



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ**

**ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ  
ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**ΤΟΥ**

**ΓΑΛΛΟΥ ΙΩΑΝΝΗ**

Προπτυχιακού Φοιτητή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών  
Ε.Μ.Π.

**Ανάλυση λειτουργικής διασυνδεσιμότητας  
εγκεφαλικών σημάτων από Λειτουργική Απεικόνιση  
Μαγνητικού Συντονισμού (fMRI) με τη χρήση  
Ανάλυσης Ανεξάρτητων Συνιστωσών (ICA) και  
Απεικόνισης Ισομετρικών Χαρακτηριστικών  
(ISOMAP)**

*Τριμελής Εξεταστική Επιτροπή*

Σιέττος Κωνσταντίνος, Αναπληρωτής Καθηγητής Ε.Μ.Π. (Επιβλέπων)

Γκούσης Δημήτριος, Καθηγητής Ε.Μ.Π.

Νικόλαος Σμυρνής, Αναπληρωτής Καθηγητής Ιατρικής Ε.Κ.Π.Α.

Αθήνα, Φεβρουάριος 2016

## Ευχαριστίες

---

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της Σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών Κωνσταντίνο Σιέττο, επειδή μου έδωσε τη δυνατότητα πραγματοποίησης της εργασίας αυτής και με καθοδήγησε μεθοδικά στην προσπάθεια μου και τους προβληματισμούς μου. Επίσης, θα ήθελα να ευχαριστήσω την οικογένεια μου που με στήριξε όλα αυτά τα χρόνια υλικά και πνευματικά αλλά και τους φίλους και συμφοιτητές μου που πάντα στέκονταν δίπλα μου σε όλη τη διάρκεια των σπουδών μου .

## Περίληψη

---

Στην εν λόγω διπλωματική εργασία θα πραγματοποιηθεί ανάλυση λειτουργικής διασυνδεσιμότητας εγκεφαλικών σημάτων με χρήση ανάλυσης ανεξάρτητων συνιστωσών (ICA) και απεικόνισης ισομετρικών χαρακτηριστικών (ISOMAP). Αρχικά θα γίνει μία εισαγωγή σε δεδομένα Νευροαπεικόνισης και στη συνέχεια η επικέντρωση σε έναν συγκεκριμένο τύπο δεδομένων, από λειτουργική απεικόνιση μαγνητικού συντονισμού (fMRI). Στο 2<sup>ο</sup> κεφάλαιο της εργασίας υπάρχει μία παρουσίαση μεθόδων τόσο γραμμικών όσο και μη γραμμικών για την ανάλυση δεδομένων fMRI. Στις γραμμικές μέθοδοι περιλαμβάνονται μεταξύ άλλων η ανάλυση σε ανεξάρτητες συνιστώσες (ICA) καθώς και η ανάλυση σε κύριες συνιστώσες (PCA) ενώ η μη γραμμική μέθοδος είναι η απεικόνιση ισομετρικών χαρακτηριστικών (ISOMAP). Στο 3<sup>ο</sup> και κύριο μέρος της εργασίας θα αναλυθούν τα βήματα και οι μέθοδοι που ακολουθήθηκαν για την ανάλυση λειτουργικής διασυνδεσιμότητας πραγματικών δεδομένων σε ένα πληθυσμό 146 ατόμων από τα οποία οι 72 είναι ασθενείς διαγνωσμένοι με σχιζοφρένεια και 74 από αυτούς είναι υγιείς. Μετά την ανάλυση, στο 4<sup>ο</sup> κεφάλαιο θα παρουσιαστούν τα αποτελέσματα και θα επιχειρηθεί μία ερμηνεία της σημασίας τους. Στο 5<sup>ο</sup> μέρος θα εκτιμηθούν και θα αξιολογηθούν τα αποτελέσματα της ανάλυσης και θα συζητηθούν παράγοντες που θα μπορούσαν να βελτιώσουν τα τελευταία, θα επισημανθούν ερευνητικοί στόχοι και θα αναδειχθούν περιοχές για περαιτέρω ανάλυση. Στο 6<sup>ο</sup> κεφάλαιο παρατίθεται όλη η σχετική βιβλιογραφία με όλες τις σχετικές αναφορές στο κυρίως κείμενο σε αλφαβητική σειρά. Τέλος, στο Παράρτημα θα παρατεθεί ο κώδικας που χρησιμοποιήθηκε για το υπολογιστικό κομμάτι της εργασίας, για το οποίο έγινε χρήση των στατιστικών πακέτων FSL για ανάλυση δεδομένων Νευροαπεικόνισης και R για στατιστικές εφαρμογές, μαζί με σχόλια για τη διευκόλυνση του αναγνώστη.

## Abstract

---

In this thesis we present functional connectivity analysis on functional Magnetic Resonance Imaging (fMRI) data. For the analysis we used Independent component analysis (ICA) and Isometric feature mapping (ISOMAP) as well. In the first place, as an introduction, we discuss Neuroimaging data and then we focus on a special type, the fMRI data. At the 2<sup>nd</sup> chapter there is a presentation of linear and non-linear methods useful for fMRI analysis. Among other methods we present the linear methods Independent Component Analysis (ICA) and Principal Component Analysis (PCA) and the one which is non-linear, the isometric-feature mapping (ISOMAP). At the 3<sup>rd</sup> and main chapter of this thesis, we analyze the steps and the methods followed for the functional connectivity analysis of a real statistical sample consisted of 146 subjects of which 72 were schizophrenia patients and the other 74 were healthy controls. The next chapter represents results taken from the analysis and a willing try for an interpretation of them. At the 5<sup>th</sup> place all results and interpretation will be evaluated, and other characteristics and factors will be discussed, in a way that they could enhance our analysis. New scientific goals will be set and new areas-of-interest will be indicated for further reading. At the sixth part, the relevant citations of all references across the entire thesis are set in an alphabetical order for the reader's convenience. Finally, there is an Appendix at the end of the book. There, lies the code we used for the computational part of our analysis, which took place on Linux environment using FSL, a package special for Neuroimaging data analysis and R, a package mainly for statistical applications. Along with the code, there are useful comments on how someone can make it work along with understanding the main concept.

# Περιεχόμενα

---

## Κεφάλαιο 1

<b>Εισαγωγή</b> .....	7
<b>1.1</b> Νευροαπεικόνιση.....	8
<b>1.2</b> Μέθοδος και δεδομένα FMRI.....	10
<b>1.2.1</b> Συμπεριφορά του σήματος BOLD.....	14
<b>1.2.2</b> Ποιοτικά και ποσοτικά χαρακτηριστικά των FMRI δεδομένων.....	15
<b>1.2.3</b> Επεξεργασία των δεδομένων FMRI.....	15
<b>1.2.4</b> FMRI scan(Χάρτες ενεργοποίησης).....	16
<b>1.2.5</b> Πλεονεκτήματα και Μειονεκτήματα της FMRI.....	17

## Κεφάλαιο 2

<b>Γραμμικές και μη γραμμικές μέθοδοι ανάλυσης FMRI</b> .....	<b>18</b>
<b>2.1</b> Γραμμικές μέθοδοι που βασίζονται στο μοντέλο (Model- based).....	18
<b>2.1.1</b> Ανάλυση πολλαπλής συσχέτισης (Cross-correlation Analysis).....	18
<b>2.1.2</b> Ανάλυση συνοχής (Coherence Analysis).....	19
<b>2.2</b> Γραμμικές μέθοδοι που βασίζονται στα δεδομένα και τη δομή τους (Data-driven).....	20
<b>2.2.1</b> Ανάλυση σε κύριες συνιστώσες και ανάλυση πίνακα σε ιδιάζουσες τιμές (PCA/SVD).....	21
<b>2.2.2</b> Ανάλυση σε ανεξάρτητες συνιστώσες (ICA).....	23
<b>2.2.3</b> Ανάλυση ICA σε ομάδες (group ICA).....	27
<b>2.3</b> Μη γραμμικές μέθοδοι που βασίζονται στα δεδομένα και τη δομή τους.....	31
<b>2.3.1</b> Ισομετρική απεικόνιση χαρακτηριστικών (Isometric feature mapping - ISOMAP).....	31

