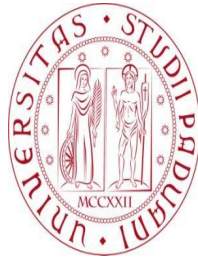


Università degli Studi di Padova
Dipartimento di Scienze Statistiche

Corso di Laurea Triennale in
Statistica per le Tecnologie e le Scienze



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

RELAZIONE FINALE
**Analisi della connettività funzionale cerebrale a
riposo: convalida dei risultati di clustering tramite
l'utilizzo di modelli nulli**

Relatore Prof. Livio Finos
Dipartimento di Scienze Statistiche

Correlatore Dott. Allegra Michele
Dipartimento di Fisica e Astronomia

Laureando: Andrea Simonetto
Matricola N 1221185

Anno Accademico 2022/2023

INDICE

1. INTRODUZIONE

2. ELEMENTI TEORICI

2.1 Introduzione alla Functional Connectivity

2.2 Stati dinamici in stato di riposo

3. ANALISI DEI DATI

3.1 - Introduzione al dataset

3.2 - Analisi preliminari

3.3 - Sliding Window Correlation

3.4 - Clustering

3.5 - Indici di cluster

4. MODELLI NULLI

4.1 - Introduzione ai modelli nulli

4.2 - Phase Randomized

4.3 - Analisi sul dataset PR

5. CONCLUSIONI

5.1 - Confronto dei Risultati

5.2 - Conclusioni

6. BIBLIOGRAFIA

1 - INTRODUZIONE

L'interesse verso lo studio del cervello è una costante all'interno della storia umana. Il fascino dei misteri che circondano la mente ha sempre alimentato la curiosità degli studiosi. E fin dagli inizi la complessità delle sue funzioni cognitive, la struttura intricata, le molteplici sfaccettature e la sua innegabile fragilità hanno reso il cervello un elemento di difficilissima indagine.

Cominciando dalla trapanazione e la raschiatura di un foro nel cranio per provare a curare un mal di testa, l'essere umano ha fatto molti progressi nel campo delle neuroscienze. Soprattutto negli ultimi decenni sono arrivate modifiche rivoluzionarie nell'analisi cerebrale, catalizzate dall'avanzamento delle tecniche di imaging, dall'impatto dell'intelligenza artificiale e anche dalla convergenza di discipline come la neuroscienza, la psicologia, l'informatica e la statistica. Questa sinergia di conoscenze ha aperto nuove porte alla comprensione del cervello, permettendo agli scienziati di esplorare i meccanismi sottili che regolano l'emozione, il pensiero, la memoria e persino la coscienza. La ricerca sul cervello non riguarda soltanto la sfera accademica, ma ha implicazioni profonde in vari ambiti della società, dall'educazione alla salute mentale, dalla progettazione tecnologica all'etica.

La sfida rimane notevole: esplorare le fondamenta della mente umana è un'impresa che richiede impegno costante e un approccio umile di fronte alla sua straordinaria complessità.

Il mio progetto di tesi si colloca all'interno di questa stimolante cornice di ricerca.

Questo elaborato si pone come obiettivo inizialmente di indagare le variazioni temporali della connettività funzionale cerebrale su un campione di 96 soggetti, andando a cercare degli stati dinamici (*patterns* di connettività ricorrenti) comuni a tutti i pazienti. Questo primo passo viene effettuato attraverso analisi di clustering.

Successivamente si andrà a verificare la significatività dei risultati ottenuti dall'analisi di clustering grazie all'utilizzo di modelli nulli (phase randomized. Si veda il capitolo 4 per ulteriori spiegazioni).

Nel primo secondo verranno introdotti dei concetti di carattere teorico riguardanti gli argomenti in esame.

Il terzo capitolo descriverà i dati utilizzati e le analisi fatte su di essi.

Nel quarto capitolo verrà introdotto il concetto di modello nullo e verrà descritta la sua implementazione all'interno di questo progetto per la convalida dei risultati.

L'ultimo capitolo verrà destinato alle conclusioni finali del progetto.

2 - ELEMENTI TEORICI

2.1 Introduzione alla Functional Connectivity

Grazie ai recenti avanzamenti tecnologici è stato possibile indagare e descrivere l'organizzazione del cervello con un dettaglio sempre maggiore. Ciò ha permesso di identificare importanti caratteristiche funzionali dell'architettura cerebrale. Questa rivoluzione ha infatti spostato l'attenzione sul funzionamento del cervello come un insieme integrato e interconnesso, ponendo un focus particolare su come i neuroni e i loro aggregati a varie scale (neuroni, popolazioni neuronali o aree di materia grigia) si influenzano l'uno con l'altro.

In questo ambito l'interdipendenza tra aree cerebrali può essere caratterizzata non soltanto misurando le loro connessioni anatomiche (*structural connectivity*), ma anche analizzando in dettaglio la tendenza di diverse aree cerebrali a attivarsi simultaneamente o in sequenza durante processi mentali e attività cognitive specifiche.

Questo tipo di interconnessione prende il nome di connettività funzionale (o *Functional Connectivity, FC*): più nello specifico può essere definita come la correlazione tra eventi neurofisiologici distanti tra loro, a prescindere dalla presenza di una connessione anatomica.

Numerose analisi hanno mostrato che questo fenomeno di interconnessione non è statico ma bensì dinamico e mutevole nel tempo, anche col cervello in uno *stato di riposo (resting state, situazione di inattività o "task negative"*, ovvero quando non viene eseguito un compito esplicito). Per questo motivo è diventato di interesse primario lo studio del cambiamento al variare del tempo della FC, ovvero della *Dynamic Functional Connectivity (DFC)*.

Mentre la connettività funzionale prende in esame le reti funzionali come eventi statici nel tempo, l'analisi della DFC permette di evidenziare le variazioni di interconnessione tra regioni cerebrali non solo con lo scorrere del tempo, ma anche cercandone le differenze a seconda dei compiti che il cervello è portato a svolgere, quali esercizi operativi, il sonno o l'apprendimento.

Le tecniche di *neuroimaging* che permettono di analizzare la DFC sono molteplici.

La fMRI (functional magnetic resonance imaging) è un metodo di visualizzazione dell'attività cerebrale in uno stato di riposo che si basa sui cambiamenti di flusso sanguigno all'interno dei vasi cerebrali. Questi cambiamenti danno origine a quello che viene denominato come

segnale BOLD (*blood-oxygen-level dependent*) e che può essere misurato tramite la risonanza magnetica. I dati utilizzati in questo elaborato provengono da fMRI effettuate su 96 pazienti, si veda il paragrafo 3.1 per maggiori dettagli.

2.2 Stati dinamici in condizioni di riposo

Di interesse in questo contesto è l'analisi del cervello a riposo (*resting state*). La connettività funzionale che sorge durante il periodo di riposo, e non è legata a specifiche attività cognitive o motorie, fornisce una visione unica sulla struttura integrata del cervello. Durante il *resting state*, si alternano differenti pattern di connettività, o *stati dinamici*, tra diverse aree cerebrali, identificati mediante il processo di clustering delle serie temporali delle attività neurali. L'analisi degli stati dinamici offre uno strumento potenzialmente prezioso per esplorare e comprendere l'attività cerebrale.

Ciò che rende questa analisi vuole andare ad indagare è la supposizione che le variazioni temporali della connettività cerebrale riflettano modifiche dinamiche nell'organizzazione cerebrale. Queste variazioni possono essere interpretate come transizioni tra stati discreti del cervello, o come cambiamenti nel modo in cui diverse regioni cerebrali interagiscono tra loro.

Questo approccio sposta l'attenzione dalla mera descrizione anatomica del cervello a una comprensione più approfondita delle sue dinamiche funzionali. Attraverso l'analisi degli stati dinamici è possibile identificare e studiare i *patterns* di comunicazione tra le diverse regioni cerebrali e come questi possano variare nel tempo o in risposta a differenti stimoli ambientali.

L'analisi degli stati dinamici ha il potenziale di fornire informazioni cruciali per lo studio di patologie neurologiche e psichiatriche, poiché in studi pregressi alterazioni nella connettività funzionale sono state correlate a diverse condizioni cliniche: si veda per esempio lo studio di Favaretto et al. (2022) ove è stato mostrato che questi stati dinamici subivano drastici cambiamenti in pazienti che erano stati colpiti da un ictus.

