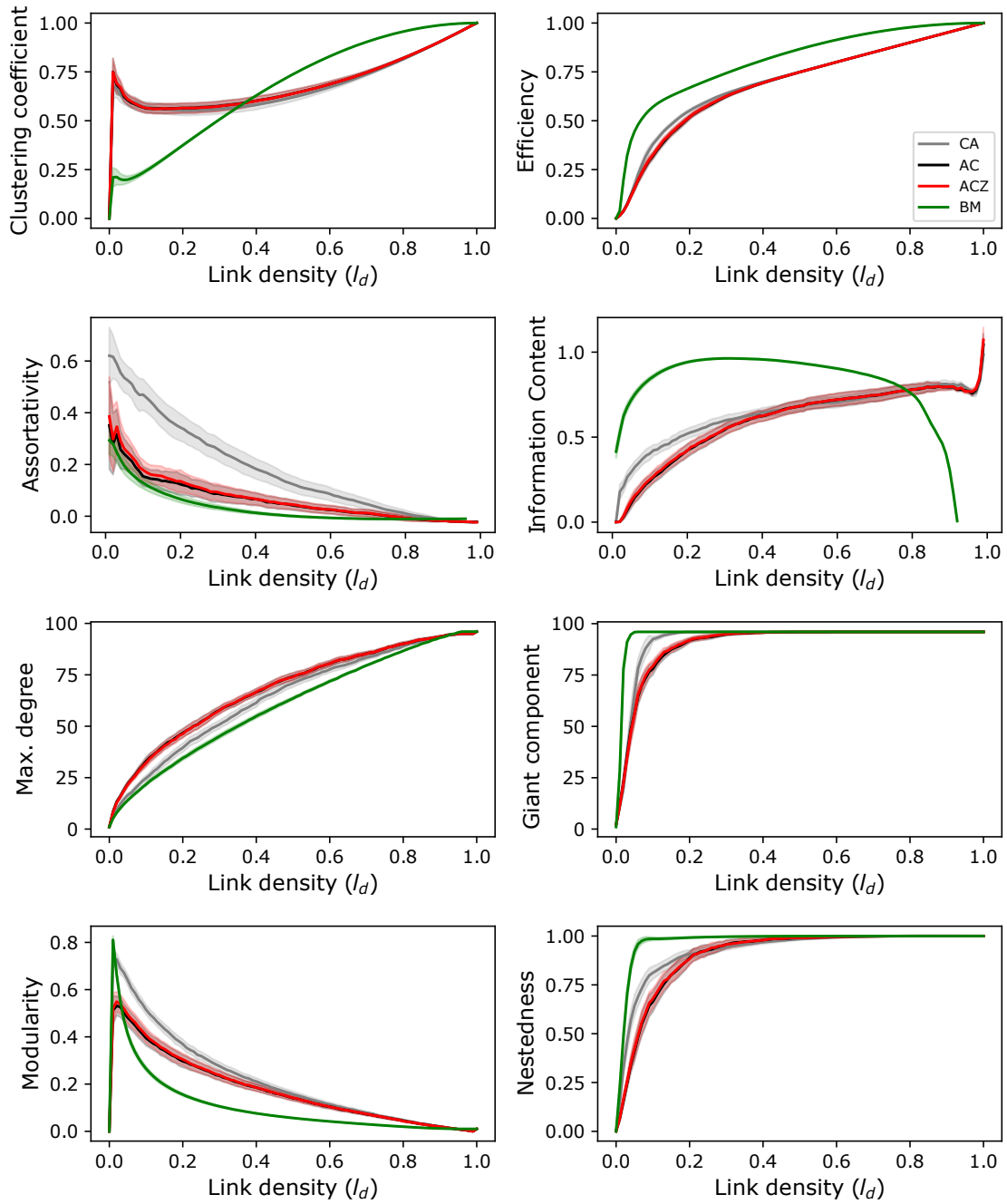
Слика А-14: ABIDE, HO, l_d , контролни субјекти.

ABIDE: HO, τ , пациенти



Слика А-15: ABIDE, HO, τ , пациенти.

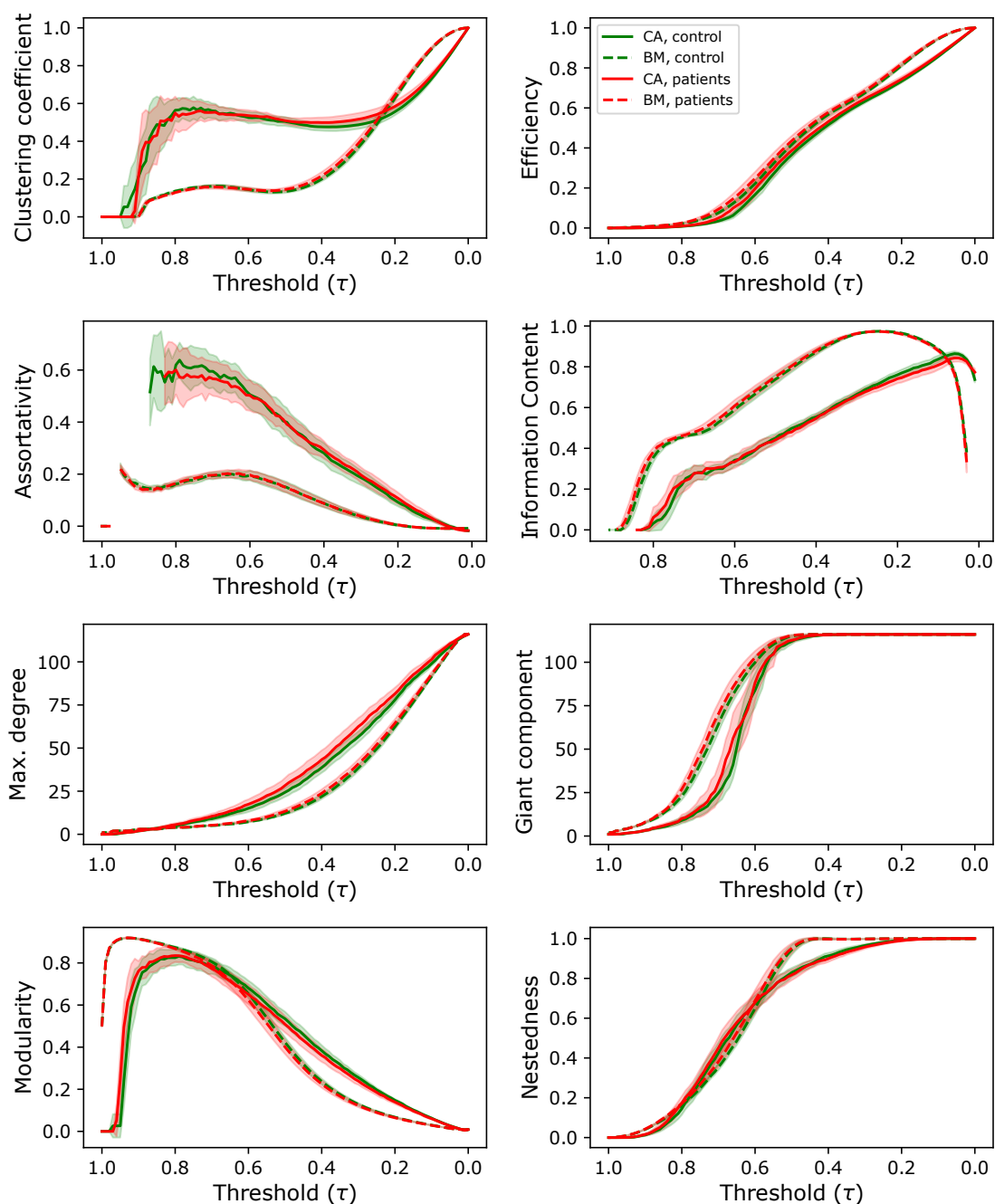
ABIDE: HO, l_d , пациенти



Слика А-16: ABIDE, HO, l_d , пациенти.

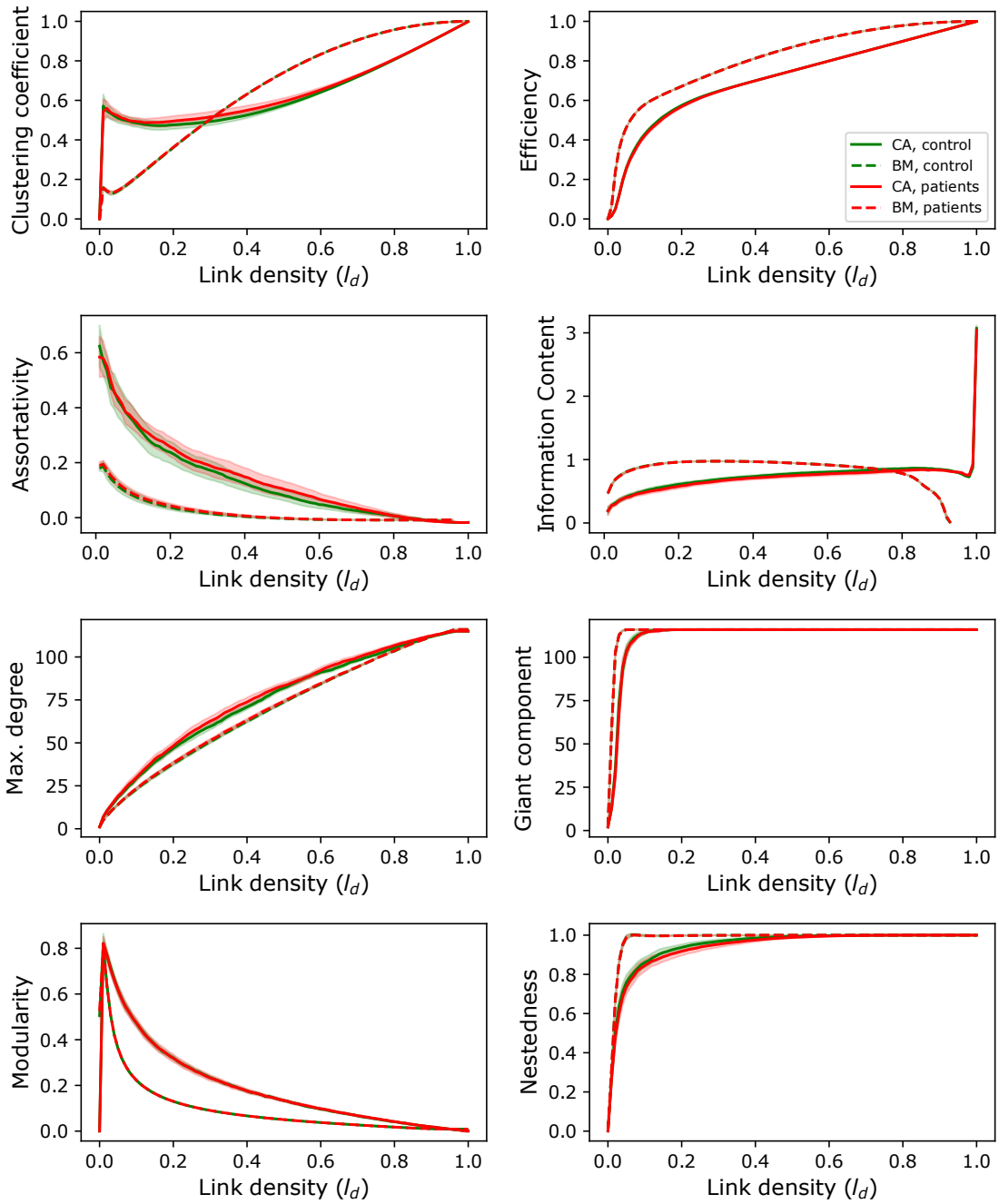
A.2 ABIDE: Споредба контролни субјекти vs. пациенти

ABIDE: AAL 1mm, τ , контролни субј. vs. пациенти



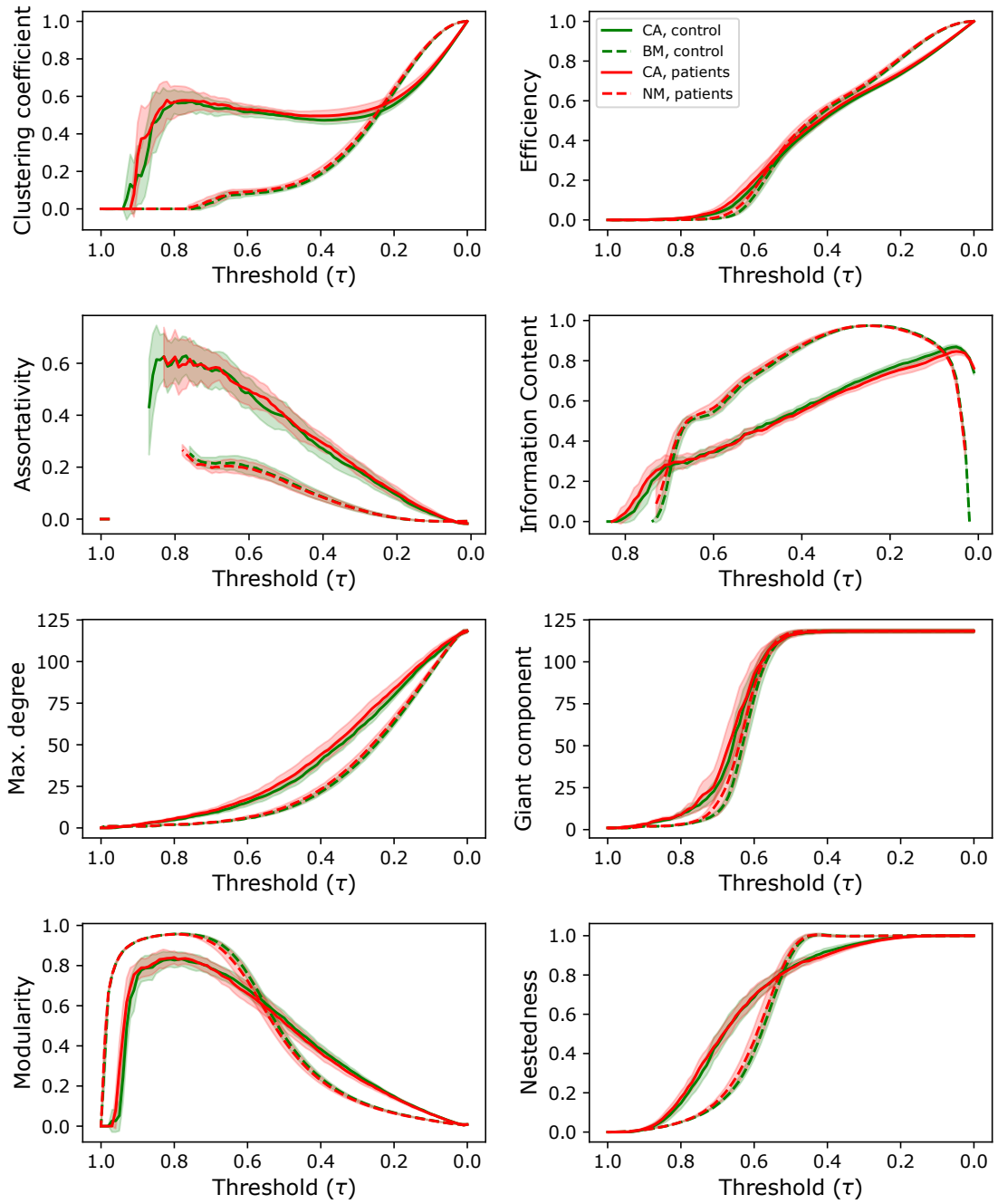
Слика А-17: ABIDE, AAL 1mm, τ , контролни субјекти vs. пациенти.