## Κεφάλαιο 3

<b>Ανάλυση λειτουργικής διασυνδεσιμότητας σε πραγματικά δεδομένα Λειτουργικής Απεικόνισης Μαγνητικού Συντονισμού (FMRI). ....</b>	<b>38</b>
<b>3.1</b> Παρουσίαση των δεδομένων FMRI του προβλήματος.....	38
<b>3.2</b> Μια επισκόπηση για τα FMRI δεδομένα και αναφορές για τις μέχρι σήμερα προσπάθειες ανάλυσης λειτουργικής διασυνδεσιμότητας.....	40
<b>3.3</b> Μέθοδοι και βήματα έγιναν για την ανάλυση των δεδομένων μας.....	42

3.3.1 Προ-επεξεργασία των δεδομένων.....	44
3.3.2 Ανάλυση ανεξάρτητων συνιστωσών για τα υποκείμενα.....	44
3.3.3 Κατασκευή πινάκων λειτουργικής διασυνδεσιμότητας.....	47
3.3.4 Κατασκευή γράφου με τη μέθοδο ISOMAP.....	49
3.3.5 Μέτρηση γραφοθεωρητικών στοιχείων.....	51
3.3.6 Ταξινόμηση χρησιμοποιώντας SVM ταξινομητή.....	53
3.3.7 Έλεγχοι υποθέσεων για τα γραφοθεωρητικά στοιχεία.....	54
<b>Κεφάλαιο 4</b> <b>Παρουσίαση αποτελεσμάτων.....</b>	<b>55</b>
<b>Κεφάλαιο 5</b> <b>Συμπεράσματα.....</b>	<b>61</b>
<b>Κεφάλαιο 6</b> <b>Βιβλιογραφία.....</b>	<b>63</b>
 <b>Παράρτημα.....</b>	<b>69</b>

# 1. Εισαγωγή

---

Στην παρούσα διπλωματική εργασία θα μας απασχολήσει το πρόβλημα της λειτουργικής διασυνδεσιμότητας εγκεφαλικών σημάτων. Για τη συστηματική μελέτη του προβλήματος αυτού χρησιμοποιείται γενικά η τεχνολογική μέθοδος της νευροαπεικόνισης ώστε να συγκεντρωθούν δεδομένα (πολλές φορές με πολύ μεγάλο μέγεθος) για τη μελέτη ενός ή περισσότερων τομέων του εγκεφάλου. Απώτερος σκοπός της διαδικασίας, είναι αφενός μεν να κατανοήσουμε τη σχέση κάποιας λειτουργίας (νοητικής ή αισθησιοκινητικής) με μία ή και περισσότερες εγκεφαλικές περιοχές, και αφετέρου να μελετήσουμε την αλληλεπίδραση αυτών των εγκεφαλικών περιοχών μεταξύ τους. Στα πλαίσια της εργασίας αυτής, θα επικεντρώσουμε τη προσοχή μας συγκεκριμένα στα δεδομένα Λειτουργικής Απεικόνισης Μαγνητικού Συντονισμού (Functional Magnetic Resonance Imaging). Η fMRI είναι μία ιατρική μέθοδος απεικόνισης που παράγει δεδομένα τεσσάρων διαστάσεων και είναι ικανή να αναδειξει διαφορές σε ένα σήμα που είναι εξαρτώμενο από το επίπεδο οξυγόνωσης του αίματος (Blood Oxygen Level Dependent). Το σήμα αυτό συνδέεται έμμεσα με την νευρωνική ενεργοποίηση. Η νευρωνική δραστηριότητα συνεπάγεται αύξηση στη ροή αίματος (λειτουργία γνωστή ως αιμοδυναμική απόκριση) (Kim et al., 1999) που συνοδεύεται με μία φυσική αύξηση της συγκέντρωσης οξυαιμοσφαιρίνης (η οποία τρέφει με οξυγόνο τους νευρώνες) γεγονός που αντιλαμβάνεται το σήμα BOLD.

Ειδικότερα, στη παρούσα εργασία, μας ενδιαφέρει η ανάλυση εγκεφαλικών σημάτων και η ανίχνευση πνευματικών ασθενειών ή νοητικών καταστάσεων βασιζόμενοι στις διαφοροποιήσεις του σήματος BOLD των fMRI δεδομένων μέσα και ανάμεσα σε εγκεφαλικά δίκτυα. Για την αντιμετώπιση του προβλήματος θα χρησιμοποιήσουμε τη μέθοδο της λειτουργικής διασυνδεσιμότητας δικτύου (Functional Network Connectivity). Η μέθοδος αυτή στην ουσία χρησιμοποιεί κάποια μεγέθη λειτουργικής διασυνδεσιμότητας που δείχνουν αν υπάρχουν διαφορές στη χρονική διασυνδεσιμότητα του εγκεφάλου που μπορεί να οφείλονται σε κάποια ασθένεια ή κάποια νοητική κατάσταση (Biswal et al., 1995; van de Ven et al., 2004). Η μέθοδος αυτή δηλαδή ερευνά αν υπάρχουν διαφορές στη χρονική διασυνδεσιμότητα ανατομικών περιοχών του εγκεφάλου ή λειτουργικά οριζόμενων δικτύων.

Ωστόσο για την εφαρμογή της μεθόδου FNC καθώς και οποιασδήποτε άλλης μεθόδου, θα πρέπει πρώτα να επεξεργαστούμε τα δεδομένα μας που πολλές φορές είναι μεγάλα σε όγκο και να τα «καθαρίσουμε» από πληροφορίες που μας εμποδίζουν στην ανάλυση καθώς και να εξάγουμε πηγές του σήματος που είναι διαφορετικές μεταξύ τους, για να διακρίνουμε π.χ. διαφορετικούς τομείς του εγκεφάλου.

Αρχικά λοιπόν στο επόμενο υποκεφάλαιο της εισαγωγής κρίθηκε σκόπιμο να αναφέρουμε κάποια στοιχεία ξεχωριστά για την νευροαπεικόνιση η οποία αποτελεί ένα πολύ σημαντικό μέρος της παρούσας εργασίας, το πως δηλαδή τα δεδομένα αυτά φτάνουν στα χέρια μας ποιες μέθοδοι υπάρχουν και πως λειτουργεί τελικά αυτή η διαδικασία. Θα επικεντρωθούμε κυρίως στη μέθοδο και τα δεδομένα fMRI.

Στο 2<sup>ο</sup> κεφάλαιο της εργασίας θα παρουσιαστούν σαν θεωρία μέθοδοι για την ανάλυση δεδομένων fMRI, τόσο γραμμικές, όπως PCA/SVD (Jolliffe, 2002) και ICA

(Hyvarinen et al., 2000) όσο και μη γραμμικές όπως το ISOMAP (Tenenbaum et al., 2000).

Στη συνέχεια στο 3<sup>ο</sup> κεφάλαιο της εργασίας θα περιγραφεί εκτενώς το κύριο πρόβλημα του πονήματος κατά τα πρότυπα των Anderson και Cohen (2013). Θα παρουσιαστούν οι μέθοδοι και τα βήματα που έγιναν για την ανάλυση πραγματικών FMRI δεδομένων από 72 ασθενείς διαγνωσμένους με σχιζοφρένεια καθώς και 74 υγιείς ανθρώπους του δείγματος.

Επειτα, στο κεφάλαιο 4 θα συζητηθούν τα αποτελέσματα και θα επιχειρηθεί η ερμηνεία τους. Θα παρουσιαστούν γραφήματα και σχηματικές αναπαραστάσεις των αποτελεσμάτων.

Στο κεφάλαιο 5 θα στηριχτούμε στα αποτελέσματα και θα εξάγουμε συμπεράσματα. Θα προτείνουμε νέες βελτιώσεις και ερευνητική δουλειά στο αντικείμενο καθώς και νέες προκλήσεις που προκύπτουν και ανάγκες του αντικειμένου που πρέπει να καλυφθούν σε μελλοντικές επιστημονικές εργασίες και αναλύσεις.

Στο 6<sup>ο</sup> και τελευταίο μέρος της εργασίας θα παραθέσουμε αναλυτικά την βιβλιογραφία που χρησιμοποιήσαμε σε όλα τα κεφάλαια ταξινομημένη σε αλφαβητική σειρά για την διευκόλυνση του αναγνώστη.

Τέλος, στο Παράρτημα μπορεί ο αναγνώστης να βρει τον κώδικα που χρησιμοποιήσαμε για την ανάλυση μας όσον αφορά το υπολογιστικό κομμάτι της εργασίας.