3 - ANALISI DEI DATI

3.1 - Introduzione al dataset

Sono stati utilizzati 96 soggetti, senza alcuna relazione tra loro, dal Human Connectome Project (HCP) (Van Essen et al. 2013).

Abbiamo utilizzato le “left-right (LR) phase encoding run” dei dati di risonanza magnetica funzionale (fMRI) in stato di riposo¹. Le registrazioni avevano una durata di circa 15 minuti con un TR (tempo di ripetizione) di 0.72 secondi. Per maggiori informazioni riguardanti il *preprocessing* dei dati ci si rifaccia a Glasser et al. 2013.

In breve, dopo la correzione per il movimento, le distorsioni di gradiente e di suscettibilità, i dati fMRI sono stati allineati a un'immagine anatomica. L'immagine funzionale allineata è stata corretta per il bias dell'intensità, normalizzata e proiettata su uno spazio superficiale comune.

Tutti i dati fMRI sono stati filtrati tra 0,1 e 0,01 Hz per mantenere la gamma di frequenza rilevante per ulteriori analisi del segnale BOLD. Sono state ottenute matrici strutturali e funzionali in diverse scale spaziali utilizzando la parcellazione corticale di Schaefer et al. (2018), che ottimizza misure di gradiente locale e similitudine globale del segnale fMRI in varie scale spaziali, arricchito con 19 regioni sottocorticali tratte dall'atlante di Freesurfer (Fischl 2012), per un totale di 119 regioni cerebrali. Le serie temporali sono state estratte con l'aiuto del comando Workbench fornito dall'HCP.

¹ La risonanza magnetica funzionale in stato di riposo (rs-fMRI o R-fMRI) è un metodo di risonanza magnetica funzionale (fMRI) utilizzato nella mappatura cerebrale per valutare le interazioni tra regioni che avvengono in uno stato di riposo.

3.2 - Analisi Preliminari

Si sono pertanto analizzate le serie storiche dell'attivazione cerebrale a riposo in 119 regioni del cervello dei 96 soggetti considerati.

Le serie storiche sono formate da 1200 punti temporali (con un intervallo costante di 0.7 secondi tra ciascun punto).

Queste sono state inizialmente centrate e standardizzate (Fig. 1).

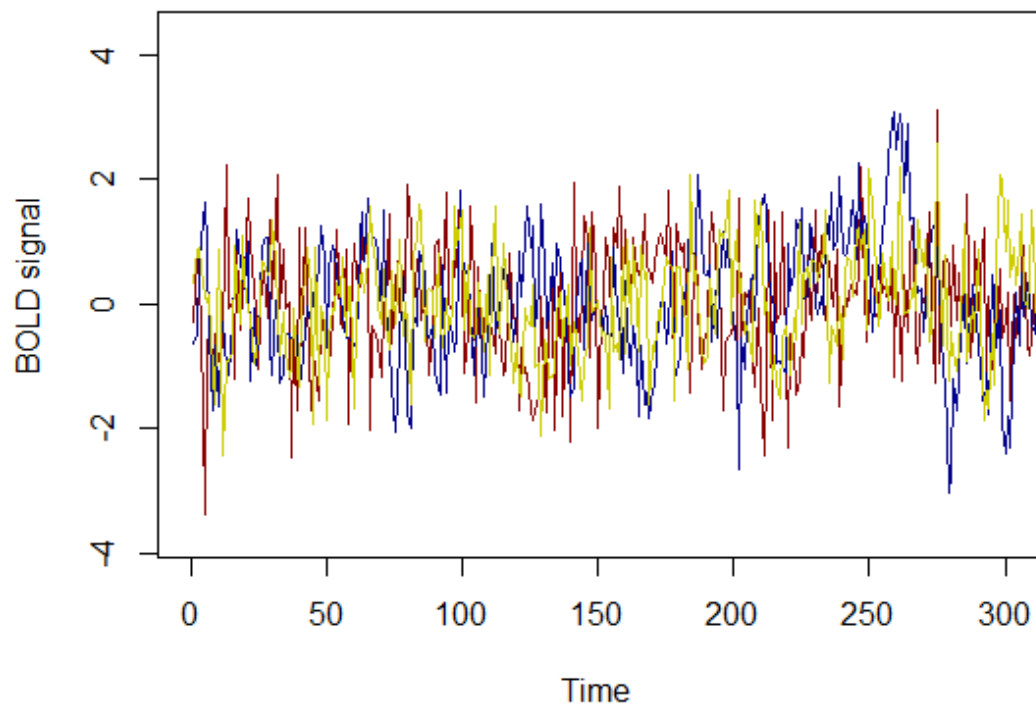


Fig.1 Rappresentazione grafica del segnale BOLD delle prime 3 regioni per il primo soggetto al variare dei 300 intervalli temporali iniziali, dopo aver standardizzato i dati. Non è stato mostrato l'intero campione dei dati per chiarezza illustrativa

Su questi dati era di primario interesse andare ad indagare la connettività funzionale (FC). Questa è stata analizzata attraverso l'approccio più comune in questo ambito : la *Sliding-Window Correlation² (SWC)*, seguita dall'analisi di clustering per andare a definire un insieme di "stati di connettività" (o stati dinamici) a riposo che si attivano e disattivano al variare del tempo.

² Si veda il paragrafo 3.3 per chiarimenti. Per una spiegazione più approfondita si rimanda a Shakil et al. "Evaluation of Sliding Window Correlation Performance for Characterizing Dynamic Functional Connectivity and Brain States."

Come prima cosa è stata calcolata la correlazione media complessiva; le 119 serie storiche delle regioni sono state concatenate tra tutti i soggetti e successivamente è stata calcolata la correlazione tra le varie coppie di regioni (Fig. 2).

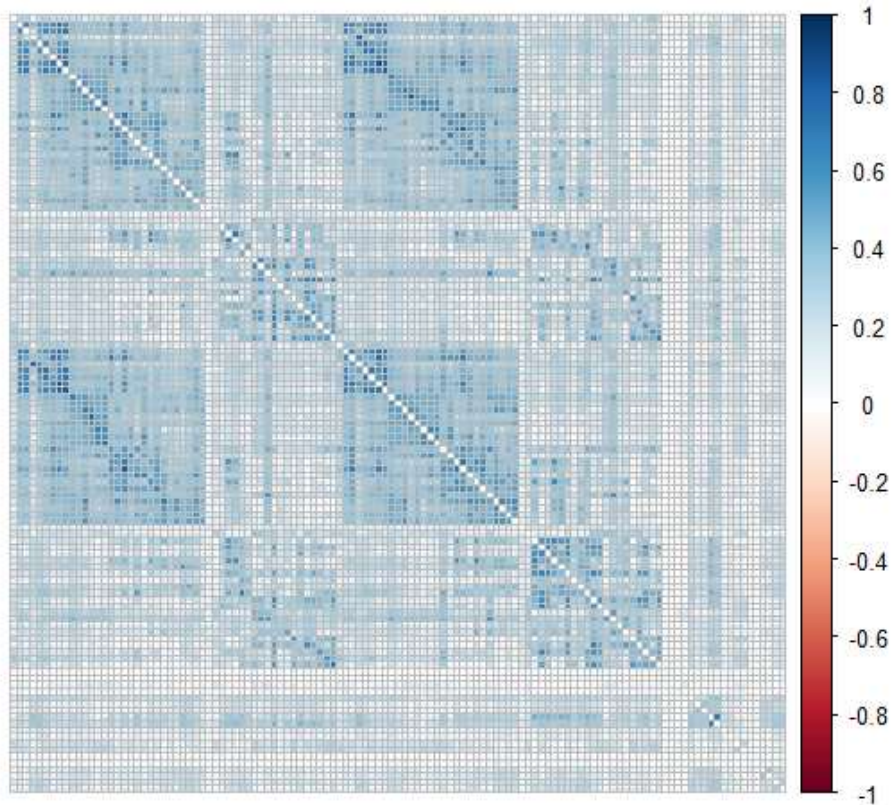


Fig.2 Il grafico è stato ottenuto tramite la funzione `corrplot.mixed()` presente nella libreria `corrplot` di R. La figura mostra la matrice di correlazione tra le 119 regioni dopo aver concatenato le serie storiche appartenenti ai 96 soggetti. L'intensità della correlazione è indicata dal colore, come mostrato dalla barra sulla destra.

Da questa matrice è possibile calcolare la correlazione media complessiva, ottenendo un valore pari a circa 0.2455.