ABIDE: AAL 1mm, l_d контролни субј. vs. пациенти



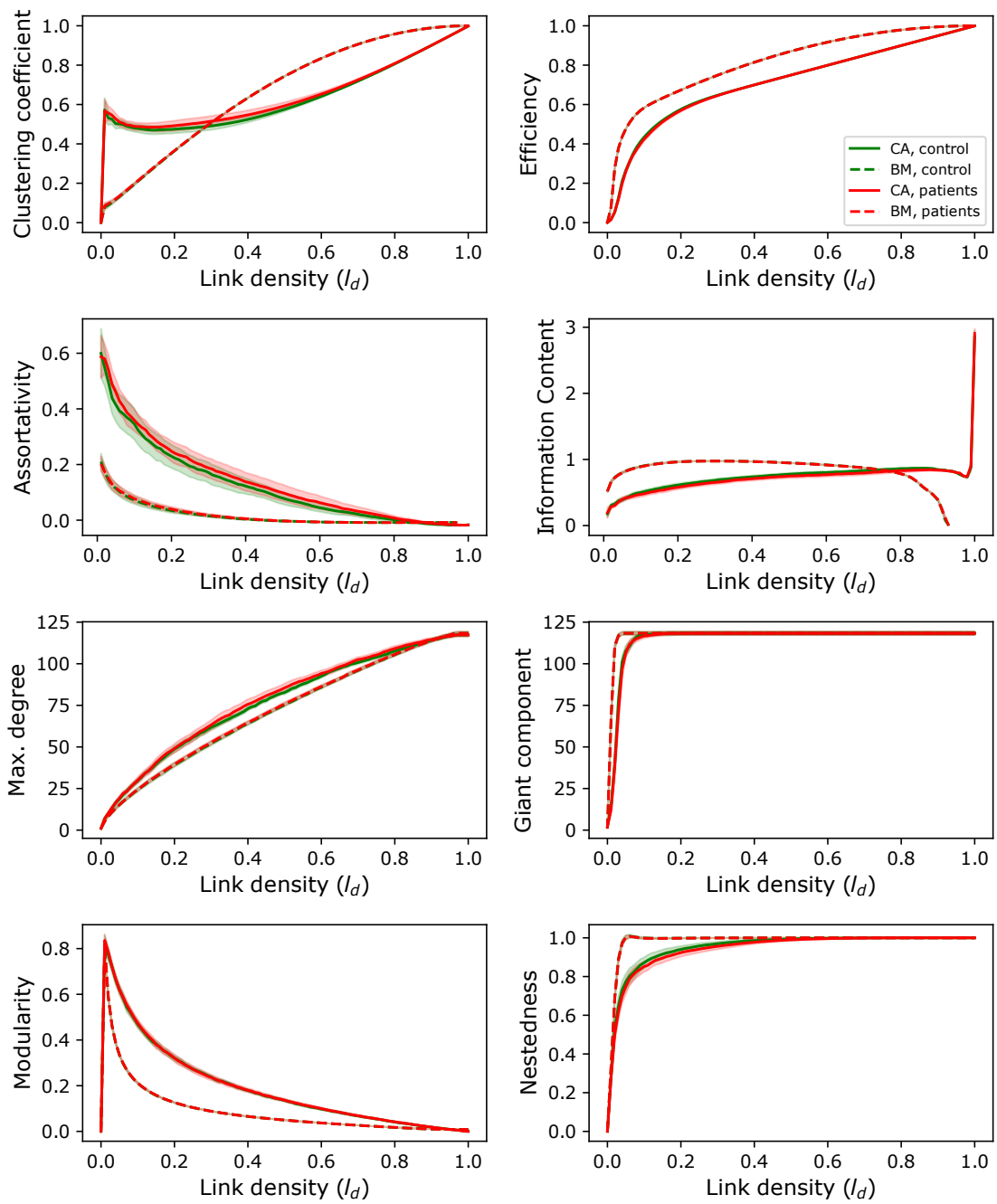
Слика А-18: ABIDE, AAL 1mm, l_d , контролни субјекти vs. пациенти.

ABIDE: AAL 2mm, τ , контролни субј. vs. пациенти



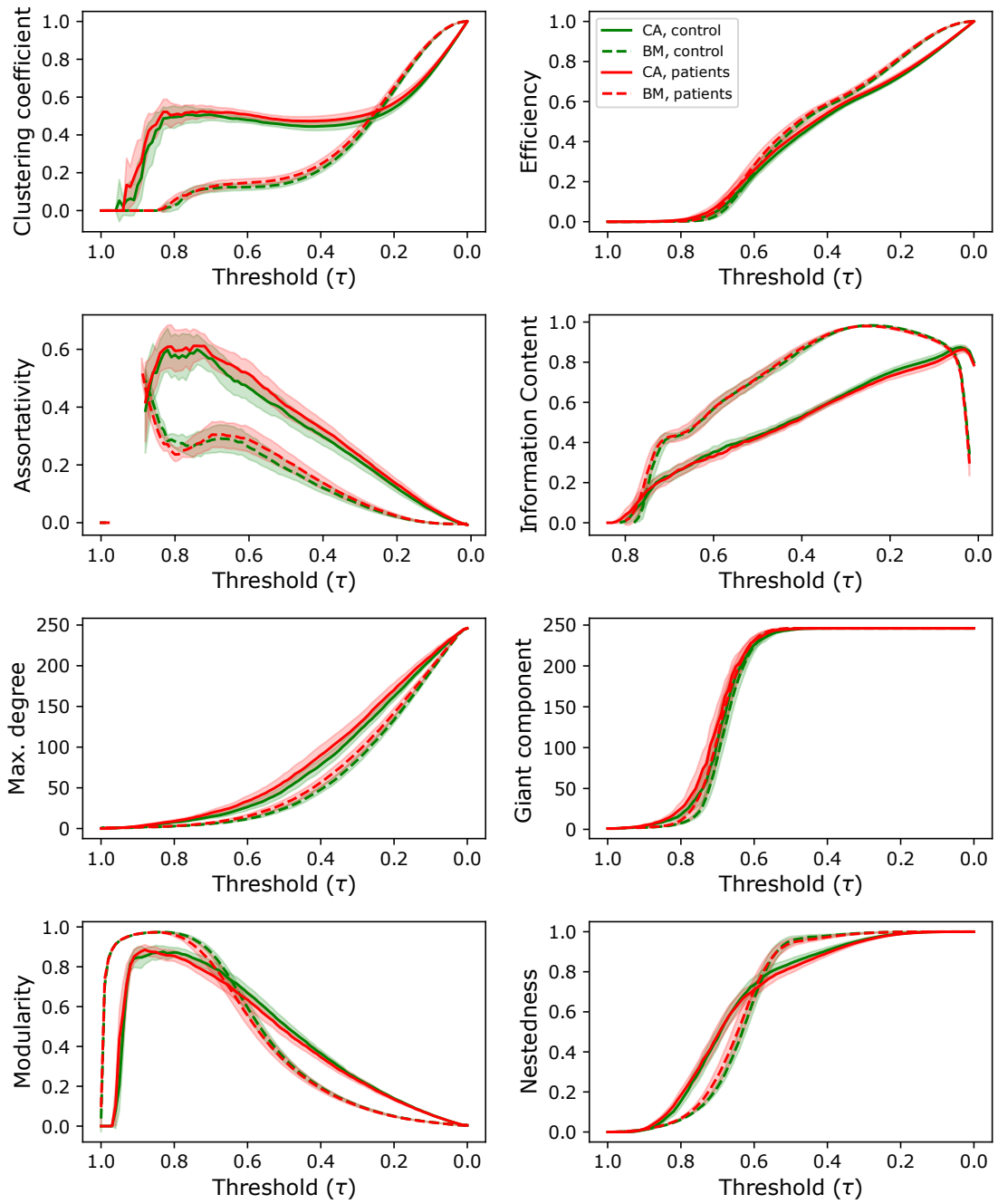
Слика А-19: ABIDE, AAL 2mm, τ , контролни субјекти vs. пациенти.

ABIDE: AAL 2mm, l_d контролни субј. vs. пациенти



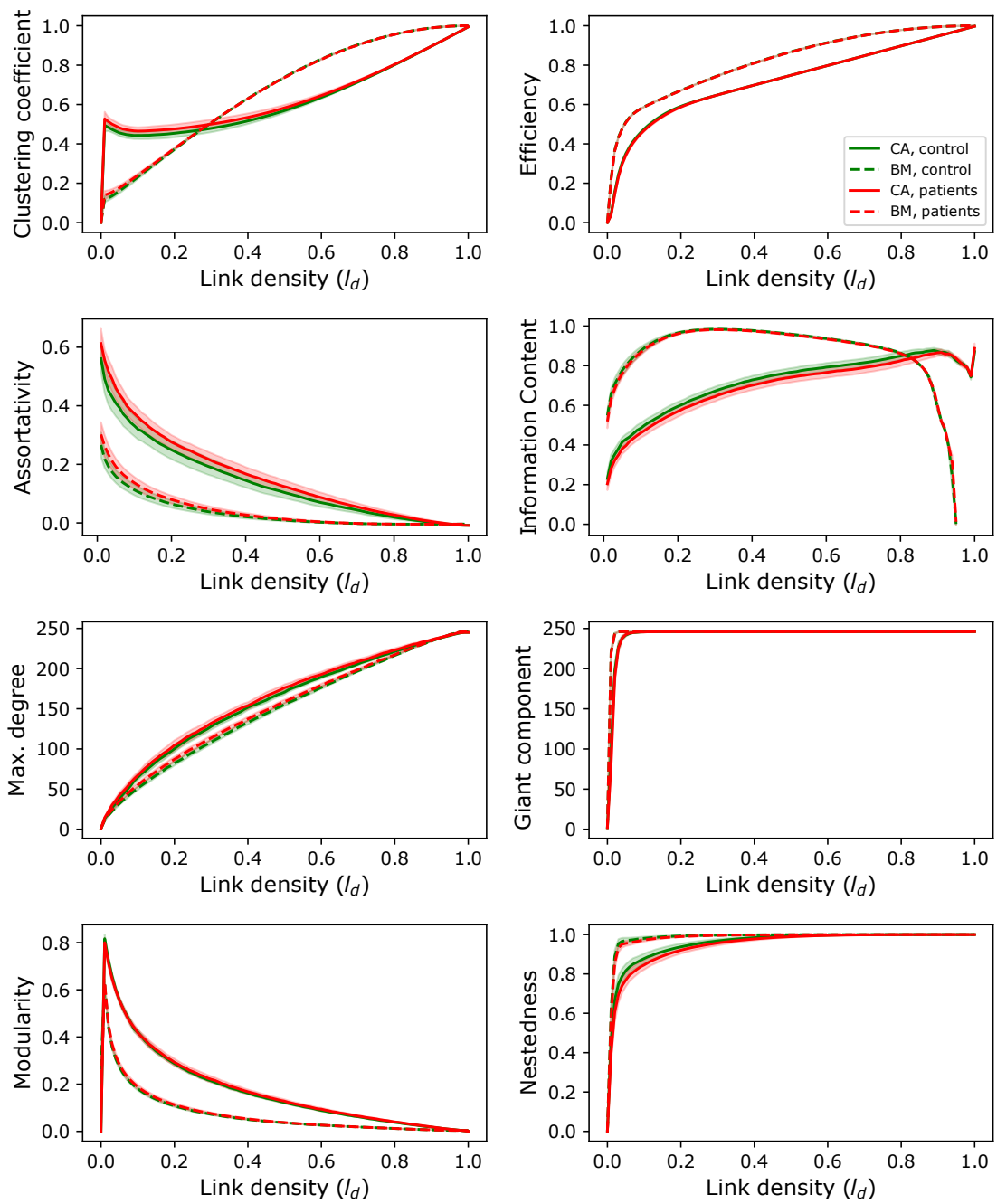
Слика А-20: ABIDE, AAL 2mm, l_d , контролни субјекти vs. пациенти.

ABIDE: Brainnetome, τ , контролни субј. vs. пациенти



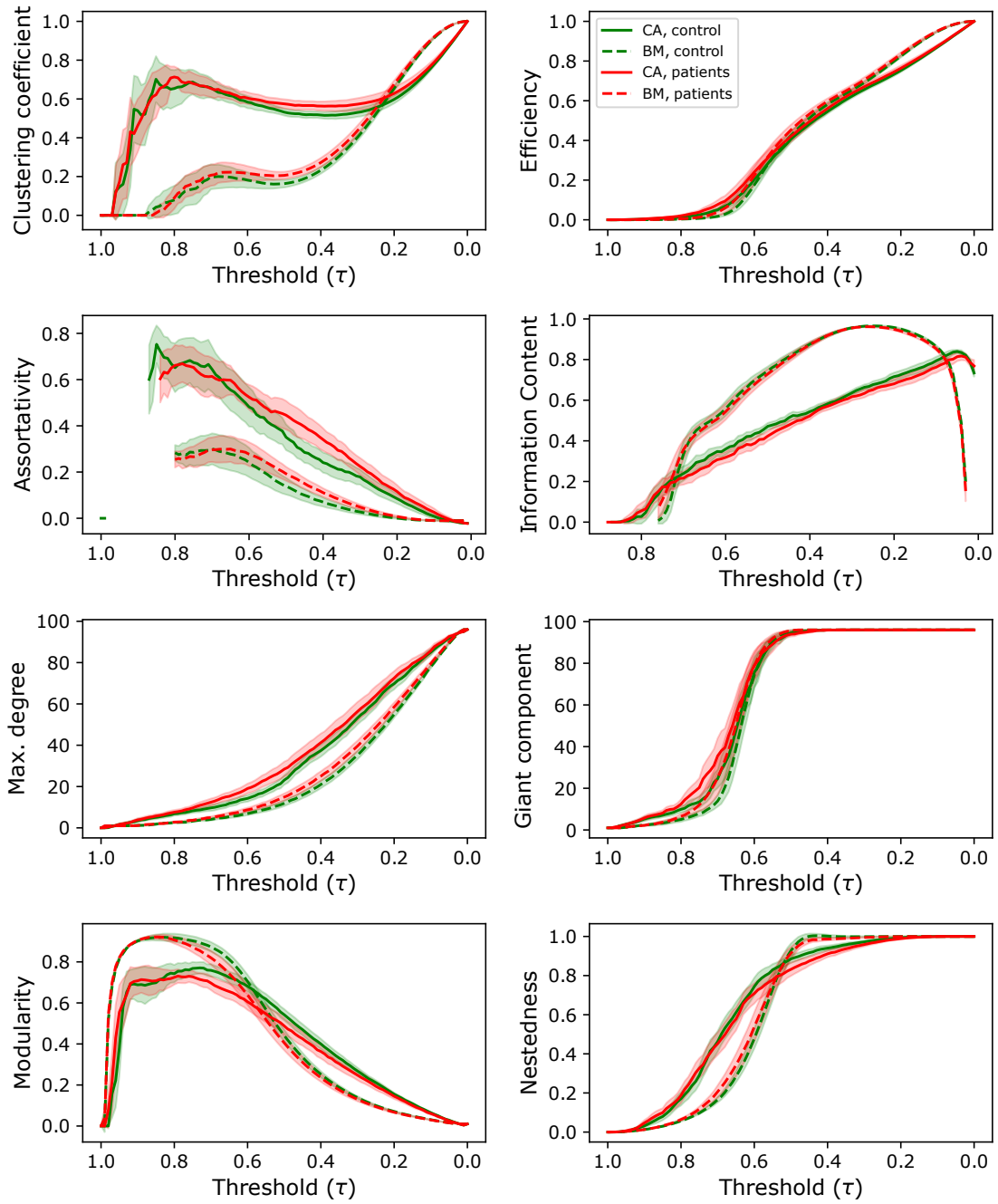
Слика А-21: ABIDE, Brainnetome, τ , контролни субјекти vs. пациенти.