## 1.1 Νευροαπεικόνιση

Με τον όρο νευροαπεικόνιση εννοούμε της χρήση διαφόρων τεχνικών για την έμμεση ή άμεση απεικόνιση της δομής ή λειτουργίας του νευρικού συστήματος του εγκεφάλου. Είναι μία σχετικά νέα επιστημονική γνώση που εμπίπτει σε επιστήμες όπως η Φαρμακολογία, η Νευροεπιστήμη και η Ψυχολογία. Γενικά, η Νευροαπεικόνιση χωρίζεται σε δύο κατηγορίες :

- ✚ Την δομική απεικόνιση, που ασχολείται με την απεικόνιση της δομής του νευρικού συστήματος και τη διάγνωση των -μεγάλης κλίμακας- ενδοκρανιακών νόσων (όπως όγκων) και τραυματισμών.
- ✚ Την λειτουργική απεικόνιση, που ασχολείται με την διάγνωση διαταραχών μεταβολισμού και ομοιόστασης (όπως η νόσος Alzheimer, απεικονίζοντας αλλοιώσεις σε μία πιο λεπτή κλίμακα) αλλά και την παροχή δεδομένων στην έρευνα σε πολλά πεδία της επιστήμης όπως η Νευροεπιστήμη και η γνωστική ψυχολογία.

Στην εργασία μας θα μας απασχολήσει η δεύτερη κατηγορία Νευροαπεικόνισης, η λειτουργική απεικόνιση.



Με τον όρο λειτουργική νευροαπεικόνιση εννοούμε της χρήση της τεχνολογίας της νευροαπεικόνισης για να μετρήσουμε/εντοπίσουμε κάποιον παράγοντα της εγκεφαλικής λειτουργίας ή και για να καταλάβουμε τη σχέση ανάμεσα σε συγκεκριμένες περιοχές του εγκεφάλου και συγκεκριμένες πνευματικές ή σωματικές λειτουργίες.

Κάποιες από τις πιο συνηθισμένες τωρινές μεθόδους νευροαπεικόνισης είναι :

- ✚ Τομογραφία εκπομπής ποζιτρονίων (PET)
- ✚ Λειτουργική απεικόνιση μαγνητικού συντονισμού (fMRI)
- ✚ Ηλεκτροεγκεφαλογράφημα (EEG)
- ✚ μαγνητικό εγκεφαλογράφημα (MEG)

Από τις παραπάνω μεθόδους οι PET και fMRI μπορούν να μετρήσουν τοπικές αλλαγές της εγκεφαλικής αιματικής ροής που σχετίζονται με τη νευρωνική δραστηριότητα . Οι αλλαγές αυτές αναφέρονται ως ενεργοποιήσεις . Οι περιοχές του εγκεφάλου που ενεργοποιούνται όταν ένα άτομο εκτελεί μια συγκεκριμένη εργασία μπορεί να παίζουν ένα σημαντικό ρόλο για τον εντοπισμό των νευρώνων που συμβάλλουν στη συμπεριφορά . Για παράδειγμα, η ευρεία ενεργοποίηση του ινιακού λοβού τυπικά εμφανίζεται σε διεργασίες που αφορούν την οπτική διέγερση. Αυτό είναι το τμήμα του εγκεφάλου που λαμβάνει σήματα από τον αμφιβληστροειδή και γενικά πιστεύεται ότι παίζει ένα σημαντικό ρόλο στην οπτική αντίληψη (**Warrington, 1986**).

Άλλες μέθοδοι νευροαπεικόνισης περιλαμβάνουν την καταγραφή των ηλεκτρικών ρευμάτων ή μαγνητικών πεδίων, για παράδειγμα οι EEG και MEG . Διαφορετικές μέθοδοι έχουν διαφορετικά πλεονεκτήματα για την έρευνα. Για παράδειγμα, η MEG μετρά τη δραστηριότητα του εγκεφάλου με υψηλή χρονική ανάλυση ( κάτω από το επίπεδο του χιλιοστού δευτερολέπτου ), αλλά περιορίζεται στην ικανότητά της να εντοπίσει την εν λόγω δραστηριότητα (την ακριβή τοποθεσία). Η fMRI αντίθετα κάνει μια πολύ καλύτερη δουλειά για την εντόπιση της εγκεφαλικής δραστηριότητας στη χωρική ανάλυση (~1-2 χιλιοστά μέγεθος pixel), αλλά με πολύ χαμηλότερη ανάλυση χρόνου (~ κάποια δευτερόλεπτα).

Η φύση του αντικειμένου μέτρησης σε μια συγκεκριμένη μελέτη γενικά σχετίζεται με το συγκεκριμένο πρόβλημα που αντιμετωπίζεται. Περιορισμοί στη μέτρηση ποικίλλουν μεταξύ των τεχνικών. Για παράδειγμα οι MEG και EEG καταγράφουν τις μαγνητικές ή ηλεκτρικές διακυμάνσεις που συμβαίνουν όταν ένας πληθυσμός νευρώνων είναι ενεργός. Αυτές οι μέθοδοι είναι άριστες στη μέτρηση του χρόνου στη πορεία των νευρωνικών εκδηλώσεων ( της τάξεως των χιλιοστών του δευτερολέπτου) αλλά όπως αναφέραμε και παραπάνω είναι γενικά κακές στην μέτρηση που αφορά την ακριβή τοποθεσία των ενεργοποιήσεων αυτών. Οι PET και fMRI είναι ικανές να μετρήσουν τις αλλαγές στη σύνθεση του αίματος κοντά σε μία ενεργοποίηση ενός πληθυσμού νευρώνων. Επειδή οι μετρήσιμες μεταβολές στο αίμα είναι αργές (της τάξης των δευτερολέπτων) (**Cohen and Bookheimer, 1994**), αυτές οι μέθοδοι είναι πολύ χειρότερες σε μέτρηση της χρονικής εξέλιξης των νευρωνικών γεγονότων, αλλά είναι γενικά καλύτερες σε μία μέτρηση σχετικά με τη θέση των γεγονότων αυτών. Επομένως, επιλέγεται η εκάστοτε μέθοδος σύμφωνα με τις ανάγκες της επικείμενης ερευνητικής εργασίας καθώς και το σκοπό της τελευταίας.

Στην παρούσα εργασία η μελέτη γίνεται πάνω σε δεδομένα λειτουργικής απεικόνισης μαγνητικού συντονισμού (fMRI) και όπως θα αναλυθεί εκτενέστερα στο κεφάλαιο 3 αφορά 74 υγιείς ανθρώπους καθώς και 72 ασθενείς που έχουν διαγνωσθεί με σχιζοφρένεια (άνδρες και γυναίκες σε ηλικίες από 18-65) (Anderson and Cohen, 2013). Ο όγκος των δεδομένων αυτών είναι πολύ μεγάλος αφού τα δεδομένα αποτελούνται από μία χρονοσειρά καθώς και 3D εικόνα με μέγεθος εικονοστοιχείου (voxel)  $3 \times 3 \times 4 \text{ mm}^3$ .

Τα δεδομένα αυτά είναι resting state δηλαδή δεδομένα που συγκεντρώθηκαν σε κατάσταση ηρεμίας του υποκειμένου, χωρίς αυτό δηλαδή να εκτελεί κάποιο ρητό καθήκον για την ανάγκη της έρευνας. Εν γένει, αυτού του τύπου τα δεδομένα fMRI μας βοηθούν να κατανοήσουμε τις περιφερειακές αλληλεπιδράσεις που συμβαίνουν στο υποκείμενο όταν αυτό δεν κάνει μία συγκεκριμένη δραστηριότητα. Το υποκείμενο λοιπόν αναπαύεται και παρατηρούνται οι διαφοροποιήσεις στο σήμα BOLD οι οποίες ανιχνεύονται μέσω της μεθόδου fMRI. Επειδή η δραστηριότητα του εγκεφάλου είναι παρούσα ακόμη και εν απουσία μιας συγκεκριμένης δραστηριότητας, κάθε περιοχή του εγκεφάλου έχει αυθόρμητη διακύμανση στο σήμα BOLD. Η προσέγγιση της μεθόδου αυτής (resting state) είναι χρήσιμη για να διερευνηθεί η λειτουργική οργάνωση του εγκεφάλου ή και να εξεταστεί αν υπάρχουν νευρολογικής ή ψυχιατρικής φύσεως ασθένειες σε ασθενείς (Jafri et al., 2008; Yu et al., 2011). Ήδη, η μέθοδος αυτή έχει αποκαλύψει μια σειρά από δίκτυα του εγκεφάλου τα οποία βρίσκονται συνεχώς σε υγιή άτομα και αφορούν π.χ. τα διάφορα στάδια της συνείδησης (Heine et al., 2012) (ακόμα και μεταξύ άλλων ειδών) και αντιπροσωπεύει συγκεκριμένο μοτίβο της σύγχρονης ερευνητικής δραστηριότητας.