3.3 - Sliding Window Correlation

La *Sliding Window Correlation* (SWC) è il metodo più semplice e più utilizzato per l'analisi della connettività funzionale a riposo. Nella SWC viene selezionata una finestra temporale, con una certa grandezza (size) e una certa forma (shape), e su questa viene calcolata la correlazione tra i segnali di interesse. La finestra viene poi fatta scorrere (sull'asse temporale) di un certo valore, detto *offset* (sfasamento). Questa procedura viene ripetuta per l'intera lunghezza delle rilevazioni.

Per ciascun soggetto è stata calcolata la SWC. Le serie storiche sono state suddivise in 1117 finestre temporali (ogni finestra aveva durata di 1 minuto, corrispondente a 83 istanti temporali consecutivi) con sfasamento di un punto (0.7 s).

In ogni finestra è stata calcolata la correlazione tra le 119 regioni, ottenendo una matrice 119x119. Questa matrice è poi stata approssimata da un vettore contenente i primi 3 autovettori riscaldati per la radice quadrata del rispettivo autovalore (il vettore risultante ha quindi dimensione 357 x 1). I vettori sono stati poi concatenati ottenendo una matrice 357 x 1117 (*sliding window matrix*) che descrive l'andamento delle correlazioni tra le regioni al variare delle finestre temporali.

In Fig. 3 si mostra l'andamento dei primi 3 autovalori al variare delle finestre temporali per 6 soggetti selezionati casualmente.

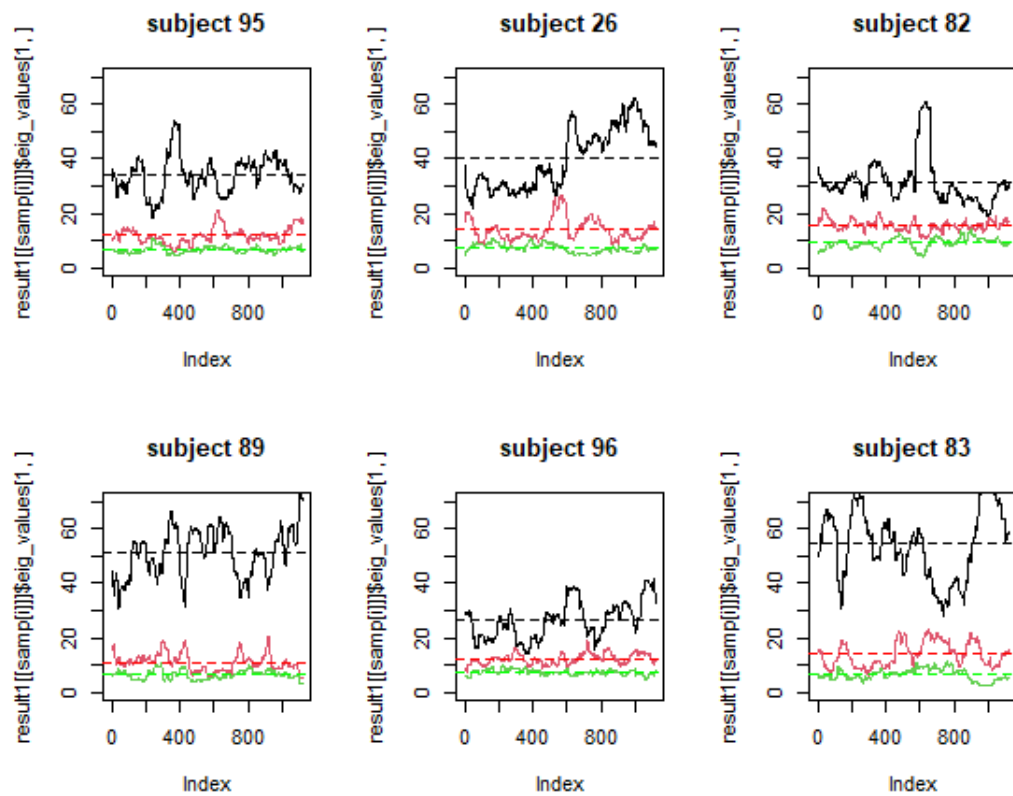


Fig.3 La figura mostra, per 6 soggetti scelti casualmente, l'andamento dei primi 3 autovalori della matrice di correlazione tra regioni al variare delle 1117 finestre temporali. In nero è riportato il primo autovalore, in rosso il secondo e in verde il terzo. La linea tratteggiata indica la media dell'autovalore del rispettivo colore.

3.4 - CLUSTERING

Durante il progetto, l'analisi di clustering è stata eseguita al fine di individuare gli stati dinamici latenti che sono comuni a tutti i soggetti sottoposti alla SWC. Per eseguire il clustering, è stato utilizzato l'algoritmo a k medie (k-means), una tecnica comune per raggruppare dati simili in cluster basati sulle loro caratteristiche (si veda per esempio Johnson, Wichern, "Applied Multivariate Statistical Analysis").

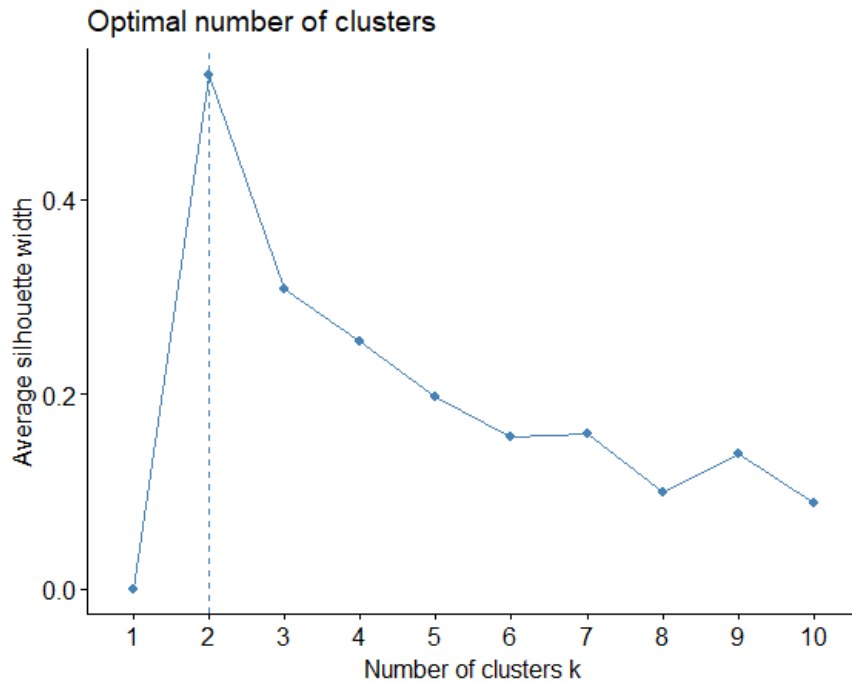
Scelta del numero di cluster:

La scelta del numero di cluster è stata una parte critica e impegnativa del processo. Per determinare il numero ottimale di cluster sono stati adottati due metodi: il metodo della silhouette e il metodo della Within-Cluster Sum of Squares (WSS). Il metodo della silhouette è finalizzato a massimizzare il coefficiente di silhouette, che misura la similarità di un punto dati all'interno del suo cluster (coesione) rispetto agli altri cluster (separazione). Una silhouette elevata indica che il punto è stato assegnato in modo corretto al cluster. Il secondo metodo, WSS, valuta la variabilità all'interno dei cluster al variare del numero di cluster utilizzati e cerca il punto in cui il guadagno marginale di variabilità spiegata non è significativo, indicando il numero ottimale di cluster.

Per eseguire l'analisi, le *sliding window matrices* di tutti e 96 soggetti sono state concatenate. Questo ha generato una matrice complessa con dimensioni di 357 righe e 107232 colonne. Poiché lavorare con tutte le finestre sarebbe stato computazionalmente oneroso, le finestre sono state sottocampionate, selezionandone una ogni 10.

I risultati ottenuti da queste analisi sono stati poi rappresentati tramite grafici, consentendo di esaminare e confrontare le diverse configurazioni di cluster e identificare il numero ottimale di cluster che meglio rappresenta gli stati dinamici latenti comuni ai soggetti sottoposti alla SWC.