ABIDE: Brainnetome, l_d контролни субј. vs. пациенти



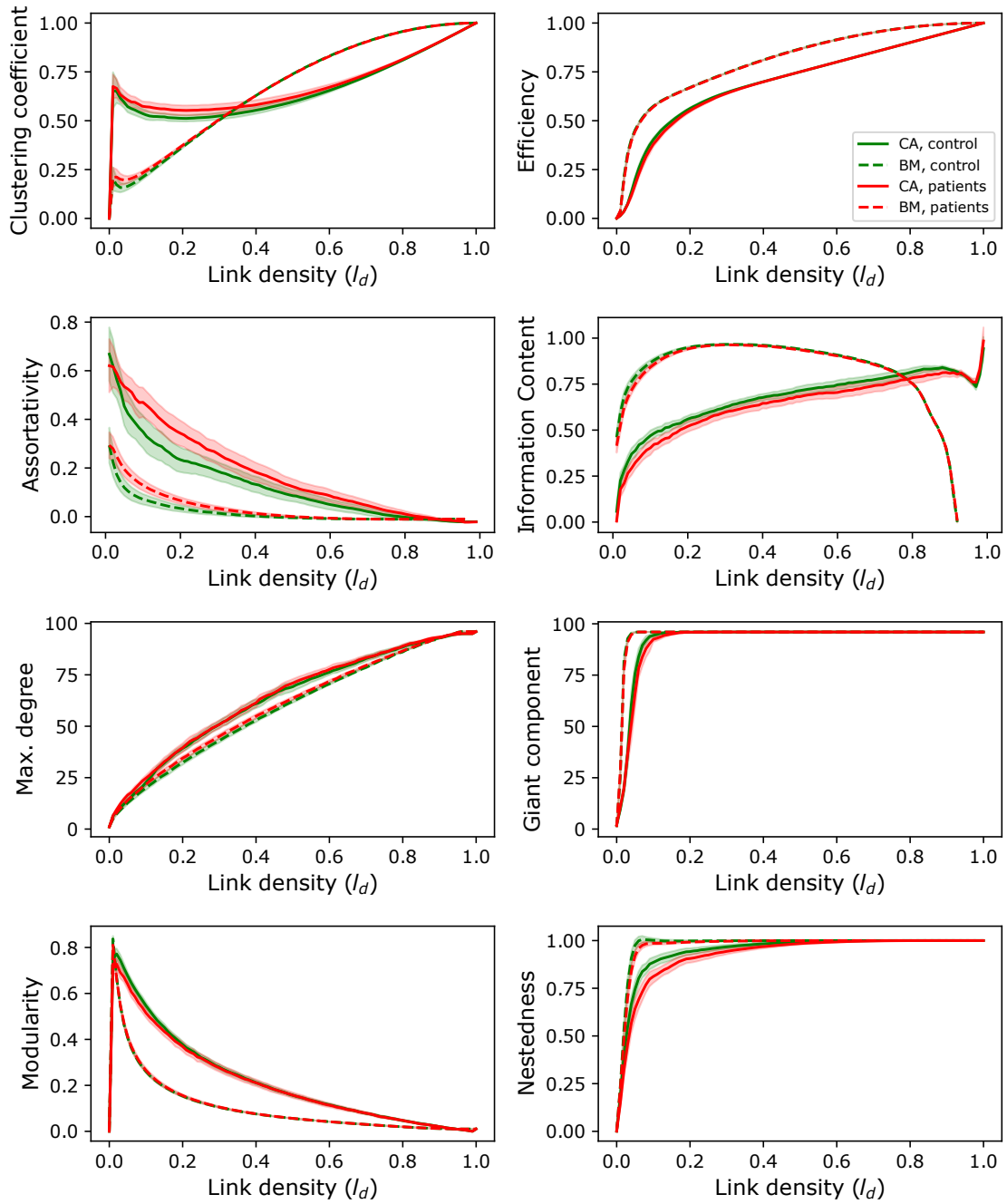
Слика А-22: ABIDE, Brainnetome, l_d , контролни субјекти vs. пациенти.

ABIDE: HO, τ , контролни субј. vs. пациенти



Слика А-23: ABIDE, HO, τ , контролни субјекти vs. пациенти.

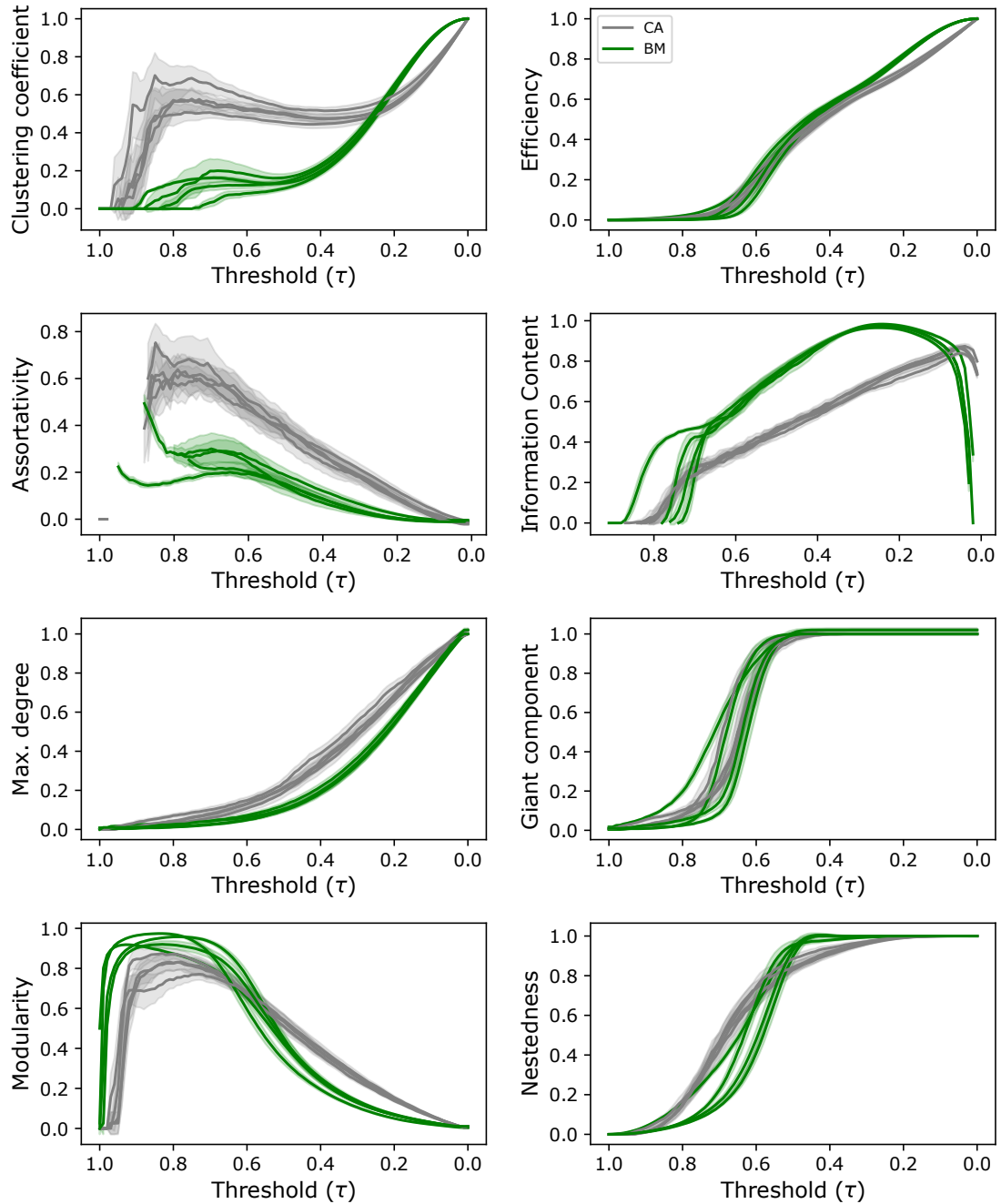
ABIDE: HO, l_d , контролни субј. vs. пациенти



Слика А-24: ABIDE, HO, l_d , контролни субјекти vs. пациенти.

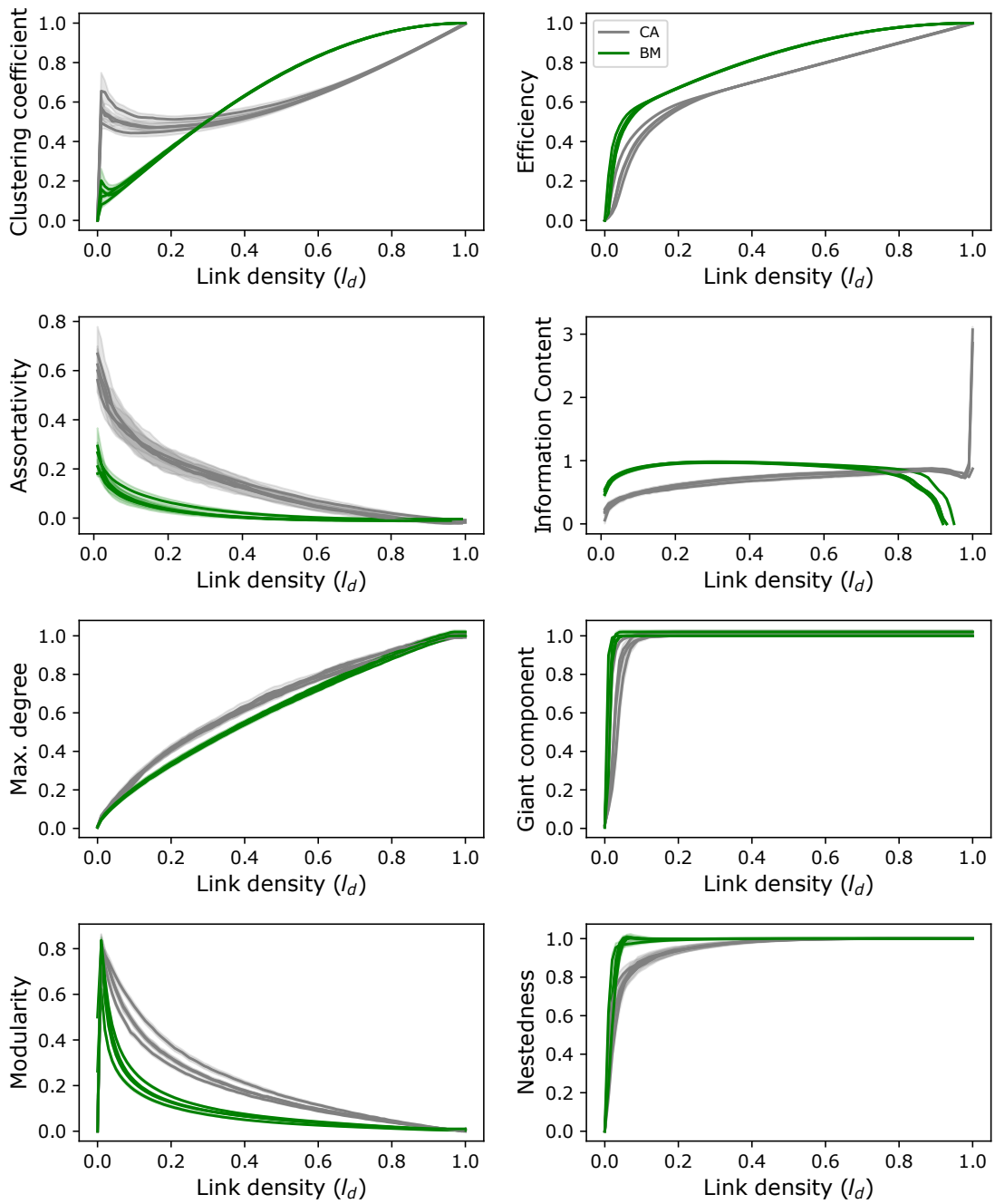
A.3 ABIDE: Споредба на различни атласи

ABIDE: Сите атласи, τ , контролни субјекти



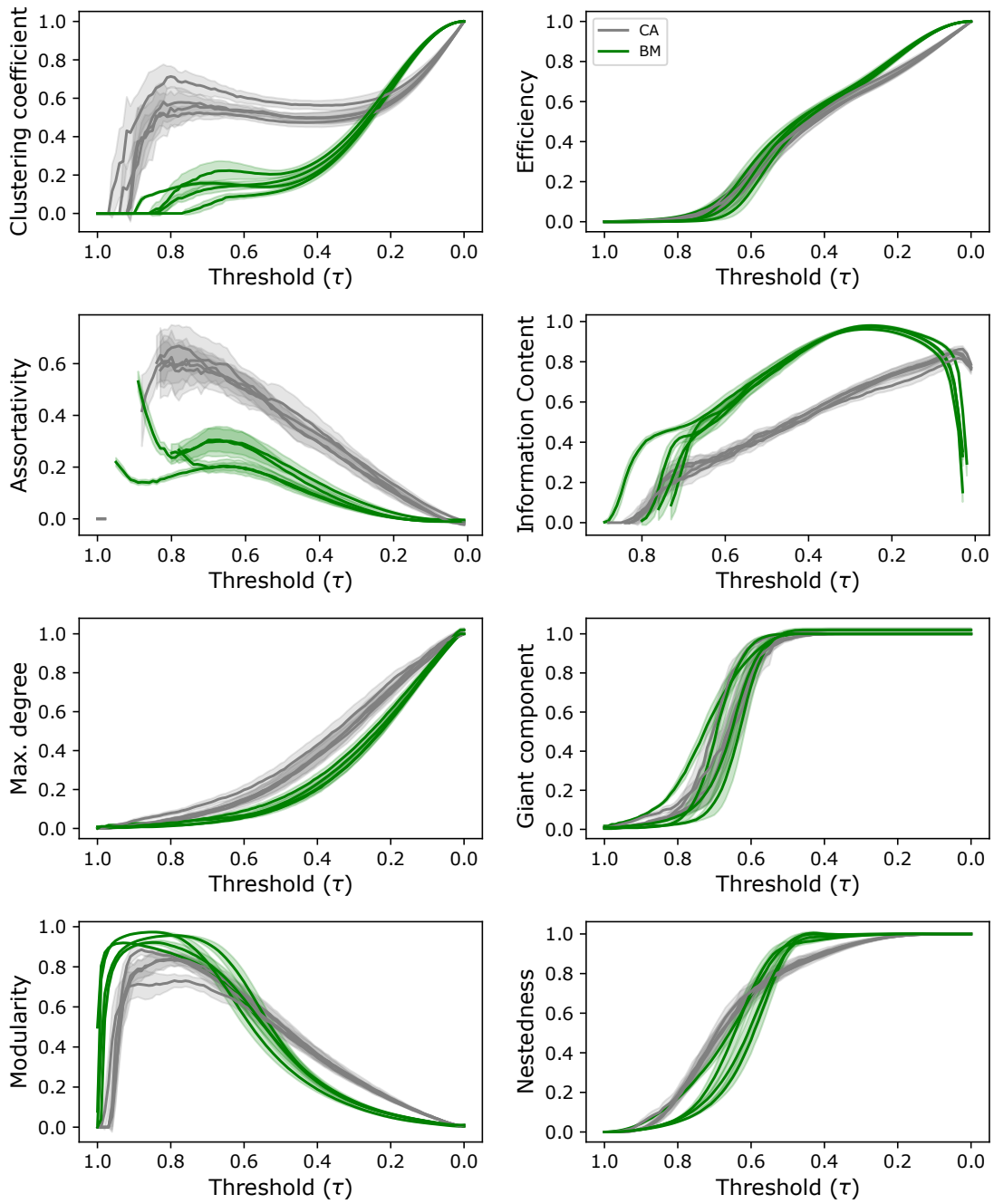
Слика А-25: ABIDE, Сите атласи, τ , контролни субјекти.