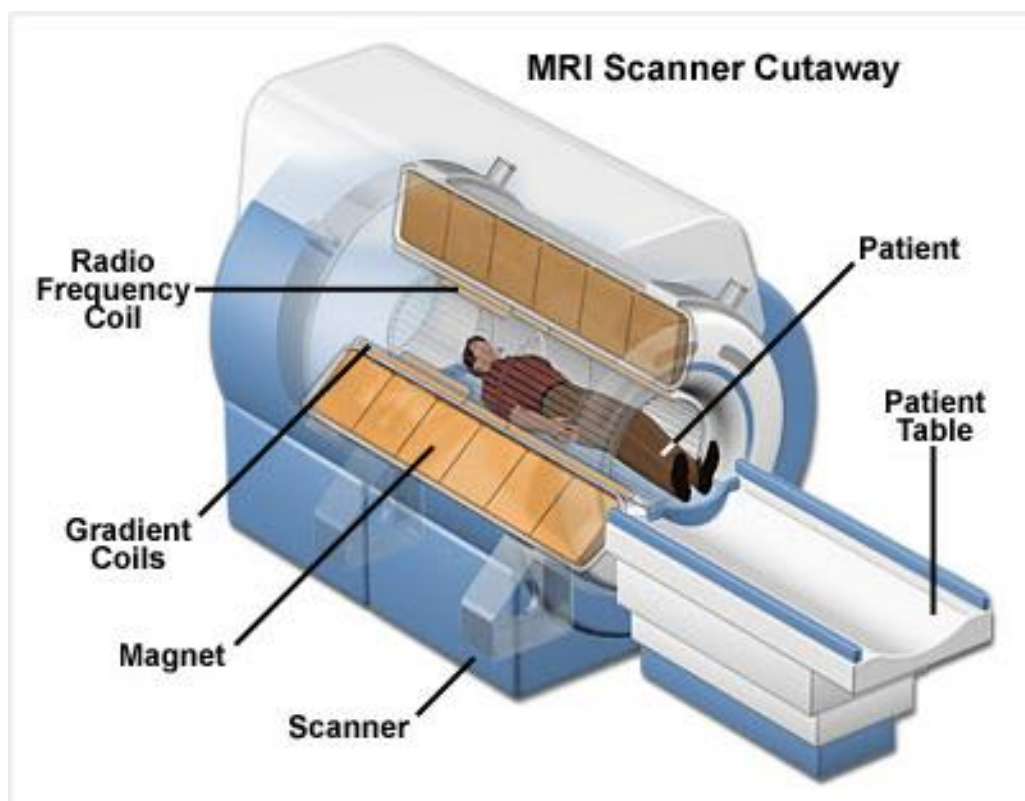
Στο επόμενο υποκεφάλαιο λοιπόν θα αναλύσουμε συγκεκριμένα τη μέθοδο fMRI και τα δεδομένα που συγκεντρώνει.

## 1.2 Μέθοδος και δεδομένα fMRI

Η Λειτουργική Απεικόνιση Μαγνητικού Συντονισμού (Functional Magnetic Resonance Imaging) απεικονίζει την αιμοδυναμική αντίδραση που σχετίζεται με τη νευρωνική δραστηριότητα στον εγκέφαλο και το Νωτιαίο Μυελό. Είναι σχετικά πρόσφατη μέθοδος Νευροαπεικόνισης και βασίζεται αφενός στο φαινόμενο πυρηνικού μαγνητικού συντονισμού (Nuclear Magnetic Resonance) όπου ένας πυρήνας μπορεί να απορροφήσει και να εκπέμψει ηλεκτρομαγνητική ακτινοβολία σε μία συγκεκριμένη συχνότητα συντονισμού (resonant frequency) και αφετέρου στην θεμελιώδη υπόθεση ότι αύξηση στην νευρωνική δραστηριότητα συμπίπτει με αύξηση στην εγκεφαλική αιματική ροή (Kim et al., 1999).

Από τη δεκαετία του 1890, χάρη στους Roy et al. (1890) ήταν γνωστό ότι αλλαγές στην αιμοδυναμική του εγκεφάλου συνδέονται στενά με τη νευρωνική δραστηριότητα. Οι νευρώνες καταναλώνουν το οξυγόνο που μεταφέρουν οι αιμοσφαιρίνες (πρωτεΐνες του αίματος) των ερυθροκυττάρων (ερυθρά αιμοσφαίρια) από γειτονικά τριχοειδή αγγεία. Η τοπική αντίδραση σε αυτή την κατανάλωση οξυγόνου είναι μια αύξηση της ροής αίματος στις περιοχές αυξημένης νευρωνικής δραστηριότητας (Kim et al., 1999), που συμβαίνει με καθυστέρηση περίπου 1-5

δευτερολέπτων. Αυτή η αιμοδυναμική απόκριση γίνεται μέγιστη μετά τα 4-5 δευτερόλεπτα και στη συνέχεια επιστρέφει στην φυσιολογική τιμή (συνήα ξεπερνώντας την ελαφρώς) κάτι που θα αναλύσουμε εκτενέστερα σε επόμενο υποκεφάλαιο. Αυτή η απόκριση έχει ως συνέπεια τοπικές αλλαγές στις σχετικές συγκεντρώσεις οξυαιμοσφαιρίνης (αιμοσφαιρίνη ενωμένη με οξυγόνο) και δεοξυαιμοσφαιρίνης (αιμοσφαιρίνη που δεν έχει δεσμευμένο οξυγόνο). Οι αλλαγές αυτές επιφέρουν εκ νέου μεταβολή στον όγκο του αίματος στην περιοχή, γεγονός που επιφέρει και μεταβολή στη ροή του αίματος (Cohen and Bookheimer, 1994).



**Εικόνα 1.** Αναπαράσταση ενός μαγνητικού τομογράφου και των δομικών μερών του για την παραγωγή δεδομένων FMRI [[www.cs.washington.edu](http://www.cs.washington.edu)].

Ο κυλινδρικός σωλήνας του MRI σκάνερ (Εικόνα 1) έχει στο εσωτερικό του έναν πανίσχυρο ήλεκτρο-μαγνήτη (Magnet). Ένα τυπικό ερευνητικό σκάνερ τέτοιου τύπου έχει ένα μαγνητικό πεδίο στα 3 Tesla (T), περίπου 50,000 φορές μεγαλύτερο από το βαρυτικό πεδίο της γης. Το μαγνητικό πεδίο μέσα στο σκάνερ επηρεάζει το μαγνητικό πυρήνα των ατόμων. Σε τυπικές συνθήκες ο πυρήνας του ατόμου είναι τυχαία οριοθετημένος αλλά υπό την επιρροή του μαγνητικού πεδίου ο πυρήνας ευθυγραμμίζεται με την κατεύθυνση του πεδίου. Όσο πιο δυνατό είναι το μέτρο του πεδίου τόσο μεγαλύτερος ο βαθμός ευθυγράμμισης. Αφού επιτευχθεί η ευθυγράμμιση των πυρήνων, εφαρμόζεται και άλλο ένα μαγνητικό πεδίο από πηνία κλίσης (gradient coils), το μαγνητικό πεδίο κλίσης (gradient field), που εφαρμόζεται για να εντοπίσει χωρικά διαφορετικούς πυρήνες. Έτσι υπάρχει και ένα δεύτερο μαγνητικό πεδίο που δεν είναι ίσου μέτρου κατά μήκος του εγκεφάλου. Ο λόγος είναι ότι όπως γνωρίζουμε από