(a)



(b)

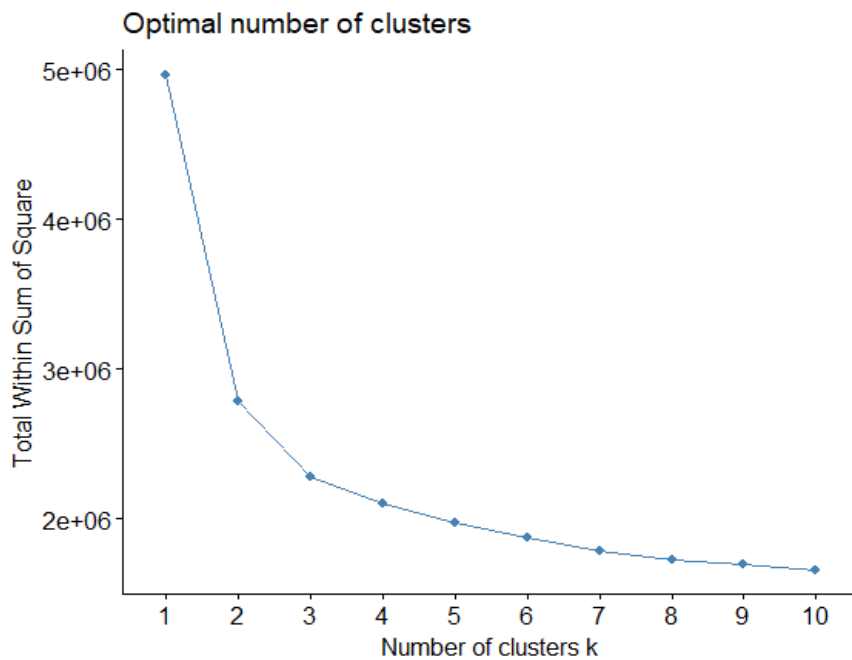


Fig. 4 I grafici mostrano i risultati dei due algoritmi utilizzati per la scelta del numero ottimale di cluster.

(a) La figura mostra l'andamento del coefficiente di silhouette al variare del numero di cluster.

(b) In figura viene mostrato l'andamento della variabilità totale all'interno dei gruppi al variare del numero di cluster

Il primo grafico (Fig. 4a), generato attraverso il metodo della silhouette, indica chiaramente il valore 2 è come il numero ottimale di cluster. Questo suggerisce che esistono due distinti raggruppamenti di dati con caratteristiche simili, ciascuno rappresentante uno stato dinamico specifico comune ai soggetti della SWC.

D'altra parte, il secondo grafico (Fig. 4b), ottenuto tramite il metodo della WSS, suggerisce invece 3 come il numero di cluster ottimale³, indicando che l'aggiunta di un terzo cluster potrebbe spiegare in modo significativo ulteriore variabilità all'interno dei dati.

Le analisi sono state poi ripetute sulla matrice completa dei dati e su un'ulteriore matrice ottenuta concatenando le matrici risultanti dalla SWC, stavolta approssimando le matrici di correlazione non con gli autovalori ma con la vettorizzazione della matrice diagonale superiore. In entrambi i casi, i risultati ottenuti sono stati confrontati con quelli precedenti e si sono rivelati comparabili, confermando quanto visto precedentemente.

Alla luce di queste analisi e delle evidenze raccolte, come scelta finale per il numero ottimale di cluster, è stato stabilito che 3 rappresenta il numero di stati dinamici più significativo e rappresentativo per il gruppo di soggetti esaminato nella SWC.

³ Si indica 3 come numero di cluster ottimale poiché si è andati a cercare il punto in cui l'aggiunta di un ulteriore cluster non riduce significativamente il WSS. Questo è il punto in cui i cluster iniziano a catturare la variazione all'interno dei dati in modo più efficace. Questo metodo è anche detto "metodo del gomito" perché si cerca il punto in cui il grafico assume una forma a gomito.

3.5 - INDICI DI CLUSTER

Dopo aver completato le analisi di clustering utilizzando l'algoritmo k-means per identificare gli stati dinamici comuni nelle serie SWC, sono stati calcolati alcuni indici di clustering significativi e utili per l'interpretazione dei risultati, come effettuato in *Favaretto et al. "Subcortical-cortical dynamical states of the human brain and their breakdown in stroke"*. Tra questi indici figurano il *fraction time*, la *transition probability* e il *dwell time*.

Il *fraction time* è una misura che indica la proporzione del tempo trascorso in ciascun cluster rispetto al tempo totale di osservazione. Questo indice permette di valutare quanto un soggetto tende a rimanere all'interno di uno specifico stato dinamico durante l'intera registrazione dei dati. Una frazione di tempo elevata per un particolare cluster suggerisce che il soggetto tende a trascorrere più tempo in quel particolare stato dinamico.

La *transition probability*, invece, fornisce un'indicazione della probabilità che un soggetto passi da uno stato dinamico ad un altro. Questo indice misura la facilità con cui un individuo transita da un cluster all'altro nel corso del tempo. Un valore elevato di *transition probability* tra due cluster indica una maggiore probabilità di passaggio tra quegli stati dinamici.

Infine, il *dwell time* rappresenta la durata media di permanenza di un soggetto in uno specifico cluster prima di passare a un altro. Questa misura consente di valutare la stabilità e la persistenza di ciascuno stato dinamico. Un *dwell time* più lungo suggerisce che il soggetto rimane in uno stato dinamico per un periodo prolungato prima di spostarsi verso un diverso stato.

-RISULTATI:

Gli indici di clustering sopra descritti sono stati calcolati sul clustering effettuato in precedenza. Il *fraction time* mostra un cluster molto più presente rispetto agli altri due. In particolare vediamo che circa il 66% delle osservazioni viene associato al cluster 3, il 21% viene associato al cluster 1 e il restante 13% al secondo cluster.

Fraction time:

Cluster 2	Cluster 1	Cluster 3
0.1335	0.2116	0.6549

Questo suggerisce la presenza di uno stato dinamico prevalente, in cui il cervello tende a passare la maggior parte del suo tempo nello stato di riposo.

La matrice qua sotto riporta la probabilità di transizione tra i diversi cluster (e quindi stati dinamici).

Transition probability:

	C1	C2	C3
C1	0.9051	0.0751	0.0198
C2	0.1191	0.7379	0.1430
C3	0.0064	0.0292	0.9644

La matrice si legge per righe e indica la probabilità, partendo da un'osservazione associata ad un cluster (indicato in riga), di passare ad un'altra osservazione associata ad un cluster (colonna). Per esempio 0.1430 rappresenta la probabilità di passare da un'osservazione associata al secondo cluster ad un'osservazione associata al terzo.

Come prima cosa si nota la tendenza di avere osservazioni consecutive che appartengono allo stesso cluster. Questo indica che, come già noto in letteratura, gli stati dinamici in generale tendono a mantenersi per un certo periodo di tempo.

I dwell times ci mostrano come la permanenza all'interno dei cluster cambia molto a seconda del cluster in esame.

Dwell time:

Cluster 1	Cluster 2	Cluster 3
10.536925	3.815881	28.099720

Si deduce che lo stato dinamico prevalente, quello associato al terzo cluster, oltre ad essere il più frequente è anche il più stabile, ovvero i soggetti tendono a restare piuttosto a lungo all'interno di questo stato. Anche lo stato dinamico associato al primo cluster si mantiene discretamente nel tempo, mentre il secondo stato tende ad essere estremamente volubile. L'indice di dwell time pari a 3.81 indica infatti che in media i soggetti tendono a restare in quello stato per sole 4 finestre temporali consecutive prima di passare ad uno stato successivo.

Il cluster numero 2 suggerisce uno stato dinamico poco frequente e molto transitorio, che quindi presenta caratteristiche molto simili a quelle del rumore. Questo dato è congruente col fatto che il metodo della silhouette suggeriva l'uso di soli 2 cluster.

4 - MODELLI NULLI

4.1 - Introduzione ai modelli nulli

La quantità di dati sul cervello a cui si può avere accesso nell'ultimo periodo è davvero incredibile. E come detto in precedenza, grazie all'innovazione tecnologica e al perfezionamento degli strumenti, è possibile andare ad indagare caratteristiche sempre più peculiari del nostro cervello. Tuttavia, aumentando la mole dei dati, aumenta anche la necessità di una convalida statistica dei risultati.

In particolare, quando si ricostruiscono reti cerebrali (ad esempio reti di connettività funzionale), utilizzando metodi provenienti dalla scienza delle reti si possono calcolare determinate statistiche per testare alcune loro caratteristiche.

Come si può mostrare se una caratteristica della rete cerebrale è più significativa di quanto ci aspetterebbe in una situazione standard? Questa caratteristica è centrale per i nostri interessi o è il risultato di una combinazione di altre caratteristiche? E può questa caratteristica essere attribuita ad un singolo meccanismo generatore?