ABIDE: Сите атласи, l_d , контролни субјекти



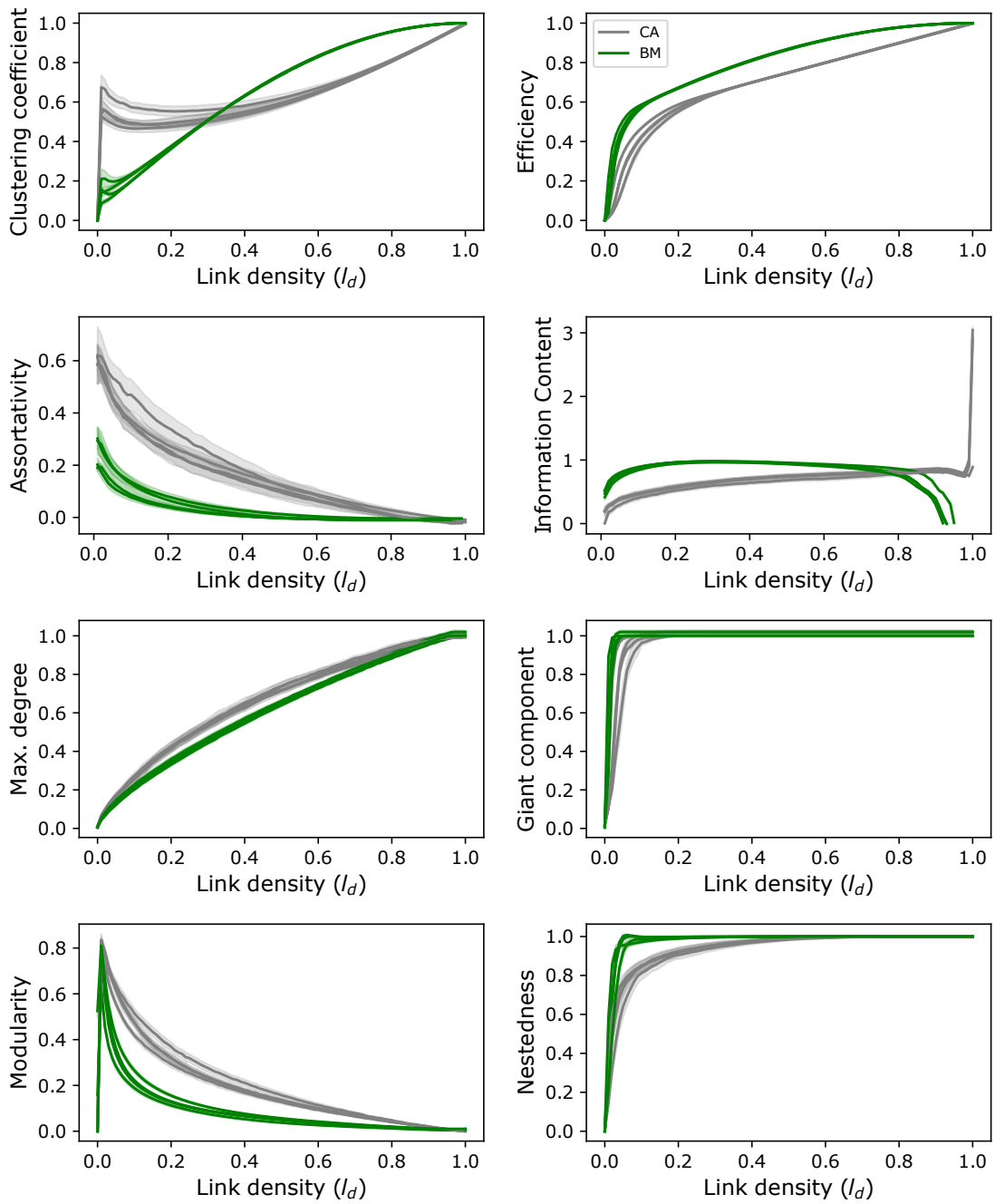
Слика А-26: ABIDE, Сите атласи, l_d , контролни субјекти.

ABIDE: Сите атласи, τ , пациенти



Слика А-27: ABIDE, Сите атласи, τ , пациенти.

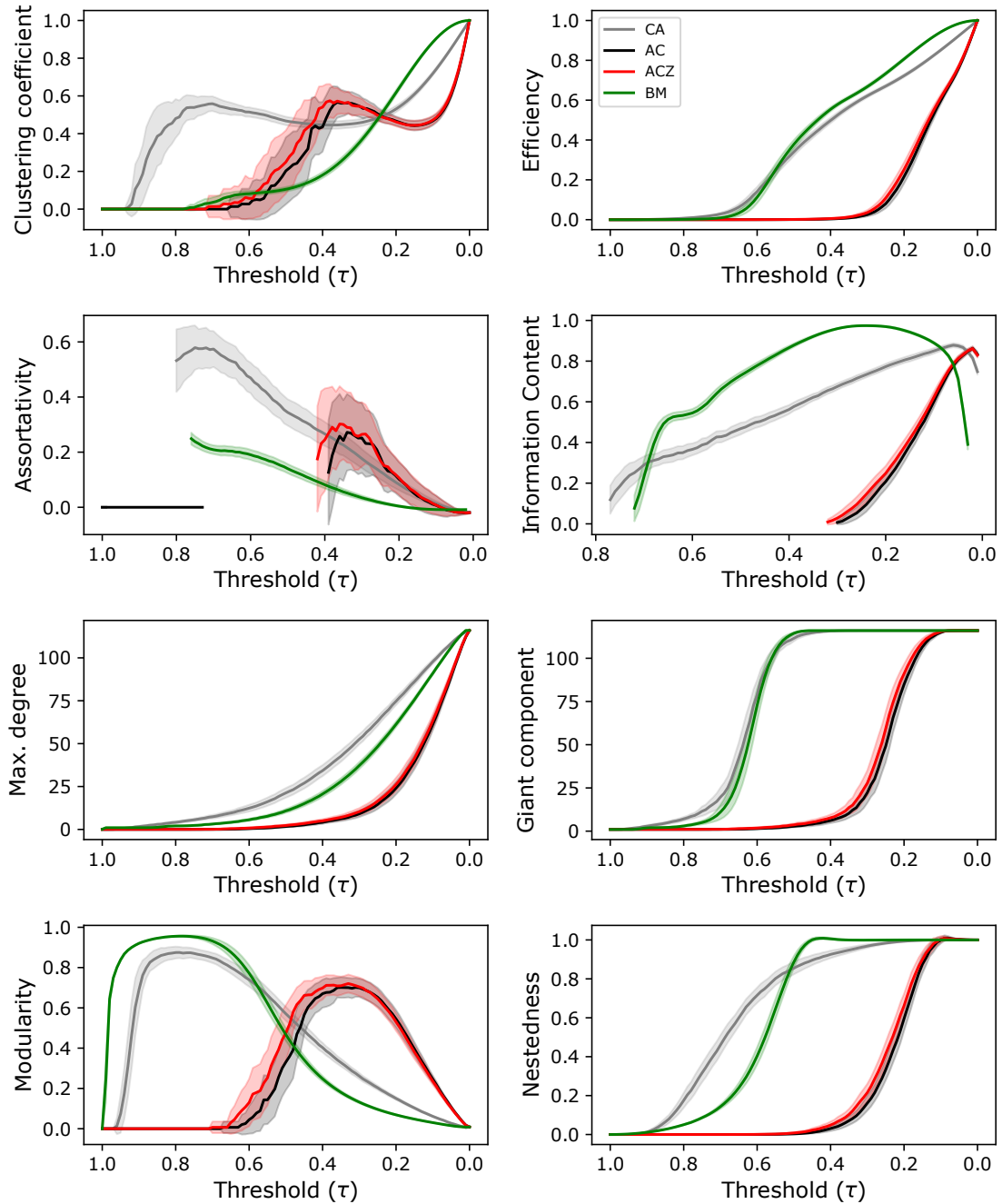
ABIDE: Сите атласи, l_d , пациенти



Слика А-28: ABIDE, Сите атласи, l_d , пациенти.

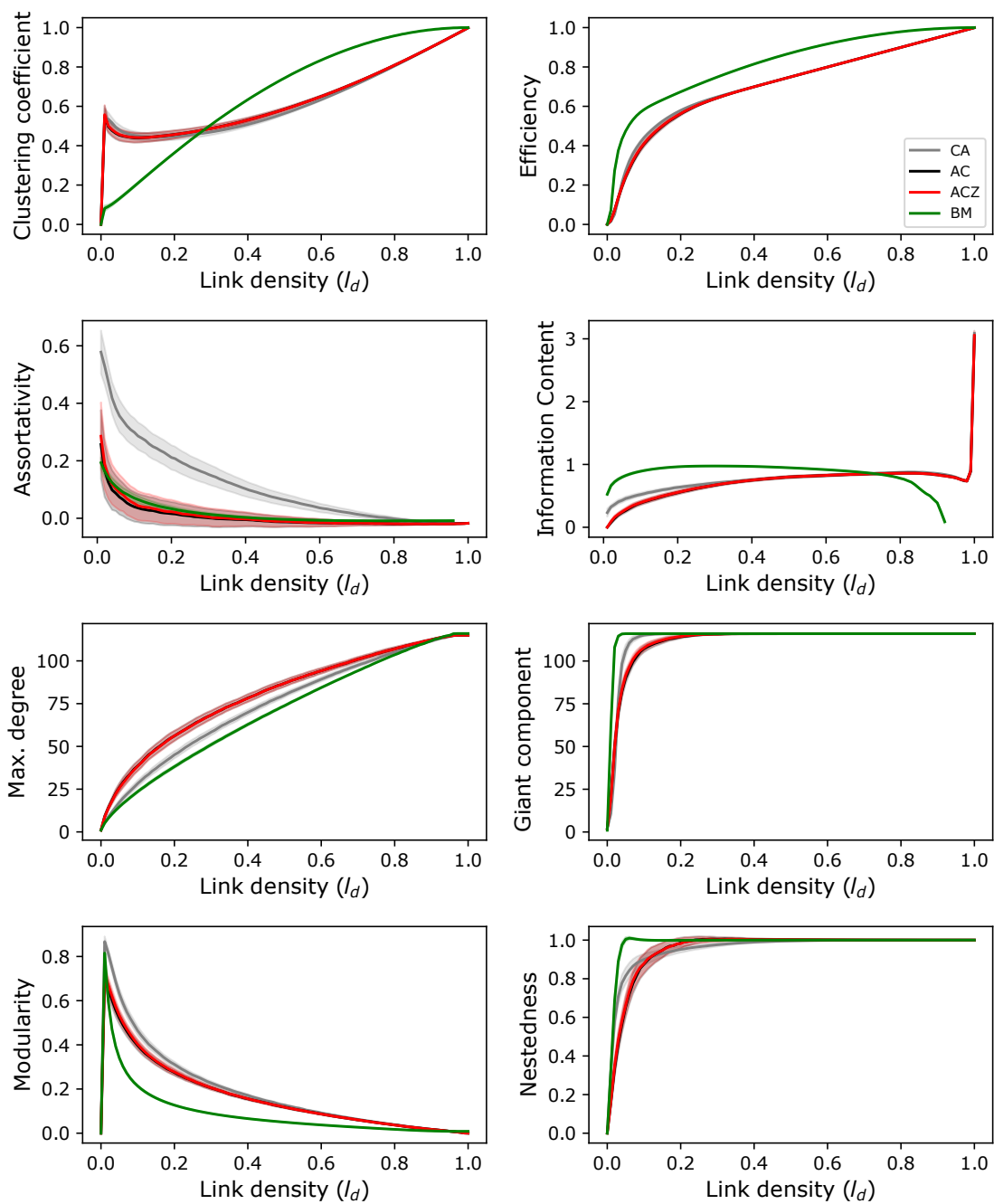
A.4 COBRE: Сите резултати

COBRE: AAL 2mm, τ , контролни субјекти



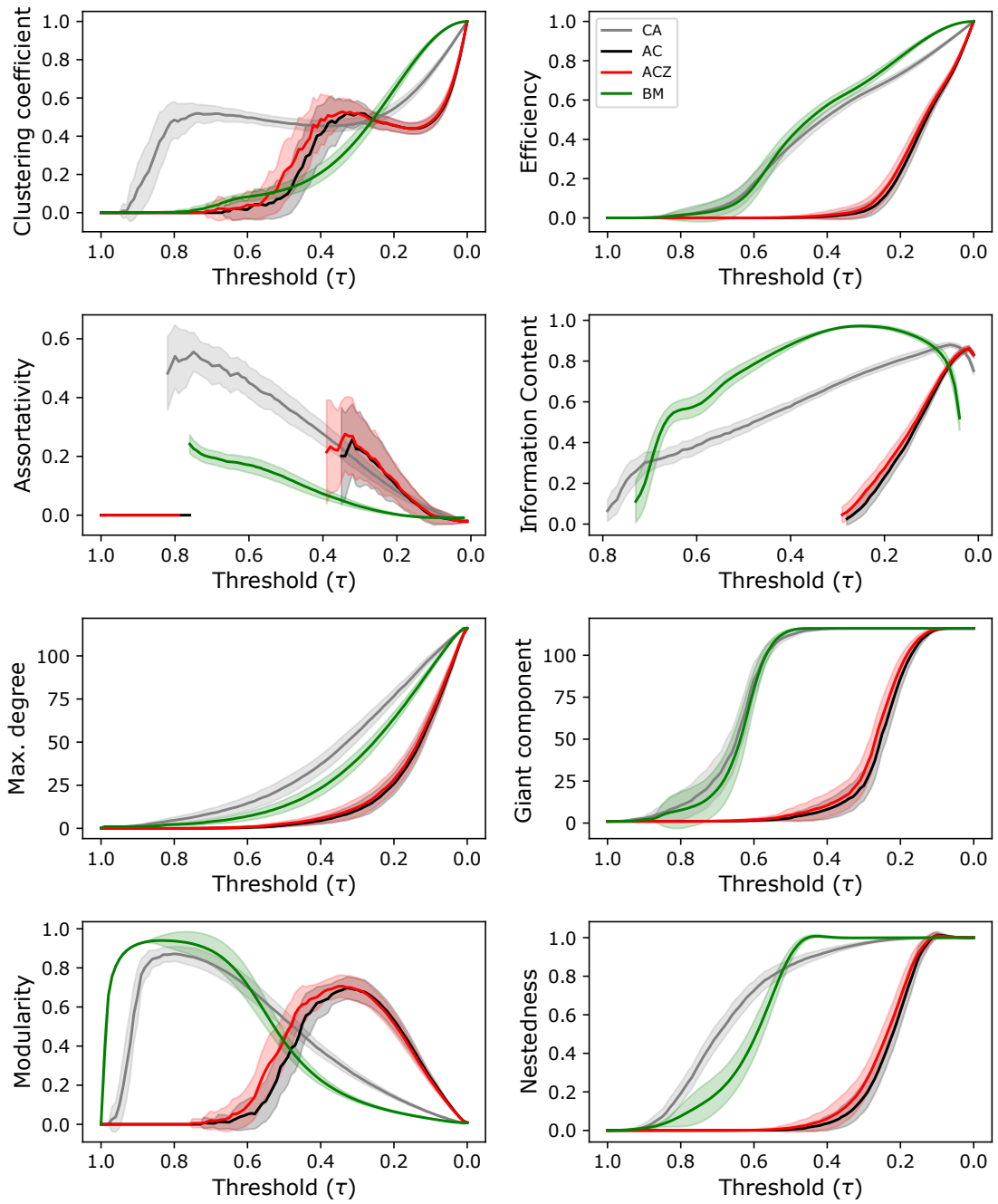
Слика А-29: COBRE, AAL 2mm, τ , контролни субјекти.

COBRE: AAL 2mm, l_d контролни субјекти



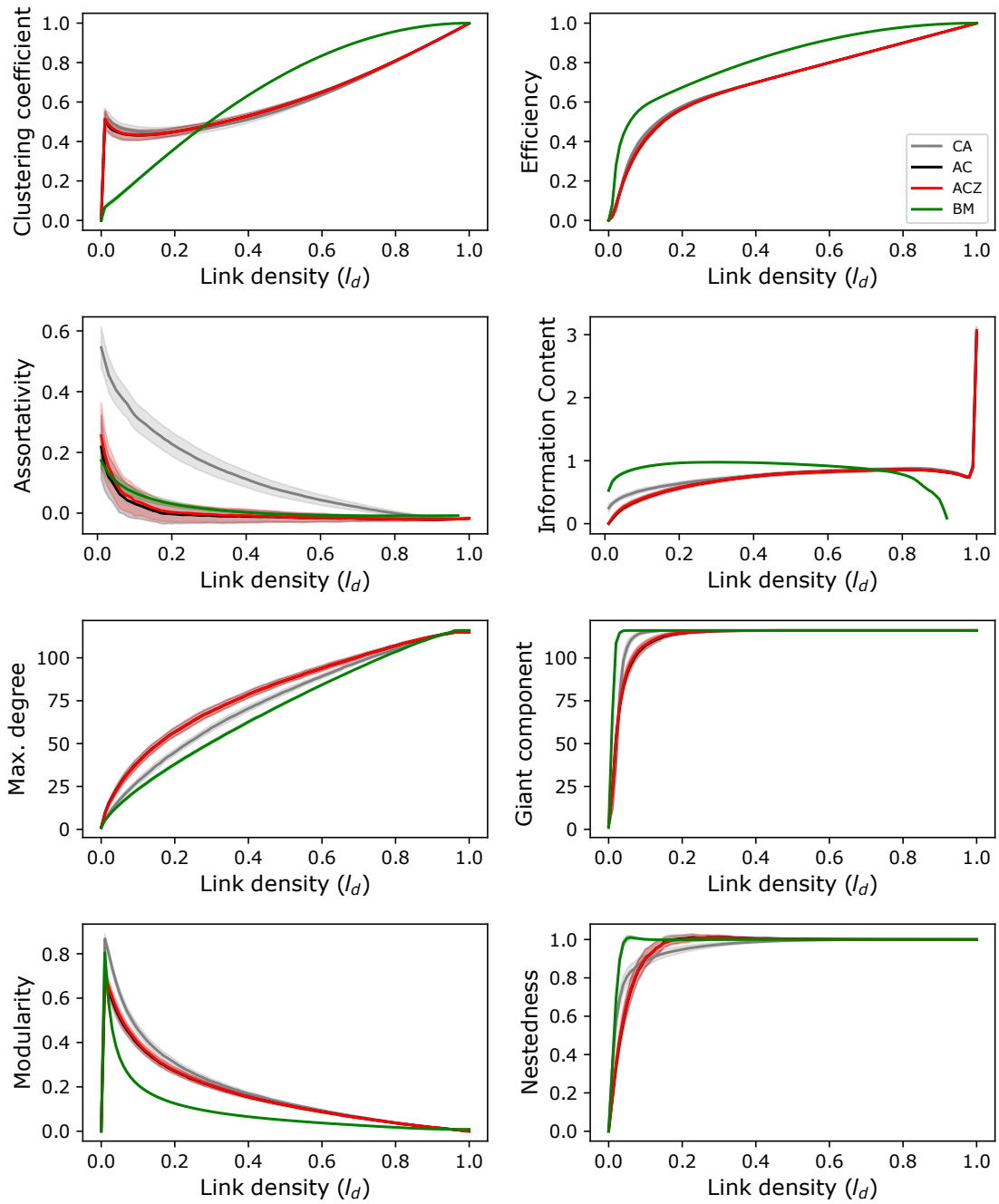
Слика А-30: COBRE, AAL 2mm, l_d контролни субјекти.