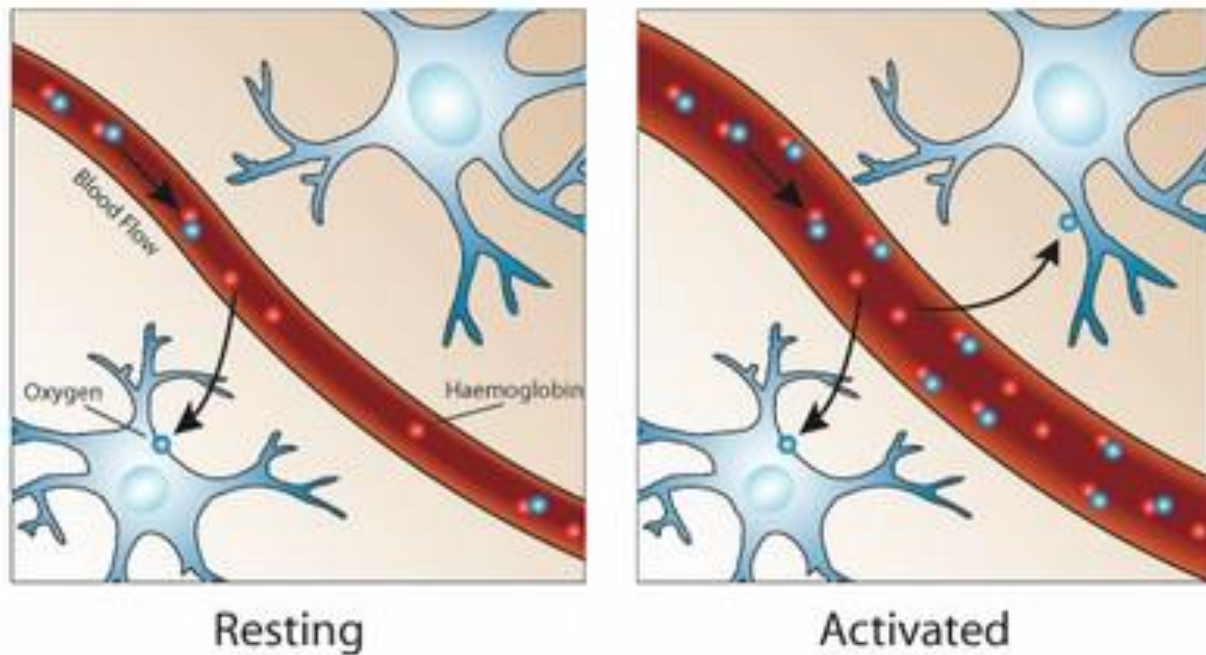
το φαινόμενο NMR, αν προκαλέσουμε κάποια ραδιοκύματα με συγκεκριμένη συχνότητα ίση με τη συχνότητα συντονισμού (resonant frequency, που είναι ανάλογη με τη δύναμη του μαγνητικού πεδίου) για κάθε πυρήνα, τότε αυτός μπορεί να τα απορροφήσει. Όταν ωστόσο σταματήσουμε την εκπομπή εκείνος μπορεί να εκπέμψει πίσω σαν μία ηχώ (αυτός είναι ο λόγος που συνήθως ένας fMRI σαρωτής λέγεται echo-planar) ραδιοκύματα με συχνότητα που βασίζεται στην δύναμη του μαγνητικού πεδίου. Έτσι αν είχαμε μόνο ένα μαγνητικό πεδίο δεν θα μπορούσαμε να πετύχουμε την εκπομπή όλων των πυρήνων σε μία δεδομένη συχνότητα  $X$ , αφού η αιμοσφαιρίνη είναι διαμαγνητική ενώ η δεοξυαιμοσφαιρίνη παραμαγνητική (Cohen and Bookheimer, 1994). Η παρουσία οποιασδήποτε ουσίας εντός ενός μαγνητικού πεδίου μεταβάλλει το μαγνητικό πεδίο της περιοχής σε ένα συγκεκριμένο βαθμό. Επομένως σε περιοχές όπου θα υπήρχε μεγαλύτερη συγκέντρωση δεοξυαιμοσφαιρίνης οι πυρήνες δεν θα εξέπεμπαν πίσω τα ραδιοκύματα στη συχνότητα  $X$ . Άρα, χρησιμοποιώντας το μαγνητικό πεδίο κλίσης καταφέρνουμε να αλλάξουμε το μέτρο του μαγνητικού πεδίου κατά μήκος του εγκεφάλου. Αν λοιπόν στη συνέχεια εκπέμψουμε ένα λευκό θόρυβο σε όλο το εύρος των συχνοτήτων (RF pulse) με τη βοήθεια του πηνίου ραδιοσυχνότητας (radio frequency coil) αλλάζουμε τις οριοθετήσεις των πυρήνων και από τις ανταποκρίσεις των τελευταίων γνωρίζουμε την ακριβή τοποθεσία τους αφού η συχνότητα που εκπέμπουν δίνεται από τη μαγνητική δύναμη στην οποία υπόκεινται. Επομένως, μετά μπορούμε με ακρίβεια να υπολογίσουμε την τοποθεσία της κάθε ανταπόκρισης που είναι συνάρτηση της δύναμης του μαγνητικού πεδίου.

Στην παραπάνω διαδικασία είναι το μαγνητικό σήμα από πυρήνες υδρογόνου που εντοπίζεται. Το γεγονός αυτό οφείλεται στο ότι το ανθρώπινο σώμα αποτελείται σε μεγάλο βαθμό από νερό. Όπως αναφέραμε και παραπάνω η παρουσία οποιασδήποτε ουσίας εντός ενός μαγνητικού πεδίου μεταβάλλει το μαγνητικό πεδίο σε ένα συγκεκριμένο βαθμό. Η διαφορά στις μαγνητικές ιδιότητες οδηγεί σε μικρές διαφορές στα MR σήματα του αίματος που εξαρτώνται από το βαθμό οξυγόνωσης της περιοχής (Cohen and Bookheimer, 1994). Συγκεκριμένα περιοχές στις οποίες υπάρχει υψηλή συγκέντρωση σε δεοξυαιμοσφαιρίνη έχουν μεγαλύτερο βαθμό εξασθένησης του δείκτη  $T_2^*$  και τα MR σήματα είναι πιο εξασθενημένα.

Γενικά ο μαγνητικός τομογράφος ανάλογα με τις ρυθμίσεις του μπορεί να λάβει διαφορετικά δεδομένα. Η λογική είναι ότι ρυθμίζουμε έτσι τον τομογράφο ώστε να μεγιστοποιήσουμε τις διαφορές σε συγκεκριμένο δείκτη που εξετάζουμε με απώτερο σκοπό να πάρουμε πληροφορίες για διαφορετικούς ιστούς. Σημαντικότεροι δείκτες, είναι ο δείκτης  $T_1$  που αφορά το χρόνο διαμήκους χαλάρωσης των πυρήνων (μετά την απόκλιση από το μαγνητικό πεδίο λόγω των ραδιοκυμάτων), ο δείκτης  $T_2$  που αφορά την εγκάρσια χαλάρωση (ουσιαστικά μετράει το πόσο γρήγορα οι πυρήνες εκπέμπουν ενέργεια μέχρι να επιτευχθεί και πάλι ισορροπία) και ο δείκτης  $T_2^*$  που σχετίζεται στενά με τον  $T_2$  αλλά αφορά την ανίχνευση ανομοιογένειας στο τοπικό μαγνητικό πεδίο των πυρήνων αλλά και τις αλληλεπιδράσεις μεταξύ των μορίων.

Βάση αυτών των δεικτών ο μαγνητικός τομογράφος παράγει εικόνες με διαφορετική στόχευση την κάθε φορά. Για παράδειγμα, αν μας ενδιαφέρει να εντοπίσουμε ανατομικές ανωμαλίες στον εγκέφαλο χρησιμοποιούμε εικόνες (δομικές εικόνες του εγκεφάλου) βασιζόμενοι στον δείκτη  $T_1$ . Ο δείκτης που σχετίζεται με τα fMRI είναι κυρίως ο  $T_2^*$  και δηλώνει το βαθμό εξασθένησης του MR σήματος (κυρίως λόγω ανομοιογένειας στο τοπικό μαγνητικό πεδίο των πυρήνων) ανιχνεύοντας έτσι

έμμεσα την νευρωνική δραστηριότητα μίας εγκεφαλικής περιοχής. Αφού η οξυγόνωση του αίματος (BOLD) διαφέρει ανάλογα με τα επίπεδα της νευρωνικής δραστηριότητας, οι διαφορές στα MR σήματα ανάλογα με τις περιοχές (εξασθενημένο σήμα  $\Rightarrow$  περιοχή με υψηλή συγκέντρωση δεοξυαιμοσφαιρίνης και το αντίθετο) αυτές μπορούν να χρησιμοποιηθούν για να ανιχνεύεται εγκεφαλική δραστηριότητα. Οξυγόνο διανέμεται στους νευρώνες από την αιμοσφαιρίνη σε τριχοειδή ερυθρών αιμοσφαιρίων. Όταν η νευρωνική δραστηριότητα αυξάνεται τότε υπάρχει αύξηση στη ροή του αίματος (Εικόνα 2) στις περιοχές με αυξημένη νευρωνική δραστηριότητα (Kim et al., 1999).