La difficoltà per i neuroscienziati sta proprio nello stabilire quanto statisticamente uniche e particolari sono le caratteristiche della rete cerebrale in esame.

Per rispondere a queste domande e a tante altre la *network neuroscience* si affida a sofisticati metodi di statistica inferenziale noti anche con il nome di *null models* (o modelli nulli).

Questi modelli servono per garantire l'affidabilità e la generalizzabilità dei risultati ottenuti nella fase di analisi (nel nostro caso l'identificazione degli stati dinamici comuni a diversi soggetti).

Con modelli nulli, nell'ambito delle neuroscienze, si intendono realizzazioni sintetiche di reti cerebrali usate per verificare la particolarità di alcune caratteristiche della rete originaria.

I modelli nulli sono così fondamentali per l'inferenza sulle reti che alcuni degli indici descrittivi di rete più utilizzati hanno dei modelli nulli all'interno della loro definizione. Tra i vari indicatori disponibili consideriamo il coefficiente di modularità Q (una misura di omofilia della rete), che è definito come la frazione di archi che connettono nodi di uno stesso tipo a cui viene sottratto il valore atteso della stessa quantità per una rete con connessioni casuali.

$$Q = \sum_{k=1}^K e_{kk} - \sum_{k=1}^K a_k^2$$

e_{ij} è il numero di archi che hanno inizio nel gruppo i e fine nel gruppo j , sul totale degli archi della rete, mentre a_i è il numero di archi che hanno inizio o fine nel gruppo i sul totale degli archi della rete

La statistica Q deve quindi essere definita sulla base di un modello nullo che quantifichi il valore atteso di connessioni tra nodi simili in una rete generica, dove le connessioni *non* sono preferenzialmente formate all'interno dei gruppi

Nella neuroscienza delle reti vi sono due principali metodi per generare modelli nulli. Il principio alla base è sempre quello di generare reti sintetiche sotto un'ipotesi nulla, randomizzando alcune caratteristiche osservate nella rete originale e mantenendo costanti altre caratteristiche.

I *Randomization Models* partono dalla rete originaria e ne creano una nuova iterando modifiche casuali alla rete (il metodo più semplice ed utilizzato è quello del *rewiring*, ovvero scegliendo randomicamente due nodi e scambiandoli ad ogni iterazione)

I *Generative Models* creano una nuova rete da zero seguendo un piccolo insieme di regole restrittive.

Ci si rifaccia a *Váša et al., 31 May 2022, "Null Models in Network Neuroscience"* per ulteriori approfondimenti.

Nel nostro caso è stato utilizzato un approccio di randomizzazione.

A seconda del tipo di rete cerebrale a disposizione e delle caratteristiche da studiare vi sono molteplici modelli che si possono utilizzare per generare surrogati dei dati originali col fine di studiarne alcune caratteristiche. Nel prossimo sottocapitolo verrà presentato uno dei modelli nulli più usati nell'ambito delle reti di correlazione.

Nonostante i modelli nulli siano la spina dorsale delle metodologie statistiche nella *network neuroscience*, questi sono soggetti a considerevoli controindicazioni. Le relazioni tra diverse proprietà delle reti possono rendere difficile, o addirittura impossibile, separare il contributo unico di ciascuna proprietà.

In particolare, la presenza di alcune caratteristiche può indurre la presenza di altre caratteristiche, rendendo impossibile randomizzare selettivamente una caratteristica mantenendo fisse le altre.

Più in generale, vista la loro grande duttilità, bisogna porre particolare attenzione alle restrizioni imposte dall'ipotesi del modello nullo. Modelli poco restrittivi che impongono pochi vincoli rischiano di dare luogo a considerazioni banali e che non indicano chiaramente l'origine dei fenomeni di interesse.

Al contrario, modelli troppo restrittivi che controllano troppe caratteristiche della rete possono portare a risultati troppo strettamente connessi con la struttura empirica di base e quindi lasciare poco spazio per la generalizzazione

Una considerazione correlata è quanto bene i modelli nulli effettivamente campionino lo spazio che cercano di esplorare. In altre parole, i modelli nulli possono campionare sistematicamente in modo insufficiente o eccessivo determinate aree dello spazio target.

Le randomizzazioni in generale coprono un'area più ristretta dello spazio campionato dai modelli nulli. I modelli di tipo generativo risolvono questo problema ma aumentano notevolmente la variabilità osservata.

Infine la diversità dei modelli nulli comporta anche difficoltà pratiche nel loro utilizzo. Le implementazioni algoritmiche di un modello nullo spesso implicano ipotesi specifiche, che possono tradursi in differenze non banali nei risultati. Questa variabilità tra diverse implementazioni dello stesso modello nullo può portare a differenze nell'interpretazione.

4.2 - Phase Randomization

Nel caso di dati provenienti da *Correlation Networks*, o in generale dove la connettività è misurata come correlazione tra diverse regioni spaziali del cervello, la generazione di dati surrogati tramite modelli nulli più semplici (come ad esempio il *rewiring* accennato in precedenza) potrebbero non essere adeguati. Infatti reti costruite sulla base di correlazioni seguono una proprietà di tipo transitivo: conoscendo il valore di un arco A-B e di quello A-C possiamo infatti stabilire un limite al valore dell'arco B-C. In questo caso scambiare casualmente tra loro gli archi potrebbe rompere questa importante proprietà (si potrebbe arrivare ad avere correlazione positiva tra A e B e tra A e C, ma negativa tra B e C).

Pertanto reti di correlazione richiedono dei modelli nulli di tipo diverso per gestire al meglio la randomizzazione. Le opzioni sono molteplici e in letteratura sono stati esaminati molti metodi a seconda della situazione⁴.

Alcuni metodi, diversamente da quanto descritto in precedenza, lavorano direttamente effettuando la randomizzazione sul segnale osservato invece che sulla struttura della rete.

In questo progetto è stato deciso di utilizzare un modello *Phase Randomization* (come in Liégeois et al. "*Interpreting temporal fluctuation in resting-state functional connectivity MRI*").

I metodi di *phase randomization* (PR) per generare dati surrogati vengono utilizzati da decine di anni nell'ambito della fisica e ultimamente sono stati applicati anche a dati di fMRI in diversi importanti articoli sull'analisi della connettività funzionale.

La procedura PR genera "dati nulli" mediante 3 passaggi:

- Viene eseguita una DFT (*Discrete Fourier Transformation*) per ciascuna serie temporale. Data la serie x_n la sua trasformata discreta di Fourier corrisponde a:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn}$$

- Ad ogni istante temporale viene aggiunta una fase random da una distribuzione uniforme in $[0, 2\pi]$
- Infine viene effettuata la trasformazione inversa della DFT:

⁴ I metodi di randomizzazione più utilizzati sono l'*Autoregressive Randomization* (ARR) e il *Phase Randomization* (PR). Si rimanda a Liégeois et al. "*Interpreting Temporal Fluctuations in Resting-State Functional Connectivity MRI*" per approfondimenti

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N} kn}$$

Questa procedura genera serie temporali stazionarie che mantengono la struttura di correlazione tra le varie regioni del cervello: a tale scopo è necessario che la medesima fase random sia aggiunta a tutti segnali analizzati dalle diverse regioni cerebrali.

Nel seguito verranno spiegati i passi effettuati per ottenere i dati PR a partire dai dati originali, le analisi effettuate e infine le conclusioni tratte.

4.3 - Analisi sul Dataset PR

Per ottenere il dataset PR sono state prese le 96 serie storiche multivariate relative ai 96 soggetti. Per ciascuna di queste è stata applicata una trasformazione di Fourier ad ognuna delle 119 serie storiche usando la funzione `fft()` presente nella libreria base del software R⁵. Successivamente, per ogni intervallo temporale è stata aggiunta una fase aleatoria generata da una distribuzione uniforme in $[0, 2\pi]$. Questa operazione è stata fatta in modo da tenere costanti le fasi randomiche tra le 119 serie temporali. Infine alle serie storiche risultanti, normalizzate per la lunghezza delle stesse, è stata applicata la trasformazione inversa di Fourier.

Il dataset ottenuto ha le stesse dimensioni di quello originale ed è stata preservata la matrice di correlazione tra le varie regioni rimuovendo viceversa possibili legami temporali tra le varie osservazioni.