COBRE: AAL 2mm, τ , пациенти



Слика А-31: COBRE, AAL 2mm, τ , пациенти.

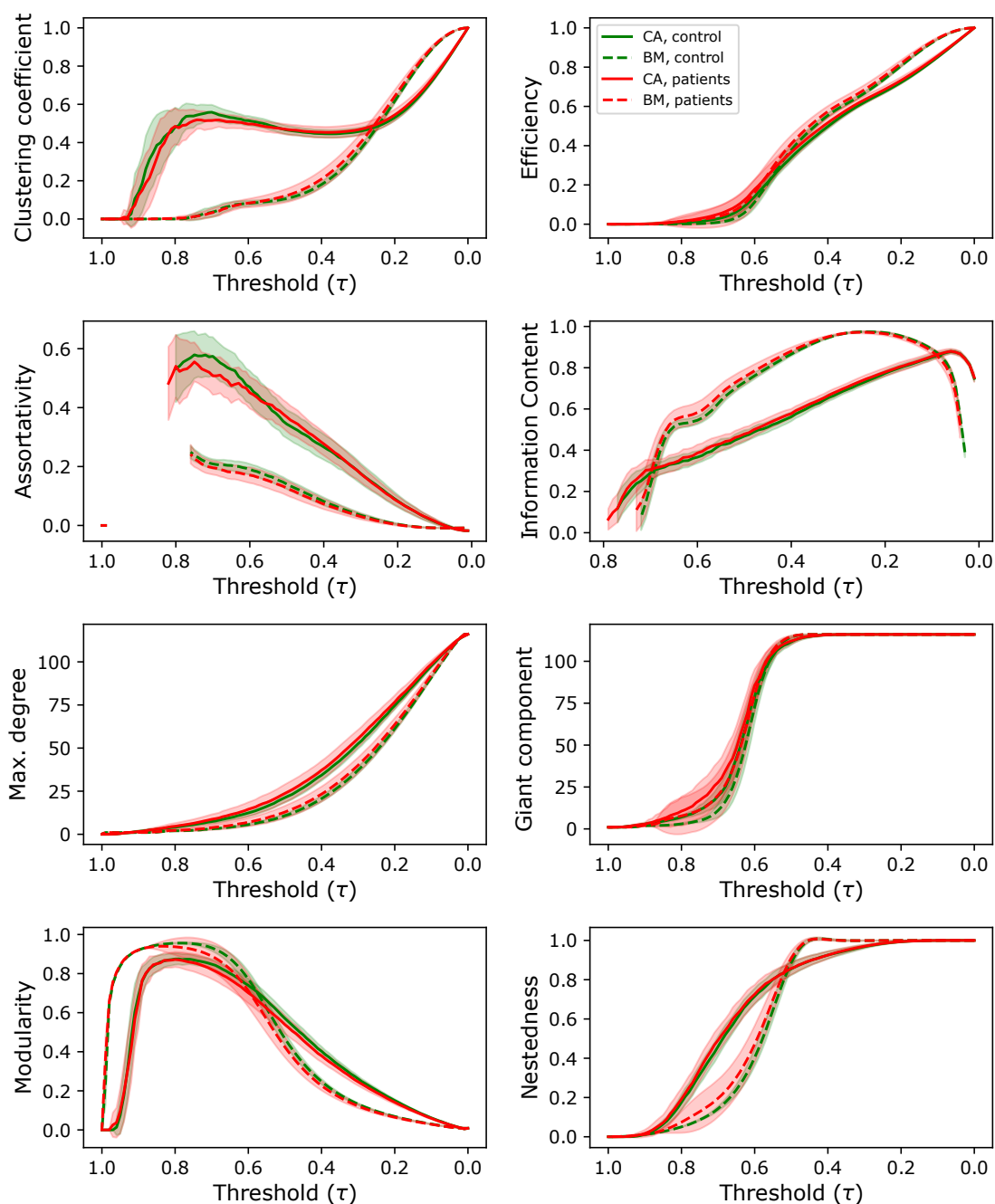
COBRE: AAL 2mm, l_d , пациенти



Слика А-32: COBRE, AAL 2mm, l_d , пациенти.

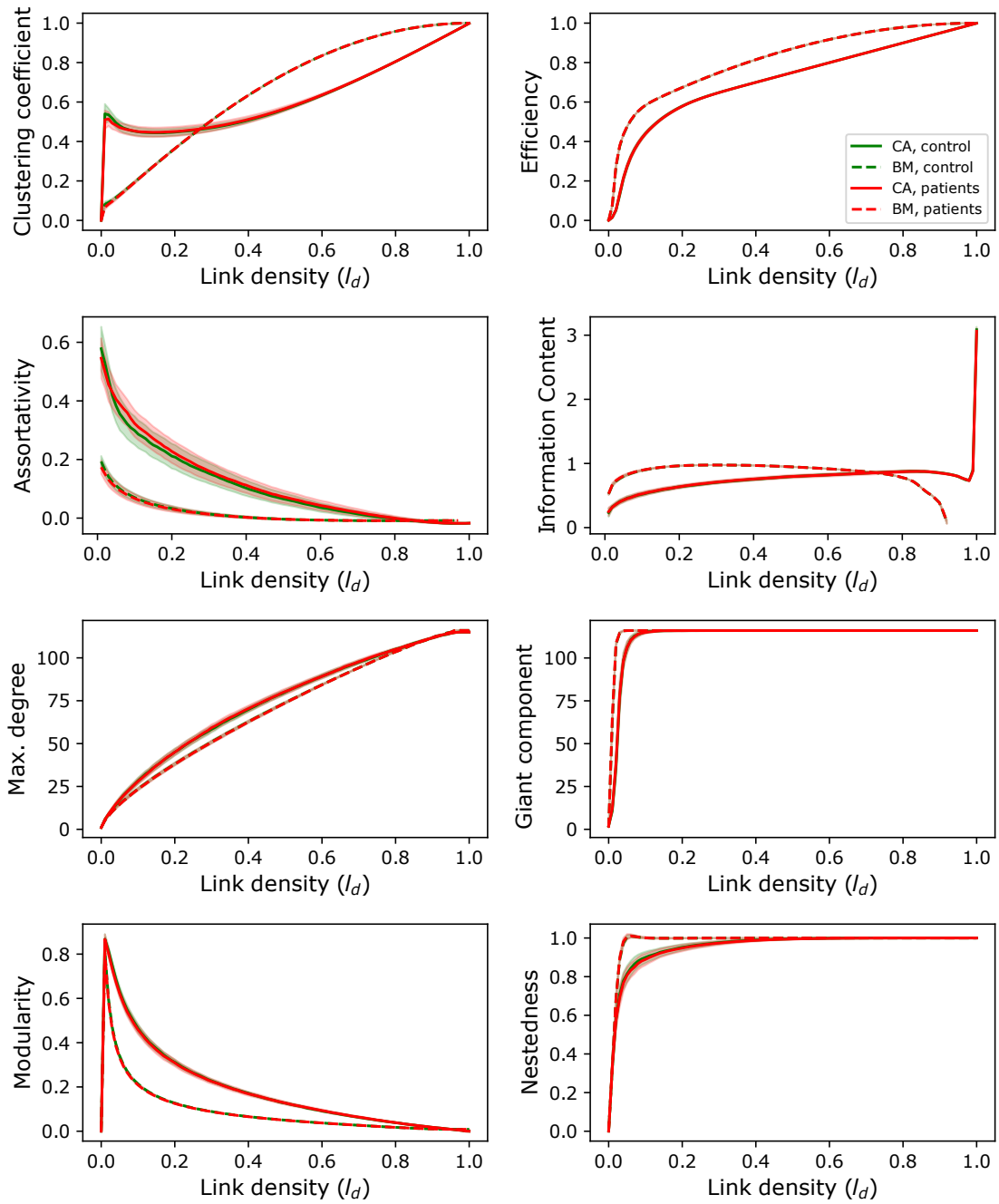
A.5 COBRE: Споредба контролни субјекти vs. пациенти

COBRE: AAL 2mm, τ , контролни субј. vs. пациенти



Слика А-33: COBRE, AAL 2mm, τ , контролни субјекти vs. пациенти.

COBRE: AAL 2mm, l_d контролни субј. vs. пациенти



Слика А-34: COBRE, AAL 2mm, l_d , контролни субјекти vs. пациенти.

Табела А.2: Листа на субјекти од AVIDE II податочното множество влучени овде.

ID	Група	Возраст	ID	Група	Возраст
29402	Контролен субјект	8.471232877	29411	Пациент	11.69041096
29405	Контролен субјект	10.19726027	29417	Пациент	8.624657534
29427	Контролен субјект	11.02465753	29430	Пациент	11.76438356
29431	Контролен субјект	12.66575342	29433	Пациент	9.093150685
29469	Контролен субјект	8.060273973	29435	Пациент	11.0109589
29303	Контролен субјект	8.556164384	29273	Пациент	8.476712329
29305	Контролен субјект	9.578082192	29276	Пациент	9.216438356
29307	Контролен субјект	9.120547945	29281	Пациент	9.331506849
29310	Контролен субјект	8.024657534	29289	Пациент	10.54246575
29329	Контролен субјект	11.72876712	29291	Пациент	11.63835616
29334	Контролен субјект	8.115068493	29292	Пациент	11.14794521
29337	Контролен субјект	12.36164384	29375	Пациент	12.60547945
29345	Контролен субјект	11.02739726	29385	Пациент	12.02465753
29346	Контролен субјект	12.66027397	29391	Пациент	8.01369863
29356	Контролен субјект	11.72876712	29393	Пациент	10.78630137
29372	Контролен субјект	11.75890411	29394	Пациент	10.8109589
29381	Контролен субјект	10.35342466	29401	Пациент	8.463013699
29388	Контролен субјект	10.59726027	29408	Пациент	12.2
29446	Контролен субјект	9.167123288	29434	Пациент	8.057534247

Библиографија

- [1] O. Sporns, *Networks of the Brain*. MIT press, 2010.
- [2] —, *Discovering the human connectome*. MIT press, 2012.
- [3] X.-N. Zuo, A. Di Martino, C. Kelly, Z. E. Shehzad, D. G. Gee, D. F. Klein, F. X. Castellanos, B. B. Biswal, and M. P. Milham, “The oscillating brain: complex and reliable,” *Neuroimage*, vol. 49, no. 2, pp. 1432–1445, 2010.
- [4] S. A. Bunge and I. Kahn, “Cognition: An overview of neuroimaging techniques,” 2009.
- [5] R. Nusslock, C. B. Young, N. Pornpattananankul, and K. S. Damme, “Neurophysiological and neuroimaging techniques,” *The Encyclopedia of Clinical Psychology*, pp. 1–9, 2014.
- [6] E. T. Bullmore and D. S. Bassett, “Brain graphs: graphical models of the human brain connectome,” *Annual review of clinical psychology*, vol. 7, pp. 113–140, 2011.
- [7] K. Trivodaliev, A. Bogojeska, and L. Kocarev, “Exploring function prediction in protein interaction networks via clustering methods,” *PloS one*, vol. 9, no. 6, p. e99755, 2014.
- [8] R. Navigli and S. P. Ponzetto, “Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network,” *Artificial intelligence*, vol. 193, pp. 217–250, 2012.
- [9] N. Ueffing, F. J. Och, and H. Ney, “Generation of word graphs in statistical machine translation,” in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, 2002, pp. 156–163.
- [10] P. J. Carrington, J. Scott, and S. Wasserman, *Models and methods in social network analysis*. Cambridge university press, 2005, vol. 28.
- [11] M. E. Newman, “The structure and function of complex networks,” *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.

- [12] J. A. Sonquist, “Communication networks: Toward a new paradigm for research.” 1984.
- [13] D. J. Cook and L. B. Holder, *Mining graph data*. John Wiley & Sons, 2006.
- [14] M. Zanin, D. Papo, P. A. Sousa, E. Menasalvas, A. Nicchi, E. Kubik, and S. Boccaletti, “Combining complex networks and data mining: why and how,” *Physics Reports*, vol. 635, pp. 1–44, 2016.
- [15] T. Takahashi, “Complexity of spontaneous brain activity in mental disorders,” *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 45, pp. 258–266, 2013.
- [16] A. C. Yang and S.-J. Tsai, “Complexity of mental illness: a new research dimension,” 2013.
- [17] F. D. Bowman, Y. Guo, and G. Derado, “Statistical approaches to functional neuroimaging data,” *Neuroimaging Clinics of North America*, vol. 17, no. 4, pp. 441–458, 2007.
- [18] E. Bullmore and O. Sporns, “Complex brain networks: graph theoretical analysis of structural and functional systems,” *Nature reviews neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.
- [19] A. Fornito, A. Zalesky, and E. Bullmore, *Fundamentals of brain network analysis*. Academic Press, 2016.
- [20] M. P. Van Den Heuvel and H. E. H. Pol, “Exploring the brain network: a review on resting-state fmri functional connectivity,” *European neuropsychopharmacology*, vol. 20, no. 8, pp. 519–534, 2010.
- [21] N. K. Logothetis and B. A. Wandell, “Interpreting the bold signal,” *Annu. Rev. Physiol.*, vol. 66, pp. 735–769, 2004.
- [22] A. Di Martino, A. Scheres, D. S. Margulies, A. Kelly, L. Q. Uddin, Z. Shehzad, B. Biswal, J. R. Walters, F. X. Castellanos, and M. P. Milham, “Functional connectivity of human striatum: a resting state fmri study,” *Cerebral cortex*, vol. 18, no. 12, pp. 2735–2747, 2008.
- [23] M. L. Schölvinck, A. Maier, Q. Y. Frank, J. H. Duyn, and D. A. Leopold, “Neural basis of global resting-state fmri activity,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 22, pp. 10 238–10 243, 2010.
- [24] Y. Chao-Gan and Z. Yu-Feng, “Dparsf: a matlab toolbox for “pipeline” data analysis of resting-state fmri,” *Frontiers in systems neuroscience*, vol. 4, 2010.