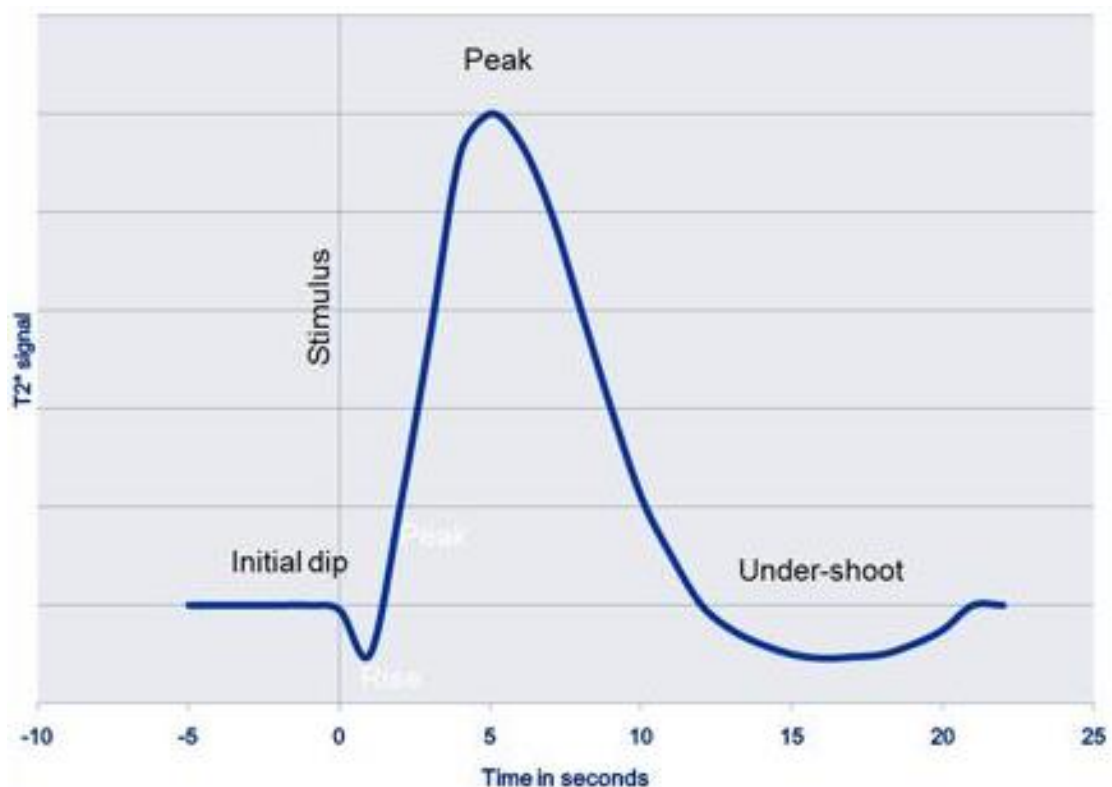


**Εικόνα 2.** Αναπαράσταση μίας εγκεφαλικής περιοχής κατά τη χαλάρωση και ενεργοποίηση των νευρώνων και ο αντίκτυπος στη ροή και την οξυγόνωση του αίματος [ [www.psychcentral.com](http://www.psychcentral.com) ].

Επομένως για μία περιοχή με εξασθενημένο MR σήμα αυτό θα σημαίνει ότι υπάρχει υψηλή συγκέντρωση δεοξυαιμοσφαιρίνης (παραμαγνητικές ιδιότητες μεγαλύτερη εξασθένηση για το  $T_2^*$ ) άρα αυτό θα κάνει την απεικόνιση στο συγκεκριμένο εικονοστοιχείο να είναι πιο σκοτεινή ενώ σε αντίθετη περίπτωση πιο λαμπερή. Με τον τρόπο αυτό διαμορφώνονται πολλές εικόνες ανάλογα με το πείραμα που γίνεται για κάποια συγκεκριμένη ενέργεια του υποκειμένου με αισθητικοκινητικές λειτουργίες ή και σε ήρεμη κατάσταση (resting state).

### 1.2.1 Συμπεριφορά του σήματος BOLD

Αξιοσημείωτη επίσης είναι η κατεύθυνση της μεταβολής οξυγόνωσης του αίματος (Εικόνα 3) για μία έντονη δραστηριότητα. Μπορεί να περιμένει κανείς ενστικτωδώς να μειωθεί η οξυγόνωση με την μείωση της δραστηριότητας (δεδομένου ότι οι νευρώνες «τρέφονται» με οξυγόνο), στην πραγματικότητα όμως τα πράγματα είναι πιο πολύπλοκα. Υπάρχει λοιπόν μία στιγμιαία μείωση της οξυγόνωσης του αίματος ακριβώς μετά από αυξημένη νευρωνική δραστηριότητα, καλούμενη και ως “the initial dip” (αρχική βουτιά) στην αιμοδυναμική απόκριση. Αυτή η μείωση ακολουθείται στη συνέχεια από μία περίοδο 4 περίπου δευτερολέπτων όπου η ροή του αίματος αυξάνεται, όχι μόνο σε επίπεδο όπου ικανοποιείται η ανάγκη σε οξυγόνο, αλλά σε επίπεδο που υπερκαλύπτεται η ζήτηση αυτή. Πρακτικά, αυτό σημαίνει ότι η οξυγόνωση του αίματος αυξάνεται πραγματικά μετά την ενεργοποίηση των νευρώνων. Η ροή του αίματος φτάνει στη μέγιστη τιμή της μετά από περίπου 6 δευτερόλεπτα και στη συνέχεια πέφτει πίσω στην αρχική της κατάσταση (συχνά ξεπερνώντας την ελαφρώς), συχνά συνοδευόμενη από μία μετά-διεγερτική υποτίναξη (post-stimulus undershoot).



**Εικόνα 3.** Αναπαράσταση της συμπεριφοράς του σήματος BOLD στο χρόνο για μία δραστηριότητα του υποκειμένου μετά από αρχικό ερέθισμα [[www.radiopaedia.org](http://www.radiopaedia.org)].

### 1.2.2 Ποιοτικά και ποσοτικά χαρακτηριστικά των FMRI δεδομένων

Τα αποτελέσματα του BOLD μετρώνται με τη χρήση ογκομετρικής λήψης εικόνων με αντίθεση σταθμισμένη (ρυθμίσεις τομογράφου) με βάση τις παραμέτρους χαλάρωσης  $T_2$  και  $T_2^*$ . Οι εικόνες αυτές έχουν συνήθως από μέτρια έως πολύ καλή ανάλυση. Πρόσφατες πρόοδοι στην τεχνική, με χρήση υψηλότερων μαγνητικών πεδίων και πολυκαναλικών ανιχνευτών ραδιοκυμάτων έχουν ανεβάσει την χωρική διακριτικότητα μέχρι και το επίπεδο του χιλιοστού. Γενικά οι εικόνες FMRI λαμβάνονται κάθε 1-4 δευτερόλεπτα και κάθε στοιχείο όγκου της εικόνας έχει μέγεθος 1-6 χιλιοστά σε κάθε ακμή. Παρόλο που μπορούν να διαχωριστούν χρονικά ερεθίσματα μέχρι και 2 δευτερόλεπτα, ο ολικός χρόνος της απόκρισης BOLD σε ένα σύντομο ερέθισμα διαρκεί περίπου 15 δευτερόλεπτα (για να θεωρούνται τυχαίες οι συνθήκες για την επόμενη μέτρηση). Αφού συλλεχθούν όλες οι φωτογραφίες, οι οποίες συλλέγονται σε φέτες του εγκεφάλου (slices) με πολλαπλές σαρώσεις, αποθηκεύονται στον 2D χώρο Fourier (στο χώρο των συχνοτήτων) σαν δεδομένα από ραδιοκύματα επί το χρόνο. Όσο καλύτερη γίνεται η ανάλυση της φωτογραφίας τόσο περισσότερες σαρώσεις πρέπει να γίνουν σε περισσότερες «φέτες» του εγκεφάλου. Στη συνέχεια μετά από επεξεργασία αυτών των ακατέργαστων (raw) δεδομένων γίνεται η μεταφορά των εικόνων από το χώρο των συχνοτήτων στον καρτεσιανό χώρο.