I grafici seguenti (Fig. 5) mostrano, per due soggetti, che la correlazione tra le varie regioni è rimasta la stessa anche dopo la randomizzazione:

⁵ <https://www.r-project.org/>

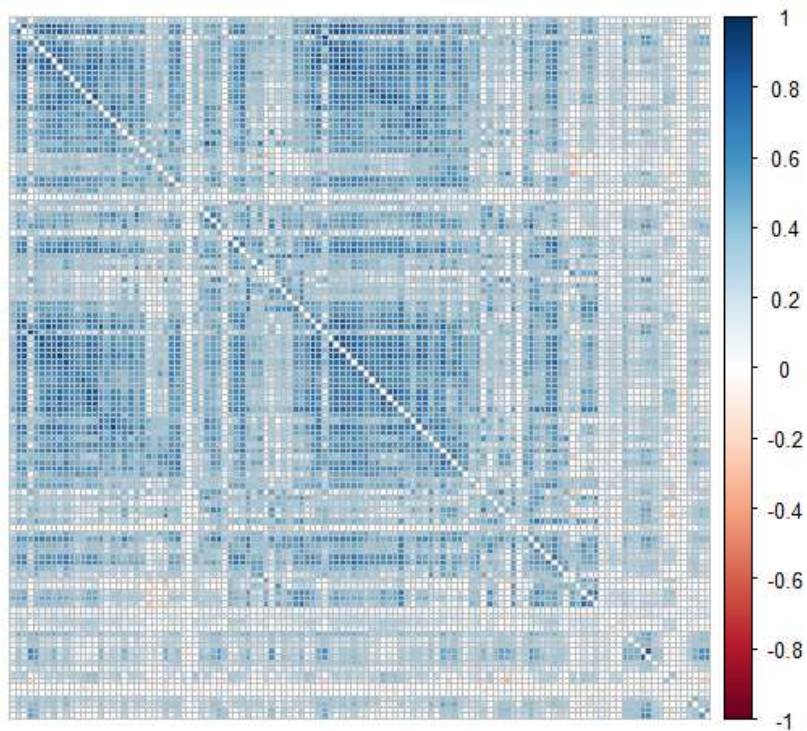


Fig. 5a Rappresentazione grafica della correlazione tra le serie storiche delle 119 regioni cerebrali per il soggetto 18

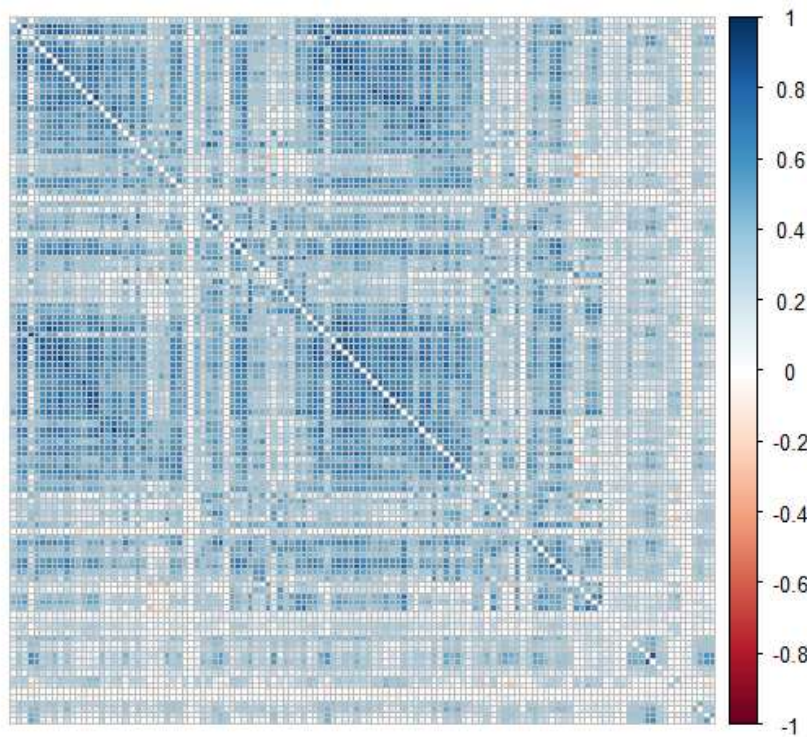


Fig. 5b Rappresentazione grafica della correlazione tra le serie storiche randomizzate eseguendo una *phase randomization* a partire dal soggetto 18

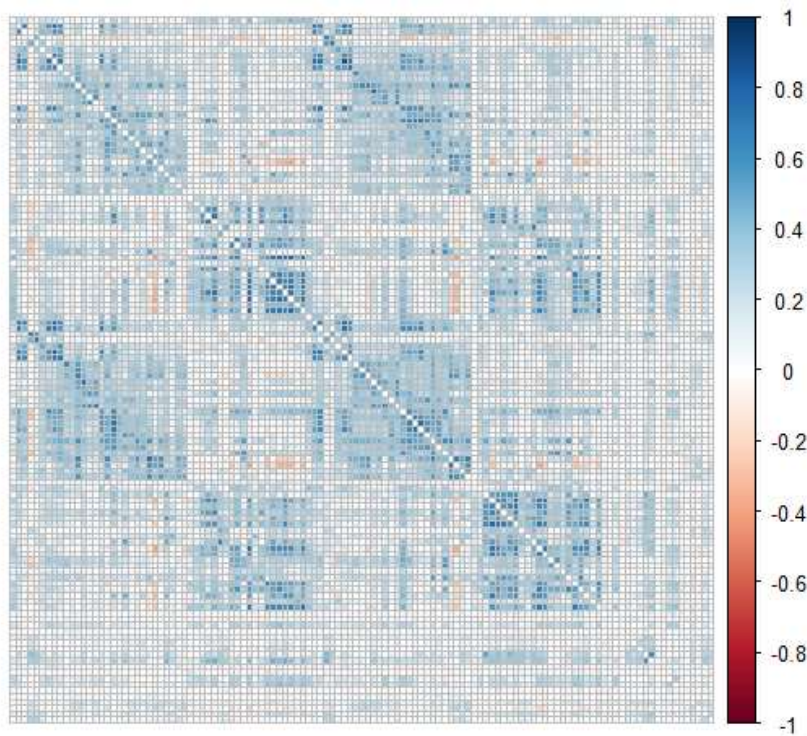


Fig. 5c Rappresentazione grafica della correlazione tra le serie storiche delle 119 regioni cerebrali per il soggetto 42

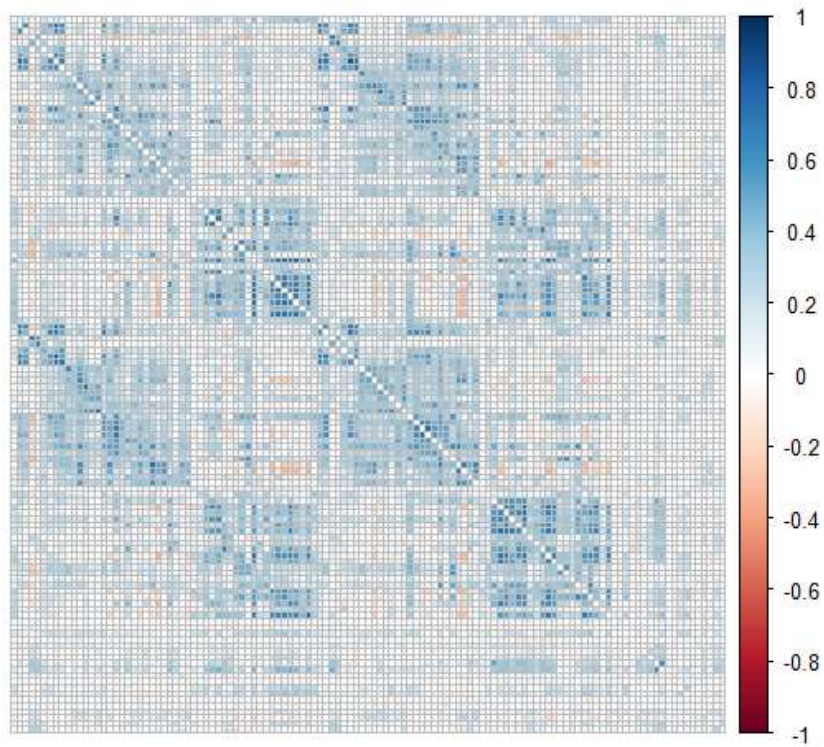


Fig. 5d Rappresentazione grafica della correlazione tra le serie storiche randomizzate eseguendo una *phase randomization* a partire dal soggetto 42

Le analisi descritte in precedenza nel capitolo 3 sono quindi state ripetute sui dati PR.