- [25] X.-W. Song, Z.-Y. Dong, X.-Y. Long, S.-F. Li, X.-N. Zuo, C.-Z. Zhu, Y. He, C.-G. Yan, and Y.-F. Zang, “Rest: a toolkit for resting-state functional magnetic resonance imaging data processing,” *PLoS one*, vol. 6, no. 9, p. e25031, 2011.
- [26] C. Sestieri, M. Corbetta, G. L. Romani, and G. L. Shulman, “Episodic memory retrieval, parietal cortex, and the default mode network: functional and topographic analyses,” *Journal of Neuroscience*, vol. 31, no. 12, pp. 4407–4420, 2011.
- [27] D. T. Jones, M. M. Machulda, P. Vemuri, E. McDade, G. Zeng, M. Senjem, J. Gunter, S. Przybelski, R. Avula, D. S. Knopman *et al.*, “Age-related changes in the default mode network are more advanced in alzheimer disease,” *Neurology*, vol. 77, no. 16, pp. 1524–1531, 2011.
- [28] D. Öngür, M. Lundy, I. Greenhouse, A. K. Shinn, V. Menon, B. M. Cohen, and P. F. Renshaw, “Default mode network abnormalities in bipolar disorder and schizophrenia,” *Psychiatry Research: Neuroimaging*, vol. 183, no. 1, pp. 59–68, 2010.
- [29] X.-Y. Long, X.-N. Zuo, V. Kiviniemi, Y. Yang, Q.-H. Zou, C.-Z. Zhu, T.-Z. Jiang, H. Yang, Q.-Y. Gong, L. Wang *et al.*, “Default mode network as revealed with multiple methods for resting-state functional mri analysis,” *Journal of neuroscience methods*, vol. 171, no. 2, pp. 349–355, 2008.
- [30] R. Kaplan, “Endogenous fmri default mode network fluctuations both positively and negatively correlate with individual transfer of learning,” *Frontiers in systems neuroscience*, vol. 8, p. 229, 2014.
- [31] Y. I. Sheline, D. M. Barch, J. L. Price, M. M. Rundle, S. N. Vaishnavi, A. Z. Snyder, M. A. Mintun, S. Wang, R. S. Coalson, and M. E. Raichle, “The default mode network and self-referential processes in depression,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 6, pp. 1942–1947, 2009.
- [32] W. Koch, S. Teipel, S. Mueller, J. Benninghoff, M. Wagner, A. L. Bokde, H. Hampel, U. Coates, M. Reiser, and T. Meindl, “Diagnostic power of default mode network resting state fmri in the detection of alzheimer’s disease,” *Neurobiology of aging*, vol. 33, no. 3, pp. 466–478, 2012.
- [33] A. Padmanabhan, C. J. Lynch, M. Schaer, and V. Menon, “The default mode network in autism,” *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 2, no. 6, pp. 476–486, 2017.
- [34] M.-L. Hu, X.-F. Zong, J. J. Mann, J.-J. Zheng, Y.-H. Liao, Z.-C. Li, Y. He, X.-G. Chen, and J.-S. Tang, “A review of the functional and anatomical default

- mode network in schizophrenia,” *Neuroscience bulletin*, vol. 33, no. 1, pp. 73–84, 2017.
- [35] X. Zhu, X. Wang, J. Xiao, J. Liao, M. Zhong, W. Wang, and S. Yao, “Evidence of a dissociation pattern in resting-state default mode network connectivity in first-episode, treatment-naive major depression patients,” *Biological psychiatry*, vol. 71, no. 7, pp. 611–617, 2012.
- [36] M. H. Lee, C. D. Hacker, A. Z. Snyder, M. Corbetta, D. Zhang, E. C. Leuthardt, and J. S. Shimony, “Clustering of resting state networks,” *PloS one*, vol. 7, no. 7, p. e40370, 2012.
- [37] Y. Wang and T.-Q. Li, “Analysis of whole-brain resting-state fmri data using hierarchical clustering approach,” *PloS one*, vol. 8, no. 10, p. e76315, 2013.
- [38] S. Teng, P. Wang, Y. Liao, T.-C. Yeh, T.-P. Su, J. Hsieh, and Y. Wu, “Small-world network for investigating functional connectivity in bipolar disorder: A functional magnetic images (fmri) study,” in *13th International Conference on Biomedical Engineering*. Springer, 2009, pp. 726–729.
- [39] D. S. Bassett and E. T. Bullmore, “Small-world brain networks revisited,” *The Neuroscientist*, vol. 23, no. 5, pp. 499–516, 2017.
- [40] J. Richiardi, S. Achard, H. Bunke, and D. Van De Ville, “Machine learning with brain graphs: predictive modeling approaches for functional imaging in systems neuroscience,” *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 58–70, 2013.
- [41] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, H. J. Johnson, J. S. Paulsen, J. A. Turner, and V. D. Calhoun, “Deep learning for neuroimaging: a validation study,” *Frontiers in neuroscience*, vol. 8, p. 229, 2014.
- [42] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh, “Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment,” *NeuroImage*, vol. 146, pp. 1038–1049, 2017.
- [43] R. R. Wilcox and G. A. Rousselet, “A guide to robust statistical methods in neuroscience,” *Current protocols in neuroscience*, vol. 82, no. 1, pp. 8–42, 2018.
- [44] G. Ma, L. He, B. Cao, J. Zhang, S. Y. Philip, and A. B. Ragin, “Multi-graph clustering based on interior-node topology with applications to brain networks,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2016, pp. 476–492.

- [45] A. Tenev *et al.*, “Connectomics based multi-modal graph measures in bipolar disorder,” in *Poster at 20th Annual Meeting of the Organization for Human Brain Mapping (OHBM)*, vol. 2163, 2014.
- [46] M. N. I. Qureshi, J. Oh, B. Min, H. J. Jo, and B. Lee, “Corrigendum: Multi-modal, multi-measure, and multi-class discrimination of adhd with hierarchical feature extraction and extreme learning machine using structural and functional brain mri,” *Frontiers in human neuroscience*, vol. 11, p. 292, 2017.
- [47] E. Rodriguez, N. George, J.-P. Lachaux, J. Martinerie, B. Renault, and F. J. Varela, “Perception’s shadow: long-distance synchronization of human brain activity,” *Nature*, vol. 397, no. 6718, pp. 430–433, 1999.
- [48] F. Varela, J.-P. Lachaux, E. Rodriguez, and J. Martinerie, “The brainweb: phase synchronization and large-scale integration,” *Nature reviews neuroscience*, vol. 2, no. 4, pp. 229–239, 2001.
- [49] M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon, “Functional connectivity in the resting brain: a network analysis of the default mode hypothesis,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 1, pp. 253–258, 2003.
- [50] R. F. Betzel, M. A. Erickson, M. Abell, B. F. O’Donnell, W. P. Hetrick, and O. Sporns, “Synchronization dynamics and evidence for a repertoire of network states in resting eeg,” *Frontiers in computational neuroscience*, vol. 6, p. 74, 2012.
- [51] J. F. Hipp, D. J. Hawellek, M. Corbetta, M. Siegel, and A. K. Engel, “Large-scale cortical correlation structure of spontaneous oscillatory activity,” *Nature neuroscience*, vol. 15, no. 6, p. 884, 2012.
- [52] T. Harmelech and R. Malach, “Neurocognitive biases and the patterns of spontaneous correlations in the human cortex,” *Trends in cognitive sciences*, vol. 17, no. 12, pp. 606–615, 2013.
- [53] E. Amico, A. Arenas, and J. Goñi, “Centralized and distributed cognitive task processing in the human connectome,” *Network Neuroscience*, vol. 3, no. 2, pp. 455–474, 2019.
- [54] D. S. Bassett and M. G. Mattar, “A network neuroscience of human learning: Potential to inform quantitative theories of brain and behavior,” *Trends in cognitive sciences*, vol. 21, no. 4, pp. 250–264, 2017.

- [55] M. Cao, H. Huang, and Y. He, “Developmental connectomics from infancy through early childhood,” *Trends in neurosciences*, vol. 40, no. 8, pp. 494–506, 2017.
- [56] C. J. Stam, “Modern network science of neurological disorders,” *Nature Reviews Neuroscience*, vol. 15, no. 10, pp. 683–695, 2014.
- [57] A. Fornito, A. Zalesky, and M. Breakspear, “The connectomics of brain disorders,” *Nature Reviews Neuroscience*, vol. 16, no. 3, pp. 159–172, 2015.
- [58] D. Papo, M. Zanin, and J. Martin Buldú, “Reconstructing functional brain networks: have we got the basics right?” *Frontiers in human neuroscience*, vol. 8, p. 107, 2014.
- [59] M. Zanin, P. Sousa, D. Papo, R. Bajo, J. García-Prieto, F. Del Pozo, E. Menasalvas, and S. Boccaletti, “Optimizing functional network representation of multivariate time series,” *Scientific reports*, vol. 2, no. 1, pp. 1–6, 2012.
- [60] M. L. Stanley, M. N. Moussa, B. Paolini, R. G. Lyday, J. H. Burdette, and P. J. Laurienti, “Defining nodes in complex brain networks,” *Frontiers in computational neuroscience*, vol. 7, p. 169, 2013.
- [61] R. Kozma and W. J. Freeman, *Cognitive phase transitions in the cerebral cortex—enhancing the neuron doctrine by modeling neural fields*. Springer, 2016.
- [62] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: a review,” *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.
- [63] T. Simas and L. M. Rocha, “Distance closures on complex networks,” *Network Science*, vol. 3, no. 2, pp. 227–268, 2015.
- [64] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, “Uncovering the overlapping community structure of complex networks in nature and society,” *nature*, vol. 435, no. 7043, pp. 814–818, 2005.
- [65] E. M. Gordon, T. O. Laumann, B. Adeyemo, J. F. Huckins, W. M. Kelley, and S. E. Petersen, “Generation and evaluation of a cortical area parcellation from resting-state correlations,” *Cerebral cortex*, vol. 26, no. 1, pp. 288–303, 2016.
- [66] O. Korhonen, H. Saarimäki, E. Glerean, M. Sams, and J. Saramäki, “Consistency of regions of interest as nodes of fmri functional brain networks,” *Network Neuroscience*, vol. 1, no. 3, pp. 254–274, 2017.
- [67] A. Zalesky, A. Fornito, I. H. Harding, L. Cocchi, M. Yücel, C. Pantelis, and E. T. Bullmore, “Whole-brain anatomical networks: does the choice of nodes matter?” *Neuroimage*, vol. 50, no. 3, pp. 970–983, 2010.

- [68] M. P. Stumpf, C. Wiuf, and R. M. May, “Subnets of scale-free networks are not scale-free: sampling properties of networks,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 12, pp. 4221–4224, 2005.
- [69] L. Gallos, M. Sigman, and H. Makse, “The conundrum of functional brain networks: small-world efficiency or fractal modularity,” *Frontiers in physiology*, vol. 3, p. 123, 2012.
- [70] A. Di Martino, D. O’Connor, B. Chen, K. Alaerts, J. Anderson, M. Assaf, J. Balsters, L. Baxter, A. Beggiato, S. Bernaerts, L. Blanken, S. Bookheimer, B. Braden, L. Byrge, F. Castellanos, M. Dapretto, R. Delorme, D. Fair, I. Fishman, and M. Milham, “Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii,” *Scientific Data*, vol. 4, p. 170010, 03 2017.
- [71] “Abide ii kennedy kriegler institute collection,” http://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html, 2019, accessed: 2019-12-30.
- [72] “Cobre dataset,” http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html, 2019, accessed: 2019-12-30.
- [73] “Statistical parametric mapping SPM12,” <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>, 2018, accessed: 2019-12-30.
- [74] S. Whitfield-Gabrieli and A. Nieto-Castanon, “Whitfield-gabrieli s, nieto-castanon a. conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. brain connect 2: 125-141,” *Brain connectivity*, vol. 2, pp. 125–41, 05 2012.
- [75] “Artifact detection tools ART,” http://www.nitrc.org/projects/artifact_detect, 2019, accessed: 2019-12-30.
- [76] Y. Behzadi, K. Restom, J. Liau, and T. Liu, “A component based noise correction method (compcor) for bold and perfusion based fmri,” *NeuroImage*, vol. 37, no. 1, pp. 90 – 101, 2007.
- [77] N. Tzourio-Mazoyer, B. Landeau, P. DF, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and J. Marc, “Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain,” *NeuroImage*, vol. 15, pp. 273–89, 02 2002.
- [78] A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux, “Machine learning for neuroimaging with scikit-learn,” *Frontiers in neuroinformatics*, vol. 8, p. 14, 12 2014.

- [79] E. Ryyppö, E. Glerean, E. Brattico, J. Saramäki, and O. Korhonen, “Regions of interest as nodes of dynamic functional brain networks,” *Network Neuroscience*, vol. 2, no. 4, pp. 513–535, 2018.
- [80] N. C. Silver and W. P. Dunlap, “Averaging correlation coefficients: Should fisher’s z transformation be used?” *Journal of Applied Psychology*, vol. 72, no. 1, p. 146, 1987.
- [81] M. Zanin, S. Belkoura, G. Javier, C. Alfaro, and J. Cano, “Topological structures are consistently overestimated in functional complex networks,” *Scientific Reports*, vol. 8, 03 2018.
- [82] N. Henze and B. Zirkler, “A class of invariant consistent tests for multivariate normality,” *Communications in statistics-Theory and Methods*, vol. 19, no. 10, pp. 3595–3617, 1990.
- [83] L. d. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas, “Characterization of complex networks: A survey of measurements,” *Advances in physics*, vol. 56, no. 1, pp. 167–242, 2007.
- [84] B. C. Van Wijk, C. J. Stam, and A. Daffertshofer, “Comparing brain networks of different size and connectivity density using graph theory,” *PloS one*, vol. 5, no. 10, 2010.
- [85] T. E. Nichols, S. Das, S. B. Eickhoff, A. C. Evans, T. Glatard, M. Hanke, N. Kriegeskorte, M. P. Milham, R. A. Poldrack, J.-B. Poline *et al.*, “Best practices in data analysis and sharing in neuroimaging using mri,” *Nature neuroscience*, vol. 20, no. 3, pp. 299–303, 2017.
- [86] F. Váša, E. T. Bullmore, and A. X. Patel, “Probabilistic thresholding of functional connectomes: Application to schizophrenia,” *Neuroimage*, vol. 172, pp. 326–340, 2018.
- [87] M. P. van den Heuvel, S. C. de Lange, A. Zalesky, C. Seguin, B. T. Yeo, and R. Schmidt, “Proportional thresholding in resting-state fmri functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations,” *Neuroimage*, vol. 152, pp. 437–449, 2017.
- [88] M. Jalili, “Functional brain networks: does the choice of dependency estimator and binarization method matter?” *Scientific reports*, vol. 6, p. 29780, 2016.
- [89] S. Achard and E. Bullmore, “Efficiency and cost of economical brain functional networks,” *PLoS Comput Biol*, vol. 3, no. 2, p. e17, 2007.
- [90] M. Rubinov and O. Sporns, “Complex network measures of brain connectivity: uses and interpretations,” *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

- [91] M. E. Newman, “Properties of highly clustered networks,” *Physical Review E*, vol. 68, no. 2, p. 026121, 2003.
- [92] V. Latora and M. Marchiori, “Efficient behavior of small-world networks,” *Physical review letters*, vol. 87, no. 19, p. 198701, 2001.
- [93] R. Noldus and P. Van Mieghem, “Assortativity in complex networks,” *Journal of Complex Networks*, vol. 3, no. 4, pp. 507–542, 2015.
- [94] M. Zanin, P. A. Sousa, and E. Menasalvas, “Information content: Assessing meso-scale structures in complex networks,” *EPL (Europhysics letters)*, vol. 106, no. 3, p. 30001, 2014.
- [95] M. E. Newman, “Modularity and community structure in networks,” *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.
- [96] S. Fortunato, “Community detection in graphs,” *Physics reports*, vol. 486, no. 3-5, pp. 75–174, 2010.
- [97] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.
- [98] A. Lancichinetti and S. Fortunato, “Community detection algorithms: a comparative analysis,” *Physical review E*, vol. 80, no. 5, p. 056117, 2009.
- [99] N. Williams, G. Arnulfo, S. H. Wang, L. Nobili, S. Palva, and J. M. Palva, “Comparison of methods to identify modules in noisy or incomplete brain networks,” *Brain connectivity*, vol. 9, no. 2, pp. 128–143, 2019.
- [100] M. S. Mariani, Z.-M. Ren, J. Bascompte, and C. J. Tessone, “Nestedness in complex networks: Observation, emergence, and implications,” *Physics Reports*, vol. 813, pp. 1–90, 2019.
- [101] D. S. Bassett and E. Bullmore, “Small-world brain networks,” *The neuroscientist*, vol. 12, no. 6, pp. 512–523, 2006.
- [102] C. C. Hilgetag and A. Goulas, “Is the brain really a small-world network?” *Brain Structure and Function*, vol. 221, no. 4, pp. 2361–2366, 2016.
- [103] D. Papo, M. Zanin, J. H. Martínez, and J. M. Buldú, “Beware of the small-world neuroscientist!” *Frontiers in human neuroscience*, vol. 10, p. 96, 2016.
- [104] J. Cohen, *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