### 1.2.3 Επεξεργασία των δεδομένων FMRI

Σκοπός της ανάλυσης δεδομένων FMRI είναι η ανίχνευση συσχετίσεων (στο σήμα BOLD) ανάμεσα σε εγκεφαλικές περιοχές κατά τη διάρκεια αισθησιοκινητικών λειτουργιών ή η ανάλυση λειτουργική διασυνδεσιμότητας εγκεφαλικών περιοχών σε ήρεμη κατάσταση. Επειδή όμως το σήμα της FMRI είναι ασθενές και οι διαφορές του BOLD που εντοπίζονται είναι επίσης ασθενείς και μικρές, ο θόρυβος είναι ένα μεγάλο πρόβλημα στα FMRI δεδομένα που μπορεί να αποπροσανατολίσει εύκολα την μετέπειτα ανάλυση.

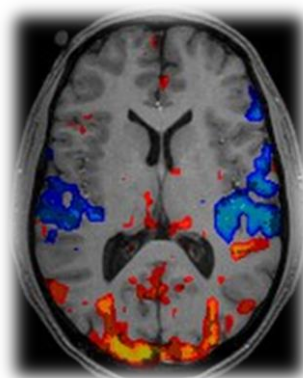
Σε μια τυπική σάρωση FMRI με χρήση απεικόνισης echo-planar (EPI) που είναι και μία από τις τεχνικές για γρήγορη συλλογή των εικόνων, ο όγκος της κεφαλής απεικονίζεται κάθε 1 ή 2 δευτερόλεπτα, με μέχρι και μερικές χιλιάδες εικόνες. Οι εικόνες λαμβάνονται στο χώρο των συχνοτήτων, και πρέπει να μετασχηματιστούν αντίστροφα στον 3D χώρο για ανάλυση όπως προαναφέρθηκε και παραπάνω. Η φύση της ανίχνευσης και ατέλειες στον ανιχνευτή όπως θερμική κίνηση, αλλά και η κίνηση και η αναπνοή του υποκειμένου παράγουν παραμορφώσεις.

Μετά την ανακατασκευή των εικόνων, η έξοδος του ανιχνευτή είναι μια σειρά 3D εικόνων του εγκεφάλου. Η πιο συχνές διορθώσεις που γίνονται σ' αυτές τις εικόνες είναι η διόρθωση κίνησης, θερμικού θορύβου, εξαίρεση τυχαίων ενεργοποιήσεων νευρώνων, η διόρθωση φυσιολογικών συνεπειών (αναπνοή του υποκειμένου κτλ.) καθώς και αφαίρεση σημάτων που δεν προέκυψαν από εγκεφαλικούς ιστούς κτλ. Με χρήση χρονικών φίλτρων επίσης μπορεί να επιχειρηθεί η από-συνέλιξη της απόκρισης

BOLD για την ανάκτηση του σήματος στην περίπτωση που αναμένεται νευρωνική δραστηριότητα ενισχύοντας το επιθυμητό σήμα. Το τελευταίο μέρος της επεξεργασίας συχνά περιλαμβάνει μείωση διάστασης των δεδομένων με κλασσικές τεχνικές όπως η ανάλυση σε κύριες συνιστώσες (PCA) (Jolliffe, 2002) λόγω του τεράστιου και θορυβώδους όγκου των δεδομένων. Τέλος, μπορεί κανείς να ακολουθήσει ανάλυση σε ανεξάρτητες συνιστώσες (ICA) (Hyvarinen et al., 2000) ή ανάλυση σε αραιές συνιστώσες (SPA) (Georgiev et al., 2007) ή και πολλές άλλες μεθόδους για να ανακαλύψει και να απομονώσει περιοχές των δεδομένων που θέλει να μελετήσει (Regions - Of - Interest).

#### 1.2.4 FMRI scan (Χάρτες ενεργοποίησης)

Όπως αναφέραμε και παραπάνω ένας από τους κυριότερους σκοπούς της ανάλυσης FMRI είναι η ανίχνευση λειτουργικής διασυνδεσιμότητας του εγκεφάλου. Η εικόνα 4 που φαίνεται στα δεξιά είναι το αποτέλεσμα ενός απλού πειράματος FMRI. Ενώ το υποκείμενο βρίσκεται σε μαγνητικό τομογράφο παρακολουθεί μία οθόνη η οποία δείχνει κάθε 30 δευτερόλεπτα (στο ενδιάμεσο είναι σκοτεινή) ένα οπτικό ερέθισμα. Εν τω μεταξύ, ο μαγνητικός τομογράφος παρακολουθεί το σήμα σε όλο τον εγκέφαλο. Σε περιοχές του εγκεφάλου που ανταποκρίνονται στο οπτικό ερέθισμα περιμένουμε το σήμα να έχει αυξομειώσεις αφού το ερέθισμα ενεργοποιείται και απενεργοποιείται, αν και ελαφρώς θολό από την καθυστέρηση στην απόκριση της ροής του αίματος που είδαμε παραπάνω. Οι ερευνητές λοιπόν, βλέπουν τη δραστηριότητα σε σάρωση, σε εικονοστοιχεία - ή στοιχειώδη εικονοστοιχεία όγκου (voxel) . Το voxel είναι το μικρότερο διακριτό σχήμα κουτιού που είναι μέρος μιας τρισδιάστατης εικόνας. Η δραστηριότητα σε ένα voxel συνήθως ορίζεται ως το πόσο κοντά είναι η χρονική σειρά του σήματος λόγω του οπτικού ερεθίσματος με εκείνη που αντιστοιχεί στην τοπική εγκεφαλική δραστηριότητα πριν το ερέθισμα. Voxel του οποίου το σήμα αντιστοιχεί σε μεγάλο βαθμό του δίνεται μια υψηλή βαθμολογία για ενεργοποίηση, ενώ voxel του οποίου το σήμα δεν δείχνει κάποια σημαντική συσχέτιση του δίνεται χαμηλή βαθμολογία και voxel που δείχνει το αντίθετο (απενεργοποίηση) του δίνεται ένα αρνητικό αποτέλεσμα. Βέβαια, για την αποφυγή λάθος εκτιμήσεων για την ενεργοποίηση των voxel, έχουν κατασκευαστεί πολλά μοντέλα με πολλαπλές συγκρίσεις για την μείωση του στατιστικού λάθους και την εξασφάλιση μίας σίγουρης ενεργοποίησης. Τελικά, αφού έχουμε για κάθε voxel κάποιο βαθμό ενεργοποίησης /απενεργοποίησης, αυτά μπορούν να μετατραπούν σε χάρτες ενεργοποίησης. Οι χάρτες αυτοί, αναπαριστούν τη διέγερση των εγκεφαλικών περιοχών κατά τη διάρκεια μίας συγκεκριμένης λειτουργίας ή και την αλληλεπίδραση μεταξύ των περιοχών αυτών σε περίπτωση ήρεμης κατάστασης (resting state).



**Εικόνα 4.** FMRI scan  
(Χάρτης ενεργοποίησης).  
[[www.psychcentral.com](http://www.psychcentral.com)]



### 1.2.5 Πλεονεκτήματα και μειονεκτήματα της μεθόδου FMRI

Όπως αναφέραμε και πιο πάνω καμία μέθοδος νευροαπεικόνισης δεν είναι ιδανική για όλες τις περιπτώσεις. Η επιλογή της μεθόδου νευροαπεικόνισης που πρόκειται να χρησιμοποιηθεί είναι σε άμεση συνάρτηση με το αντικείμενο και τη φύση της μελέτης που πρόκειται να πραγματοποιηθεί. Ας δούμε λοιπόν παρακάτω συγκεντρωτικά τα χαρακτηριστικά της μεθόδου FMRI :