Il calcolo della correlazione media, come ci si aspettava per costruzione, risulta analogo a quello dei dati reali, ottenendo un valore pari a 0.2458.

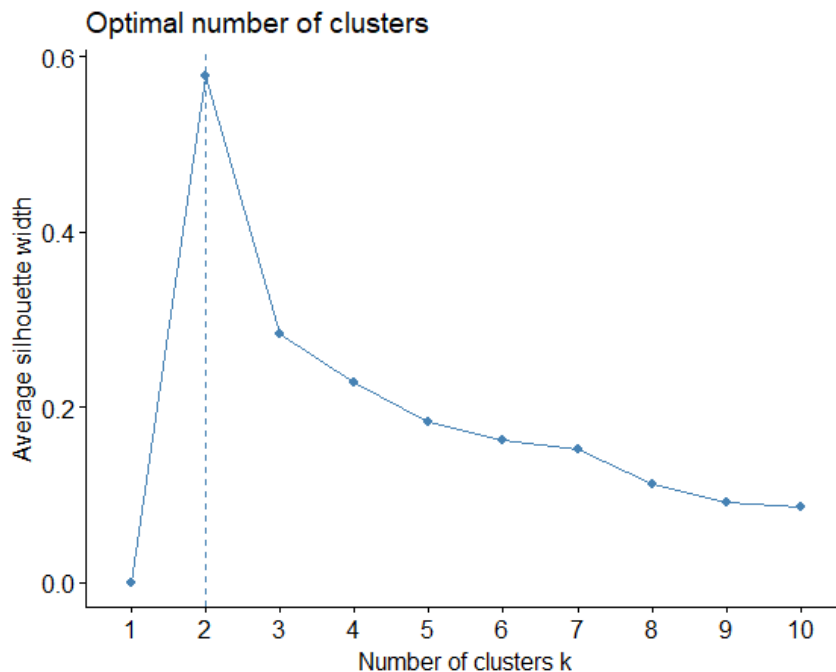
Ad ogni matrice di pseudo dati è stata poi effettuata una SWC coerentemente con quanto fatto in precedenza.

Il risultato sono quindi 96 matrici di dimensione 357×1117 dove 1117 sono il numero di finestre temporali analizzate e le rispettive colonne sono un'approssimazione delle matrici di correlazione tra le regioni in ogni finestra temporale.

Queste matrici sono poi state concatenate tra loro per riga al fine di effettuare le analisi di clustering.

Come in precedenza anche qui la scelta del numero ottimale di cluster ha richiesto del tempo. Interessante è che sia il metodo della silhouette (Fig. 6a) che quello della wss (Fig. 6b) diano risultati molto simili a quelli dei dati reali.

Il metodo della silhouette suggerisce 2 i maniera abbastanza marcata:



Mentre il metodo della WSS mostra indecisione tra 2 e 3:

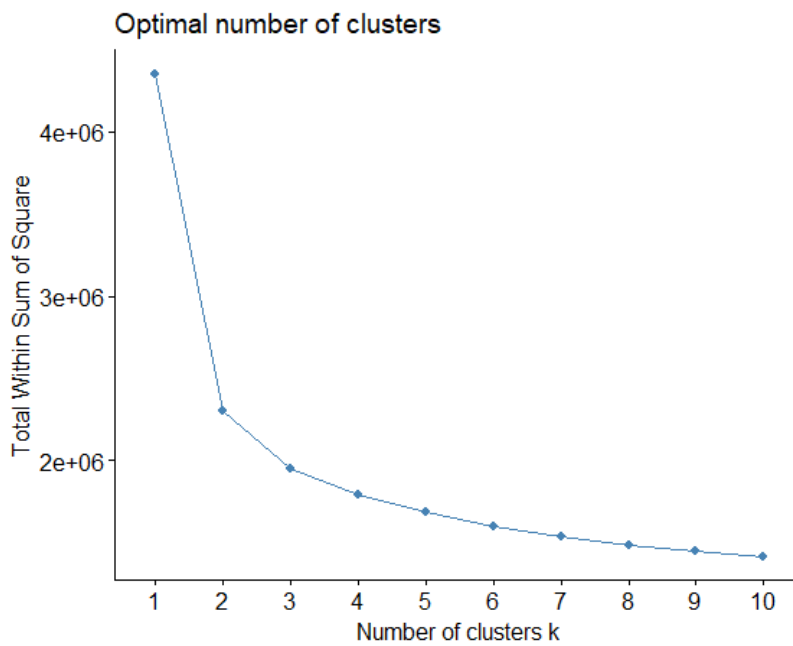


Fig. 6 I grafici mostrano i risultati dei due algoritmi utilizzati per la scelta del numero ottimale di cluster per i dati PR.

(a) La figura mostra l'andamento del coefficiente di silhouette al variare del numero di cluster.

(b) In figura viene mostrato l'andamento della variabilità totale all'interno dei gruppi al variare del numero di cluster

Per analogia e semplicità poi nel confronto dei risultati anche questa volta è stato scelto 3 come numero di cluster ottimale.

Qui in seguito vengono riportati i risultati ottenuti su R per gli indici di cluster calcolati sugli pseudo dati phase randomized. Gli indici sono gli stessi usati anche in precedenza. Per le definizioni si riguardi il paragrafo 3.5.

Fraction Time:

Cluster 2	Cluster 3	Cluster 1
0,1165	0.1984	0.6851

Transition probability:

	C1	C2	C3
C1	0.9658	0.0281	0.0061
C2	0.1648	0.7214	0.1138
C3	0.0211	0.0668	0.9121

Dwell Time:

Cluster 1	Cluster 2	Cluster 3
29.280590	3.588742	11.380417

Guardando gli indici di cluster, e ricordando che si suppone una corrispondenza tra i cluster e degli stati dinamici latenti, si nota la presenza di uno stato dinamico prevalente (corrispondente al primo cluster) a cui sono associate il 68% circa delle finestre temporali. Come si evince dalla *transition probability* e dal *dwell time* questo stato dinamico tende ad essere anche piuttosto stabile nel tempo.

Un secondo stato dinamico, corrispondente al terzo cluster, ha una presenza meno impattante (solo il 20% circa delle finestre temporali vengono associate a questo stato) e ha anche una tendenza più aleatoria nella durata.

Il terzo stato dinamico (cluster 2) è poco frequente e di natura estremamente transitoria (ha una durata media di circa sole 4 finestre temporali).

5 - CONCLUSIONI

5.1 - Confronto dei risultati

Come già probabilmente evidente nel precedente capitolo, i risultati ottenuti dal clustering nel dataset reale e in quello con dati *phase randomized* dimostrano notevole somiglianza tra di loro.

La distribuzione delle finestre nei 3 cluster mantiene più o meno le stesse proporzioni (70-20-10). Inoltre gli indici di cluster mostrano caratteristiche comparabili negli stati dinamici. Si riportano le tabelle degli indici di cluster calcolate nei capitoli precedenti per il confronto.

-Fraction Times:

Dati originari			Dati PR		
C2	C1	C3	C2	C3	C1
0.1335	0.2116	0.6549	0.1165	0.1984	0.6851

-Dwell Time:

Dati originari			Dati PR		
C1	C2	C3	C1	C2	C3
10.537	3.816	28.100	29.280	3.589	11.380

-Transition Probability:

	Dati originari			Dati PR		
	C1	C2	C2	C1	C2	C3
C1	0.9051	0.0751	0.0198	0.9658	0.0281	0.0061
C2	0.1191	0.7379	0.1430	0.1648	0.7214	0.1138
C2	0.0064	0.0292	0.9644	0.0211	0.0668	0.9121

Si nota che in entrambi i dataset lo stato dinamico predominante tende ad essere più stabile e a durare di più durante l'attivazione. Mentre lo stato meno presente ha una tendenza molto aleatoria ed è di breve durata.

E' stato anche effettuato un confronto grafico sui centroidi ottenuti dalle operazioni di clustering nei due gruppi, tra i quali è stata riscontrata una notevole similarità. Si riportano qui in seguito nella Fig.7 i grafici usati per il confronto.