- [105] H. D. Rozenfeld, C. Song, and H. A. Makse, “Small-world to fractal transition in complex networks: a renormalization group approach,” *Physical review letters*, vol. 104, no. 2, p. 025701, 2010.
- [106] P. Csermely, “Strong links are important, but weak links stabilize them,” *Trends in biochemical sciences*, vol. 29, no. 7, pp. 331–334, 2004.
- [107] M. Karsai, N. Perra, and A. Vespignani, “Time varying networks and the weakness of strong ties,” *Scientific reports*, vol. 4, p. 4001, 2014.
- [108] V. Braitenberg and A. Schüz, *Cortex: statistics and geometry of neuronal connectivity*. Springer Science & Business Media, 2013.
- [109] K. J. Friston, “The disconnection hypothesis,” *Schizophrenia research*, vol. 30, no. 2, pp. 115–125, 1998.
- [110] A. Hahamy, M. Behrmann, and R. Malach, “The idiosyncratic brain: distortion of spontaneous connectivity patterns in autism spectrum disorder,” *Nature neuroscience*, vol. 18, no. 2, p. 302, 2015.
- [111] F. G. Hillary and J. H. Grafman, “Injured brains and adaptive networks: the benefits and costs of hyperconnectivity,” *Trends in cognitive sciences*, vol. 21, no. 5, pp. 385–401, 2017.
- [112] C. Hohenfeld, C. J. Werner, and K. Reetz, “Resting-state connectivity in neurodegenerative disorders: Is there potential for an imaging biomarker?” *NeuroImage: Clinical*, vol. 18, pp. 849–870, 2018.
- [113] D. Meunier, S. Achard, A. Morcom, and E. Bullmore, “Age-related changes in modular organization of human brain functional networks,” *Neuroimage*, vol. 44, no. 3, pp. 715–723, 2009.
- [114] L. K. Gallos, H. A. Makse, and M. Sigman, “A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 8, pp. 2825–2830, 2012.
- [115] E. Olejarczyk and W. Jernajczyk, “Graph-based analysis of brain connectivity in schizophrenia,” *PloS one*, vol. 12, no. 11, p. e0188629, 2017.
- [116] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, “Clinical diagnosis of alzheimer’s disease: Report of the nincds-adrda work group* under the auspices of department of health and human services task force on alzheimer’s disease,” *Neurology*, vol. 34, no. 7, pp. 939–939, 1984.

- [117] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack Jr, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux *et al.*, “The diagnosis of dementia due to alzheimer’s disease: recommendations from the national institute on aging-alzheimer’s association workgroups on diagnostic guidelines for alzheimer’s disease,” *Alzheimer’s & dementia*, vol. 7, no. 3, pp. 263–269, 2011.
- [118] A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.
- [119] L. Berg, “Clinical dementia rating (cdr),” *Psychopharmacol Bull*, vol. 24, pp. 637–639, 1988.
- [120] İ. Gurvit and B. Baran, “Scales in dementia and cognitive disorders,” *Arch Neuropsychiatry*, vol. 44, pp. 58–65, 2007.
- [121] J. Morris, “The clinical dementia rating (cdr): current version and scoring rules.” *Neurology*, vol. 43, no. 11, pp. 2412–2414, 1993.
- [122] J. C. Morris, “Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the alzheimer type,” *International psychogeriatrics*, vol. 9, no. S1, pp. 173–176, 1997.
- [123] M. F. Folstein, L. N. Robins, and J. E. Helzer, “The mini-mental state examination,” *Archives of general psychiatry*, vol. 40, no. 7, pp. 812–812, 1983.
- [124] C. Gungen, T. Ertan, E. Eker, R. Yaşar, and F. Engin, “Toplumunda hafif demans tanısında geçerlik ve güvenilirliği [standardize mini mental test],” *Türk Psikiyatri Dergisi*, vol. 13, pp. 273–281, 2002.
- [125] S. Daniel and A. Lees, “Parkinson’s disease society brain bank, london: overview and research.” *Journal of neural transmission. Supplementum*, vol. 39, pp. 165–172, 1993.
- [126] A. E. T. Lang and F. S, “Assessment of parkinson’s disease,” in *Quantification of neurological deficit*. Butterworths, 1989, pp. 285–309.
- [127] M. M. Hoehn and M. D. Yahr, “Parkinsonism: onset, progression, and mortality,” *Neurology*, vol. 17, no. 5, pp. 427–427, 1967.
- [128] X. L. Zhang, H. Begleiter, B. Porjesz, W. Wang, and A. Litke, “Event related potentials during object recognition tasks,” *Brain Research Bulletin*, vol. 38, no. 6, pp. 531–538, 1995.

- [129] R. Cao, Z. Wu, H. Li, J. Xiang, and J. Chen, “Disturbed connectivity of eeg functional networks in alcoholism: a graph-theoretic analysis,” *Bio-Medical Materials and Engineering*, vol. 24, no. 6, pp. 2927–2936, 2014.
- [130] J. G. Snodgrass and M. Vanderwart, “A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity.” *Journal of Experimental Psychology: Human Learning and Memory*, vol. 6, no. 2, p. 174, 1980.
- [131] A. Di Martino, D. O’connor, B. Chen, K. Alaerts, J. S. Anderson, M. Assaf, J. H. Balsters, L. Baxter, A. Beggiato, S. Bernaerts *et al.*, “Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii,” *Scientific data*, vol. 4, no. 1, pp. 1–15, 2017.
- [132] W. Neuroimaging, “Spm12-statistical parametric mapping,” *The FIL Methods group*. <http://www.fil.ion.ucl.ac.uk/spm/software/spm12>, 2016.
- [133] S. Whitfield-Gabrieli and A. Nieto-Castanon, “Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks,” *Brain connectivity*, vol. 2, no. 3, pp. 125–141, 2012.
- [134] S. Whitfield-Gabrieli, A. Nieto-Castanon, and S. Ghosh, “Artifact detection tools (art),” *Camb., Ma. Release Version*, 2011.
- [135] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, “Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain,” *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.
- [136] J. D. Rudie, J. Brown, D. Beck-Pancer, L. Hernandez, E. Dennis, P. Thompson, S. Bookheimer, and M. Dapretto, “Altered functional and structural brain network organization in autism,” *NeuroImage: clinical*, vol. 2, pp. 79–94, 2013.
- [137] J. A. Brown, J. D. Rudie, A. Bandrowski, J. D. Van Horn, and S. Y. Bookheimer, “The ucla multimodal connectivity database: a web-based platform for brain connectivity matrix sharing and analysis,” *Frontiers in neuroinformatics*, vol. 6, p. 28, 2012.
- [138] K. B. Nooner, S. Colcombe, R. Tobe, M. Mennes, M. Benedict, A. Moreno, L. Panek, S. Brown, S. Zavitz, Q. Li *et al.*, “The nki-rockland sample: a model for accelerating the pace of discovery science in psychiatry,” *Frontiers in neuroscience*, vol. 6, p. 152, 2012.
- [139] S. L. Bressler and A. K. Seth, “Wiener–granger causality: a well established methodology,” *Neuroimage*, vol. 58, no. 2, pp. 323–329, 2011.

- [140] A. Kraskov, H. Stögbauer, and P. Grassberger, “Estimating mutual information,” *Physical review E*, vol. 69, no. 6, p. 066138, 2004.
- [141] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, “Transfer entropy - a model-free measure of effective connectivity for the neurosciences,” *Journal of computational neuroscience*, vol. 30, no. 1, pp. 45–67, 2011.
- [142] A. L. Goldberger, L. A. Amaral, J. M. Hausdorff, P. C. Ivanov, C.-K. Peng, and H. E. Stanley, “Fractal dynamics in physiology: alterations with disease and aging,” *Proceedings of the national academy of sciences*, vol. 99, no. suppl 1, pp. 2466–2472, 2002.
- [143] B. J. West, “Fractal physiology and the fractional calculus: a perspective,” *Frontiers in physiology*, vol. 1, p. 12, 2010.
- [144] B. J. Roach and D. H. Mathalon, “Event-related eeg time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking in schizophrenia,” *Schizophrenia bulletin*, vol. 34, no. 5, pp. 907–926, 2008.
- [145] D. V. Moretti, C. Babiloni, G. Binetti, E. Cassetta, G. Dal Forno, F. Ferreric, R. Ferri, B. Lanuzza, C. Miniussi, F. Nobili *et al.*, “Individual analysis of eeg frequency and band power in mild alzheimer’s disease,” *Clinical Neurophysiology*, vol. 115, no. 2, pp. 299–308, 2004.
- [146] S. Finger, *Origins of neuroscience: a history of explorations into brain function*. Oxford University Press, USA, 2001.
- [147] S. H. Strogatz, “Exploring complex networks,” *nature*, vol. 410, no. 6825, pp. 268–276, 2001.
- [148] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: Structure and dynamics,” *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
- [149] D. S. Bassett and E. T. Bullmore, “Human brain networks in health and disease,” *Current opinion in neurology*, vol. 22, no. 4, p. 340, 2009.
- [150] N. Altman and M. Krzywinski, “The curse (s) of dimensionality,” *Nat Methods*, vol. 15, no. 6, pp. 399–400, 2018.
- [151] G. W. Snedecor and W. G. Cochran, “Statistical methods, 8th edn,” *Ames: Iowa State Univ. Press Iowa*, vol. 54, pp. 71–82, 1989.
- [152] D. W. Zimmerman, “A note on preliminary tests of equality of variances,” *British Journal of Mathematical and Statistical Psychology*, vol. 57, no. 1, pp. 173–181, 2004.

- [153] Y. Benjamini and Y. Hochberg, “Controlling the false discovery rate: a practical and powerful approach to multiple testing,” *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.
- [154] S. Holm, “A simple sequentially rejective multiple test procedure,” *Scandinavian journal of statistics*, pp. 65–70, 1979.
- [155] K. Korthauer, P. K. Kimes, C. Duvallet, A. Reyes, A. Subramanian, M. Teng, C. Shukla, E. J. Alm, and S. C. Hicks, “A practical guide to methods controlling false discoveries in computational biology,” *Genome biology*, vol. 20, no. 1, pp. 1–21, 2019.
- [156] O. J. Dunn, “Multiple comparisons among means,” *Journal of the American statistical association*, vol. 56, no. 293, pp. 52–64, 1961.
- [157] G. Hommel, “A stagewise rejective multiple test procedure based on a modified bonferroni test,” *Biometrika*, vol. 75, no. 2, pp. 383–386, 1988.
- [158] Y. Hochberg, “A sharper bonferroni procedure for multiple tests of significance,” *Biometrika*, vol. 75, no. 4, pp. 800–802, 1988.
- [159] C. Bonferroni, “Statistical class theory and probability calculation,” *Publications of the R Higher Institute of Economic and Commercial Sciences of Florence*, vol. 8, pp. 3–62, 1936.
- [160] A. Zalesky, A. Fornito, and E. T. Bullmore, “Network-based statistic: identifying differences in brain networks,” *Neuroimage*, vol. 53, no. 4, pp. 1197–1207, 2010.
- [161] H. C. Baggio, A. Abos, B. Segura, A. Campabadal, A. Garcia-Diaz, C. Uribe, Y. Compta, M. J. Marti, F. Valldeoriola, and C. Junque, “Statistical inference in brain graphs using threshold-free network-based statistics,” *Human Brain Mapping*, vol. 39, no. 6, pp. 2289–2302, 2018. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.24007>
- [162] G. Salimi-Khorshidi, S. M. Smith, and T. E. Nichols, “Adjusting the effect of nonstationarity in cluster-based and tfce inference,” *Neuroimage*, vol. 54, no. 3, pp. 2006–2019, 2011.
- [163] J. Kim, W. Pan, A. D. N. Initiative *et al.*, “Highly adaptive tests for group differences in brain functional connectivity,” *NeuroImage: Clinical*, vol. 9, pp. 625–639, 2015.
- [164] W. Pan, J. Kim, Y. Zhang, X. Shen, and P. Wei, “A powerful and adaptive association test for rare variants,” *Genetics*, vol. 197, no. 4, pp. 1081–1095, 2014.

- [165] J. Kim, J. R. Wozniak, B. A. Mueller, X. Shen, and W. Pan, “Comparison of statistical tests for group differences in brain functional networks,” *NeuroImage*, vol. 101, pp. 681–694, 2014.
- [166] B. L. Welch, “The generalization of student’s’ problem when several different population variances are involved,” *Biometrika*, vol. 34, no. 1/2, pp. 28–35, 1947.
- [167] C. E. Ginestet, J. Li, P. Balachandran, S. Rosenberg, E. D. Kolaczyk *et al.*, “Hypothesis testing for network data in functional neuroimaging,” *The Annals of Applied Statistics*, vol. 11, no. 2, pp. 725–750, 2017.
- [168] J. Fan, Y. Liao, and H. Liu, “An overview of the estimation of large covariance and precision matrices,” 2016.
- [169] J. Bai and S. Shi, “Estimating high dimensional covariance matrices and its applications,” 2011.
- [170] J. Bien and R. J. Tibshirani, “Sparse estimation of a covariance matrix,” *Biometrika*, vol. 98, no. 4, pp. 807–820, 2011.
- [171] F. Oztoprak, J. Nocedal, S. Rennie, and P. A. Olsen, “Newton-like methods for sparse inverse covariance estimation,” *Advances in neural information processing systems*, vol. 25, pp. 755–763, 2012.
- [172] O. Ledoit and M. Wolf, “A well-conditioned estimator for large-dimensional covariance matrices,” *Journal of multivariate analysis*, vol. 88, no. 2, pp. 365–411, 2004.
- [173] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, “Scikit-learn: Machine learning in python,” *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.
- [174] I. A. Higgins, S. Kundu, K. S. Choi, H. S. Mayberg, and Y. Guo, “A difference degree test for comparing brain networks,” *Human brain mapping*, vol. 40, no. 15, pp. 4518–4536, 2019.
- [175] M. Hirschberger, Y. Qi, and R. E. Steuer, “Randomly generating portfolio-selection covariance matrices with specified distributional characteristics,” *European Journal of Operational Research*, vol. 177, no. 3, pp. 1610–1625, 2007.
- [176] D. Fraiman and R. Fraiman, “An anova approach for statistical comparisons of brain networks,” *Scientific Reports*, vol. 8, 03 2018.

- [177] A. Sanfeliu and K.-S. Fu, “A distance measure between attributed relational graphs for pattern recognition,” *IEEE transactions on systems, man, and cybernetics*, no. 3, pp. 353–362, 1983.
- [178] I. Lovato, A. Pini, A. Stamm, and S. Vantini, “Model-free two-sample test for network-valued data,” *Computational Statistics & Data Analysis*, vol. 144, p. 106896, 2020.
- [179] D. E. Meskaldji, M.-C. Ottet, L. Cammoun, P. Hagmann, R. Meuli, S. Eliez, J. P. Thiran, and S. Morgenthaler, “Adaptive strategy for the statistical analysis of connectomes,” *PloS one*, vol. 6, no. 8, p. e23009, 2011.
- [180] A. Zalesky, L. Cocchi, A. Fornito, M. M. Murray, and E. Bullmore, “Connectivity differences in brain networks,” *Neuroimage*, vol. 60, no. 2, pp. 1055–1062, 2012.
- [181] S. Chen, J. Kang, Y. Xing, and G. Wang, “A parsimonious statistical method to detect groupwise differentially expressed functional connectivity networks,” *Human brain mapping*, vol. 36, no. 12, pp. 5196–5206, 2015.
- [182] H. Chen and J. H. Friedman, “A new graph-based two-sample test for multivariate and object data,” *Journal of the American statistical association*, vol. 112, no. 517, pp. 397–409, 2017.
- [183] D. Durante and D. B. Dunson, “Bayesian inference and testing of group differences in brain networks,” *Bayesian Anal.*, vol. 13, no. 1, pp. 29–58, 03 2018. [Online]. Available: <https://doi.org/10.1214/16-BA1030>
- [184] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.
- [185] H. Liu and H. Motoda, *Feature selection for knowledge discovery and data mining*. Springer Science & Business Media, 2012, vol. 454.
- [186] I. H. Witten and E. Frank, “Data mining: practical machine learning tools and techniques with java implementations,” *Acm Sigmod Record*, vol. 31, no. 1, pp. 76–77, 2002.
- [187] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [188] J. L. Gastwirth, “The Estimation of the Lorenz Curve and Gini Index,” *The Review of Economics and Statistics*, vol. 54, no. 3, pp. 306–316, August 1972.
- [189] M. Robnik-Šikonja and I. Kononenko, “Theoretical and empirical analysis of relieff and rrelieff,” vol. 53, no. 1–2, 2003.

- [190] K. Kira and L. A. Rendell, “A practical approach to feature selection,” in *Proceedings of the Ninth International Workshop on Machine Learning*, ser. ML92. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1992, p. 249–256.
- [191] I. Kononenko, “Estimating attributes: Analysis and extensions of relief,” in *Machine Learning: ECML-94*, F. Bergadano and L. De Raedt, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 171–182.
- [192] C. Greene, N. Penrod, J. Kiralis, and J. Moore, “Spatially uniform relief (surf) for computationally-efficient filtering of gene-gene interactions,” *BioData mining*, vol. 2, p. 5, 09 2009.
- [193] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” *Mach. Learn.*, vol. 46, no. 1–3, p. 389–422, Mar. 2002. [Online]. Available: <https://doi.org/10.1023/A:1012487302797>
- [194] M. Kursa, A. Jankowski, and W. Rudnicki, “Boruta - a system for feature selection,” *Fundam. Inform.*, vol. 101, pp. 271–285, 01 2010.
- [195] A. Crimi, L. Giancardo, F. Sambataro, A. Gozzi, V. Murino, and D. Sona, “Multilink analysis: brain network comparison via sparse connectivity analysis,” *Scientific reports*, vol. 9, no. 1, pp. 1–13, 2019.
- [196] L. Clemmensen, T. Hastie, D. Witten, and B. Ersbøll, “Sparse discriminant analysis,” *Technometrics*, vol. 53, no. 4, pp. 406–413, 2011.
- [197] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society (Series B)*, vol. 58, pp. 267–288, 1996.
- [198] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society, Series B*, vol. 67, pp. 301–320, 2005.
- [199] J. A. Brown and J. D. Van Horn, “Connected brains and minds—the umcd repository for brain connectivity matrices,” *Neuroimage*, vol. 124, pp. 1238–1241, 2016.
- [200] J. A. Brown, A. Y. Hua, A. Trujillo, S. Attygalle, R. J. Binney, S. Spina, S. E. Lee, J. H. Kramer, B. L. Miller, H. J. Rosen *et al.*, “Advancing functional dysconnectivity and atrophy in progressive supranuclear palsy,” *NeuroImage: Clinical*, vol. 16, pp. 564–574, 2017.
- [201] I. Rish *et al.*, “An empirical study of the naive bayes classifier,” in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22, 2001, pp. 41–46.

- [202] S. R. Safavian and D. Landgrebe, “A survey of decision tree classifier methodology,” *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
- [203] A. Liaw, M. Wiener *et al.*, “Classification and regression by randomforest,” *R news*, vol. 2, no. 3, pp. 18–22, 2002.
- [204] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” *Machine learning*, vol. 63, no. 1, pp. 3–42, 2006.
- [205] L. Bottou, “Large-scale machine learning with stochastic gradient descent,” in *Proceedings of COMPSTAT’2010*. Springer, 2010, pp. 177–186.
- [206] C. Ying, M. Qi-Guang, L. Jia-Chen, and G. Lin, “Advance and prospects of adaboost algorithm,” *Acta Automatica Sinica*, vol. 39, no. 6, pp. 745–758, 2013.
- [207] Y. Freund, R. E. Schapire *et al.*, “Experiments with a new boosting algorithm,” in *icml*, vol. 96. Citeseer, 1996, pp. 148–156.
- [208] M. Riedmiller, “Advanced supervised learning in multi-layer perceptrons—from backpropagation to adaptive learning algorithms,” *Computer Standards & Interfaces*, vol. 16, no. 3, pp. 265–278, 1994.
- [209] L. Wang, *Support vector machines: theory and applications*. Springer Science & Business Media, 2005, vol. 177.
- [210] A. E. Hoerl, R. W. Kannard, and K. F. Baldwin, “Ridge regression: some simulations,” *Communications in Statistics-Theory and Methods*, vol. 4, no. 2, pp. 105–123, 1975.
- [211] J. H. Friedman, “Stochastic gradient boosting,” *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367–378, 2002.
- [212] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013, vol. 398.
- [213] G. Batista, D. F. Silva *et al.*, “How k-nearest neighbor parameters affect its performance,” in *Argentine symposium on artificial intelligence*. Citeseer, 2009, pp. 1–12.
- [214] D. Ruta and B. Gabrys, “Classifier selection for majority voting,” *Information fusion*, vol. 6, no. 1, pp. 63–81, 2005.
- [215] L. Breiman, “Bagging predictors,” *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

- [216] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [217] L. Deng and D. Yu, “Deep learning: methods and applications,” *Foundations and trends in signal processing*, vol. 7, no. 3–4, pp. 197–387, 2014.
- [218] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [219] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv:1609.02907*, 2016.
- [220] M. Zhang, Z. Cui, M. Neumann, and Y. Chen, “An end-to-end deep learning architecture for graph classification,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.
- [221] A. Kalousis, J. Prados, and M. Hilario, “Stability of feature selection algorithms: A study on high-dimensional spaces,” *Knowl. Inf. Syst.*, vol. 12, pp. 95–116, 05 2007.
- [222] N. C. Andreasen, “The role of the thalamus in schizophrenia,” *The Canadian Journal of Psychiatry*, vol. 42, no. 1, pp. 27–33, 1997.
- [223] W. Byne, E. A. Hazlett, M. S. Buchsbaum, and E. Kemether, “The thalamus and schizophrenia: current status of research,” *Acta neuropathologica*, vol. 117, no. 4, pp. 347–368, 2009.
- [224] D. Pinault, “Dysfunctional thalamus-related networks in schizophrenia,” *Schizophrenia bulletin*, vol. 37, no. 2, pp. 238–243, 2011.
- [225] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra, “Circos: an information aesthetic for comparative genomics,” *Genome research*, vol. 19, no. 9, pp. 1639–1645, 2009.
- [226] Q. K. Telesford, A. R. Morgan, S. Hayasaka, S. L. Simpson, W. Barret, R. A. Kraft, J. L. Mozolic, and P. J. Laurienti, “Reproducibility of graph metrics in fmri networks,” *Frontiers in neuroinformatics*, vol. 4, p. 117, 2010.
- [227] H. E. Wang, C. G. Bénar, P. P. Quilichini, K. J. Friston, V. K. Jirsa, and C. Bernard, “A systematic framework for functional connectivity measures,” *Frontiers in neuroscience*, vol. 8, p. 405, 2014.
- [228] R. Botvinik-Nezer, F. Holzmeister, C. F. Camerer, A. Dreber, J. Huber, M. Johannesson, M. Kirchler, R. Iwanir, J. A. Mumford, R. A. Adcock *et al.*,

- “Variability in the analysis of a single neuroimaging dataset by many teams,” *Nature*, vol. 582, no. 7810, pp. 84–88, 2020.
- [229] L. Deuker, E. T. Bullmore, M. Smith, S. Christensen, P. J. Nathan, B. Rockstroh, and D. S. Bassett, “Reproducibility of graph metrics of human brain functional networks,” *Neuroimage*, vol. 47, no. 4, pp. 1460–1468, 2009.
- [230] M. Hardmeier, F. Hatz, H. Bousleiman, C. Schindler, C. J. Stam, and P. Fuhr, “Reproducibility of functional connectivity and graph measures based on the phase lag index (pli) and weighted phase lag index (wpli) derived from high resolution eeg,” *PloS one*, vol. 9, no. 10, p. e108648, 2014.
- [231] Y. Höller, A. Uhl, A. Bathke, A. Thomschewski, K. Butz, R. Nardone, J. Fell, and E. Trinka, “Reliability of eeg measures of interaction: a paradigm shift is needed to fight the reproducibility crisis,” *Frontiers in human neuroscience*, vol. 11, p. 441, 2017.
- [232] B. J. Harrison, J. Pujol, M. López-Solà, R. Hernández-Ribas, J. Deus, H. Ortiz, C. Soriano-Mas, M. Yücel, C. Pantelis, and N. Cardoner, “Consistency and functional specialization in the default mode brain network,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 28, pp. 9781–9786, 2008.
- [233] M. E. Raichle, “The brain’s default mode network,” *Annual review of neuroscience*, vol. 38, pp. 433–447, 2015.
- [234] R. F. Algunaid, A. H. Algumaei, M. A. Rushdi, and I. A. Yassine, “Schizophrenic patient identification using graph-theoretic features of resting-state fmri data,” *Biomedical Signal Processing and Control*, vol. 43, pp. 289–299, 2018.
- [235] M. N. I. Qureshi, J. Oh, and B. Lee, “3d-cnn based discrimination of schizophrenia using resting-state fmri,” *Artificial intelligence in medicine*, vol. 98, pp. 10–17, 2019.
- [236] K. Dadi, M. Rahim, A. Abraham, D. Chyzyk, M. Milham, B. Thirion, G. Varoquaux, A. D. N. Initiative *et al.*, “Benchmarking functional connectome-based predictive models for resting-state fmri,” *NeuroImage*, vol. 192, pp. 115–134, 2019.
- [237] B. M. Adhikari, N. Jahanshad, D. Shukla, J. Turner, D. Grotegerd, U. Dannlowski, H. Kugel, J. Engelen, B. Dietsche, A. Krug *et al.*, “A resting state fmri analysis pipeline for pooling inference across diverse cohorts: an enigma rs-fmri protocol,” *Brain imaging and behavior*, vol. 13, no. 5, pp. 1453–1467, 2019.

- [238] D. Papo, M. Zanin, J. A. Pineda-Pardo, S. Boccaletti, and J. M. Buldú, “Functional brain networks: great expectations, hard times and the big leap forward,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1653, p. 20130525, 2014.
- [239] J. Kim, V. D. Calhoun, E. Shim, and J.-H. Lee, “Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia,” *Neuroimage*, vol. 124, pp. 127–146, 2016.
- [240] X. Guo, K. C. Dominick, A. A. Minai, H. Li, C. A. Erickson, and L. J. Lu, “Diagnosing autism spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method,” *Frontiers in neuroscience*, vol. 11, p. 460, 2017.
- [241] X. Bi, X. Zhao, H. Huang, D. Chen, and Y. Ma, “Functional brain network classification for alzheimer’s disease detection with deep features and extreme learning machine,” *Cognitive Computation*, vol. 12, no. 3, pp. 513–527, 2020.
- [242] J. Richiardi and B. Ng, “Recent advances in supervised learning for brain graph classification,” in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 907–910.
- [243] J. D. A. Reli3n, D. Kessler, E. Levina, and S. F. Taylor, “Network classification with applications to brain connectomics,” *The annals of applied statistics*, vol. 13, no. 3, p. 1648, 2019.
- [244] G. Chen, B. D. Ward, C. Xie, W. Li, Z. Wu, J. L. Jones, M. Franczak, P. Antuono, and S.-J. Li, “Classification of alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional mr imaging,” *Radiology*, vol. 259, no. 1, pp. 213–221, 2011.
- [245] J. A. Nielsen, B. A. Zielinski, P. T. Fletcher, A. L. Alexander, N. Lange, E. D. Bigler, J. E. Lainhart, and J. S. Anderson, “Multisite functional connectivity mri classification of autism: Abide results,” *Frontiers in human neuroscience*, vol. 7, p. 599, 2013.
- [246] M. Plitt, K. A. Barnes, and A. Martin, “Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards,” *NeuroImage: Clinical*, vol. 7, pp. 359–366, 2015.

- [247] A. Khazaei, A. Ebrahimzadeh, and A. Babajani-Feremi, “Application of advanced machine learning methods on resting-state fmri network for identification of mild cognitive impairment and alzheimer’s disease,” *Brain imaging and behavior*, vol. 10, no. 3, pp. 799–817, 2016.
- [248] D. Mastrovito, C. Hanson, and S. J. Hanson, “Differences in atypical resting-state effective connectivity distinguish autism from schizophrenia,” *NeuroImage: Clinical*, vol. 18, pp. 367–376, 2018.
- [249] L.-L. Zeng, H. Shen, L. Liu, L. Wang, B. Li, P. Fang, Z. Zhou, Y. Li, and D. Hu, “Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis,” *Brain*, vol. 135, no. 5, pp. 1498–1507, 2012.
- [250] S. Hiwa, S. Obuchi, and T. Hiroyasu, “Automated extraction of human functional brain network properties associated with working memory load through a machine learning-based feature selection algorithm,” *Computational intelligence and neuroscience*, vol. 2018, 2018.
- [251] X. Zhang, B. Hu, X. Ma, and L. Xu, “Resting-state whole-brain functional connectivity networks for mci classification using l2-regularized logistic regression,” *IEEE transactions on nanobioscience*, vol. 14, no. 2, pp. 237–247, 2015.
- [252] O. Sporns, C. J. Honey, and R. Kötter, “Identification and classification of hubs in brain networks,” *PloS one*, vol. 2, no. 10, p. e1049, 2007.
- [253] B. Jie, D. Zhang, C.-Y. Wee, and D. Shen, “Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification,” *Human brain mapping*, vol. 35, no. 7, pp. 2876–2897, 2014.
- [254] G. Ma, C.-T. Lu, L. He, S. Y. Philip, and A. B. Ragin, “Multi-view graph embedding with hub detection for brain network analysis,” in *2017 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2017, pp. 967–972.
- [255] C. Morris and I. Rekić, “Autism spectrum disorder diagnosis using sparse graph embedding of morphological brain networks,” in *Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics*. Springer, 2017, pp. 12–20.
- [256] L. Tang, S. Mostafa, B. Liao, and F.-X. Wu, “A network clustering based feature selection strategy for classifying autism spectrum disorder,” *BMC medical genomics*, vol. 12, no. 7, pp. 1–10, 2019.

- [257] J. D. Wilson, M. Baybay, R. Sankar, P. Stillman, and A. M. Popa, “Analysis of population functional connectivity data via multilayer network embeddings,” *Network Science*, vol. 9, no. 1, pp. 99–122, 2021.
- [258] P. Goyal and E. Ferrara, “Graph embedding techniques, applications, and performance: A survey,” *Knowledge-Based Systems*, vol. 151, pp. 78–94, 2018.
- [259] A. Azran, “The rendezvous algorithm: Multiclass semi-supervised learning with markov random walks,” in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 49–56.
- [260] S. Bhagat, I. Rozenbaum, and G. Cormode, “Applying link-based classification to label blogs,” in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, 2007, pp. 92–101.
- [261] A. Clauset, C. Moore, and M. E. Newman, “Hierarchical structure and the prediction of missing links in networks,” *Nature*, vol. 453, no. 7191, pp. 98–101, 2008.
- [262] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer, “Learning probabilistic relational models,” in *IJCAI*, vol. 99, 1999, pp. 1300–1309.
- [263] D. Heckerman, C. Meek, and D. Koller, “Probabilistic entity-relationship models, prms, and plate models,” *Introduction to statistical relational learning*, pp. 201–238, 2007.
- [264] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [265] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [266] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [267] A. Ahmed, N. Shervashidze, S. Narayanamurthy, V. Josifovski, and A. J. Smola, “Distributed large-scale natural graph factorization,” in *Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 37–48.
- [268] M. Ou, P. Cui, J. Pei, Z. Zhang, and W. Zhu, “Asymmetric transitivity preserving graph embedding,” in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 1105–1114.

- [269] D. Wang, P. Cui, and W. Zhu, “Structural deep network embedding,” in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 1225–1234.
- [270] A. Grover and J. Leskovec, “node2vec: Scalable feature learning for networks,” in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 855–864.