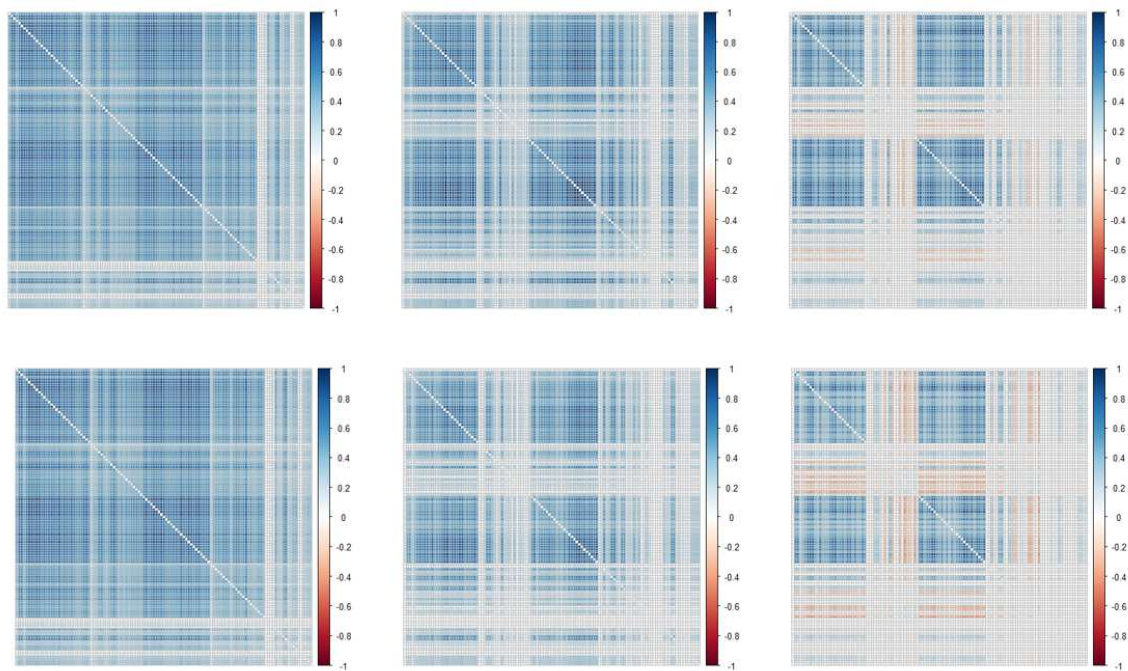


Fig.7 Il grafico mostra una rappresentazione grafica dei 3 centroidi ottenuti dal clustering effettuato sui dati PR sopra e sui dati reali sotto. La rappresentazione è stata ottenuta prendendo il vettore-centroide e ricostruendo la struttura di correlazione tra le regioni utilizzando gli autovettori.

Per dare più affidabilità ai risultati è stato effettuato un test per verificare se effettivamente i risultati nei due dataset sono statisticamente equivalenti.

Per ciascuno dei 96 soggetti sono stati presi i *fraction times* ricavati sia dai dati originali sia dai dati *phase randomized* e su questi due gruppi è stato effettuato un test di Kolmogorov-Smirnov⁶. Si riportano qui in seguito i risultati ottenuti con la funzione `ks.test()` della libreria base di R:

⁶ Il test di Kolmogorov-Smirnov è un test statistico non parametrico usato per verificare se ci sono differenze statisticamente significative tra le distribuzioni dei due campioni. Dati due campioni X_1 e X_2 , il test verifica l'ipotesi che i due campioni provengano da una stessa distribuzione contro l'ipotesi alternativa che i due campioni non abbiano lo stesso meccanismo generatore. La statistica test è calcolata come la distanza tra le funzioni di ripartizione empiriche dei due campioni.

```
> ks.test(fr_times, fr_times_pr)
```

```
Asymptotic two-sample Kolmogorov-Smirnov test
```

```
data: fr_times and fr_times_pr
```

```
D = 0.09375, p-value = 0.159
```

```
alternative hypothesis: two-sided
```

Il p-value elevato ci suggerisce che non vi è sufficiente evidenza statistica per rigettare l'ipotesi nulla di equivalenza tra i dati originali e quelli *phase randomized*. In altre parole, i risultati ottenuti nei due dataset non sono statisticamente differenti.

5.1 - Considerazioni finali

Come si è visto nel paragrafo precedente, i risultati del clustering ottenuti nel dataset reale mostrano un notevole grado di concordanza con quelli ottenuti sui dati *phase randomized*. Questo mostra che la struttura di cluster dei dati reali è presente anche nel modello con randomizzazione delle fasi, dando conferma alle analisi effettuate nel capitolo 3.

Ricordando che la *phase randomization* genera dati lineari, stazionari e gaussiani che mantengono la struttura di correlazione dei dati reali, si possono trarre informazioni interessanti sulla natura della *functional connectivity* e degli stati dinamici.

La presenza dei cluster individuati nel capitolo 3 potrebbe portare a pensare che l'organizzazione della connettività cerebrale evolva attraverso un insieme di stati metastabili, che tipicamente emergono in modelli non lineari e/o non stazionari.

Tuttavia, la presenza di cluster anche nel modello di randomizzazione delle fasi potrebbe suggerire che questa interpretazione non sia necessariamente veritiera. In altre parole, l'idea che vi siano stati dinamici sembra poter essere compatibile con un modello lineare e stazionario in cui la connettività funzionale non è soggetta a transizioni discrete ma a variazioni continue.

6 - BIBLIOGRAFIA

- Battaglia, Demian, et al. "Dynamic Functional Connectivity between Order and Randomness and Its Evolution across the Human Adult Lifespan." *NeuroImage*, vol. 222, 15 Nov. 2020, p. 117156, www.sciencedirect.com/science/article/pii/S105381192030642X, <https://doi.org/10.1016/j.neuroimage.2020.117156>. Accessed 10 Aug. 2023.
- Betzel, Richard F., and Danielle S. Bassett. "Generative Models for Network Neuroscience: Prospects and Promise." *Journal of the Royal Society Interface*, vol. 14, no. 136, Nov. 2017, p. 20170623, <https://doi.org/10.1098/rsif.2017.0623>. Accessed 23 Feb. 2023.
- Brunner, Manuela I, et al. "Technical Note: Stochastic Simulation of Streamflow Time Series Using Phase Randomization." *Hydrology and Earth System Science*, vol. 23, no. 8, 2 Aug. 2019, pp. 3175–3187, <https://doi.org/10.5194/hess-23-3175-2019>. Accessed 17 July 2023.
- Cao, Zhu, et al. "Discrete-Phase-Randomized Coherent State Source and Its Application in Quantum Key Distribution." *New Journal of Physics*, vol. 17, no. 5, 13 May 2015, pp. 053014–053014, <https://doi.org/10.1088/1367-2630/17/5/053014>. Accessed 11 Aug. 2023.
- Favaretto, Chiara, et al. "Subcortical-Cortical Dynamical States of the Human Brain and Their Breakdown in Stroke." *Nature Communications*, vol. 13, no. 1, 29 Aug. 2022, <https://doi.org/10.1038/s41467-022-32304-1>.
- Fischl, B. (2012). FreeSurfer. *Neuroimage*, 62(2), 774-781.

- Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage*. 2013;80:105–124.

- Raphaël Liégeois, et al. “Interpreting Temporal Fluctuations in Resting-State Functional Connectivity MRI.” *NeuroImage*, vol. 163, 1 Dec. 2017, pp. 437–455, <https://doi.org/10.1016/j.neuroimage.2017.09.012>. Accessed 23 Apr. 2023.

- Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X. N., Holmes, A. J., ... & Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral cortex*, 28(9), 3095-3114.

- Shakil, Sadia, et al. “Evaluation of Sliding Window Correlation Performance for Characterizing Dynamic Functional Connectivity and Brain States.” *NeuroImage*, vol. 133, June 2016, pp. 111–128, <https://doi.org/10.1016/j.neuroimage.2016.02.074>. Accessed 4 Apr. 2019.

- Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, et al. The WU-Minn human connectome project: an overview. *Neuroimage*. 2013; 80:62–79

- Stephan, K. E., and K. J. Friston. “Functional Connectivity.” *ScienceDirect*, Academic Press, 1 Jan. 2009, www.sciencedirect.com/science/article/pii/B9780080450469003089. Accessed 2 Apr. 2023.

- Váša, František, and Bratislav Mišić. “Null Models in Network Neuroscience.” *Nature Reviews Neuroscience*, vol. 23, no. 8, 31 May 2022, pp. 493–504, <https://doi.org/10.1038/s41583-022-00601-9>. Accessed 1 Dec. 2022.