- ✚ Δεν χρειάζονται ραδιενεργοί ανιχνευτές, γεγονός που κάνει τη μέθοδο αυτή μη επεμβατική (non-invasive) για το υποκείμενο και ασφαλής (δεν υπάρχουν γνωστές παρενέργειες μέχρι σήμερα), αντίθετα με άλλες μεθόδους όπως η PET.
- ✚ Υψηλή χωρική ανάλυση στα 1-6 χιλιοστά στις περισσότερες εφαρμογές, που θεωρείται πραγματικά πολύ ικανοποιητική για τον εντοπισμό της ακριβούς τοποθεσίας της νευρωνικής δραστηριότητας.
- ✚ Αρκετά γρήγορη χρονική ανάλυση της τάξης των δευτερολέπτων, ώστε να γίνει διάκριση μεταξύ των δοκιμών (να θεωρηθούν δηλαδή οι συνθήκες τυχαίες για την επόμενη μέτρηση, event related design randomization).
- ✚ Χρονική ανάλυση που δεν είναι αρκετά γρήγορη για να γίνει διάκριση μεταξύ προτύπων ενεργοποίησης που συνδέονται με τα διάφορα στάδια της επεξεργασίας ερεθίσματος, αντίθετα με άλλες μεθόδους όπως η EEG.
- ✚ Η μέθοδος FMRI μετράει έμμεσα την νευρωνική δραστηριότητα βασισμένη στην υπόθεση ότι αύξηση στη νευρωνική δραστηριότητα συνεπάγεται μία αύξηση της αιματικής ροής. Πρακτικά λοιπόν μπορεί να μελετήσει μόνο την αιματική ροή και όχι ατομικά, τα εγκεφαλικά κύτταρα.
- ✚ Η μέθοδος FMRI είναι σχετικά ακριβή γεγονός που βάζει περιορισμούς στην ευρεία χρήση τις μεθόδου.
- ✚ Κατά τη διάρκεια του πειράματος το υποκείμενο πρέπει να μένει κυριολεκτικά ακίνητο (πολλές φορές μία αλλά και δύο ώρες) κάτι που στην πραγματικότητα είναι αδύνατο, δημιουργώντας τον κίνδυνο θορύβου που είναι ικανός να παραμορφώσει τα δεδομένα σε μεγάλο βαθμό και να αποπροσανατολίσει τη μετέπειτα ανάλυση.
- ✚ Δεν μπορεί ένα υποκείμενο να υποβληθεί στο πείραμα αν φέρει εσωτερικά μεταλλικά στοιχεία (όπως ο βηματοδότης και άλλα εμφυτεύματα). Γεγονός που ισχύει και για την απλή απεικόνιση μαγνητικού συντονισμού (μαγνητικός τομογράφος, λόγω του πανίσχυρου μαγνητικού πεδίου).

## 2. Γραμμικές και μη γραμμικές μέθοδοι ανάλυσης FMRI δεδομένων

---

Μετά από μία εκτενής προσπάθεια για πάνω από μία δεκαετία, οι επιστήμονες – ερευνητές εφηύραν αρκετές μεθόδους για την ανάλυση δεδομένων FMRI. Παρακάτω, θα γίνει μία παρουσίαση κάποιων από αυτές τις μεθόδους κατά τα πρότυπα των **Li et al. (2009)**. Γενικά οι μέθοδοι αυτοί μπορούν να διαχωριστούν σε δύο βασικές κατηγορίες. Τις μεθόδους που βασίζονται στο μοντέλο της ανάλυσης και τις μεθόδους που βασίζονται απευθείας στα δεδομένα αυτά καθ' αυτά. Κάθε κατηγορία έχει τα προτερήματα και τους περιορισμούς της.

### 2.1 Γραμμικές μέθοδοι που βασίζονται στο μοντέλο (Model-based)

Μέχρι σήμερα, πολλές ανακαλύψεις μεθόδων ανάλυσης λειτουργικής διασυνδεσιμότητας είναι βασιζόμενες στο εκάστοτε μοντέλο. Αυτές είναι μέθοδοι που επιλέγουν τις λεγόμενες περιοχές ενδιαφέροντος (Regions Of Interest) και στη συνέχεια αποφασίζουν αν και πότε άλλες περιοχές συνδέονται μαζί τους. Για τη διαδικασία αυτή ορίζονται συγκεκριμένες μετρικές που στην ουσία παράγουν έναν πίνακα διασυνδεσιμότητας του ανθρώπινου εγκεφάλου. Συνήθως, αυτές οι μέθοδοι βασίζονται σε μία προχωρημένη γνώση πάνω στη Νευροεπιστήμη και μεγάλη εμπειρία στο αντικείμενο.

#### 2.1.1 Ανάλυση πολλαπλής συσχέτισης (Cross-Correlation Analysis)

Η ανάλυση πολλαπλής συσχέτισης (CCA) είναι μία ώριμη τεχνική και έχει δοκιμαστεί και εφαρμοστεί σε πολλά πεδία της επιστήμης. Το πεδίο αυτό με το οποίο θα ασχοληθούμε εμείς είναι η πολλαπλή συσχέτιση σε ανάλυση χρονοσειρών που προκύπτουν από δεδομένα FMRI. Συγκεκριμένα, η CCR για την ανάλυση FMRI προτάθηκε από τους **Cao και Worsley (1999)**. Εγγενώς αν ένα μέρος του εγκεφάλου είναι λειτουργικά συνδεδεμένο με μία συγκεκριμένη περιοχή ενδιαφέροντος (ROI), τότε θα έπρεπε να υπάρχει συσχέτιση όσον αφορά τη χρονοσειρά του δείκτη BOLD (Blood Oxygen Level Dependent) ανάμεσα στις δύο χρονοσειρές. Για μία χρονοσειρά BOLD μίας περιοχής του εγκεφάλου έστω  $F_x(k)$  και τη χρονοσειρά μίας ROI έστω  $F_y(k)$ , η συνάρτηση πολλαπλής συσχέτισης μπορεί να εκτιμηθεί για καθυστέρηση  $\mu$  (lags) από τη σχέση :

$$Cor_{x,y}(\mu) = \frac{Cov_{x,y}(\mu)}{\sqrt{Var(x) \times Var(y)}} \quad (1)$$

Όπου,  $Var(x)$  και  $Var(y)$  είναι η διασπορά των  $F_x(k)$  και  $F_y(k)$  αντίστοιχα και  $Cov_{x,y}(\mu)$  είναι η συνδιακύμανση των  $F_x(k)$  και  $F_y(k)$  αντίστοιχα με καθυστέρηση  $\mu$  και δίνεται από τη σχέση :

$$Cov_{x,y}(\mu) = E\{(F_x(k) - E(F_x)) \times (F_y(k) - E(F_y))\} \quad (2)$$

Εδώ, το  $E$  είναι η μέση τιμή και τα  $E(F_x)$  και  $E(F_y)$  είναι οι αναμενόμενες τιμές ή μέσες τιμές των  $F_x(k)$  και  $F_y(k)$ . Στην ανάλυση πολλαπλής συσχέτισης λοιπόν αν η συνδιακύμανση  $Cov_{x,y}(\mu)$  είναι πάνω από κάποια οριακή τιμή (threshold), τότε θεωρούμε ότι οι δύο BOLD χρονοσειρές  $F_x(k)$  και  $F_y(k)$  είναι λειτουργικά συνδεδεμένες. Με τον όρο lags εννοούμε την μεταφορά στο χρόνο (timeshifts) μίας εκ των δύο χρονοσειρών κρατώντας την άλλη χρονοσειρά σταθερή.

Βέβαια, ολόκληρος ο υπολογισμός της πολλαπλής συσχέτισης (cross-correlation) για όλες τις πιθανές καθυστερήσεις (lags) θα ήταν υπολογιστικά εξαντλητικός (Cecchi and Guillermo et al., 2007; Almasi et al., 2005). Ευτυχώς, η αιμοδυναμική απόκριση του αίματος κάνει τον υπολογισμό αυτόν μη αναγκαίο. Παρόλο που η συνάρτηση αιμοδυναμικής απόκρισης (HRF) διαφέρει για τα διάφορα υποκείμενα ή ακόμα και για διαφορετικές περιοχές του εγκεφάλου στο ίδιο υποκείμενο, η διάρκεια της συνάρτησης αυτής είναι περιορισμένη. Αυτό σημαίνει ότι θα επιστρέψει στο αρχικό σημείο αναφοράς μετά από ένα περιορισμένο χρόνο που συνήθως είναι περίπου 12 δευτερόλεπτα (Ogawa et al., 1990; Blamire et al., 1992; Friston et al., 1994). Έτσι γενικά, χρειάζεται να υπολογίζουμε την πολλαπλή συσχέτιση με μόνο μερικές χρονικές μονάδες καθυστέρησης (ο ακριβής αριθμός υπολογίζεται σε συνάρτηση με τις ρυθμίσεις και τα χαρακτηριστικά της λειτουργίας του μαγνητικού τομογράφου και συγκεκριμένα τον χρόνο επανάληψης των συνεχόμενων παλμών TR). Στην πραγματικότητα σε πολλές μελέτες ανάλυσης πολλαπλής συσχέτισης υπολογίζουμε μόνο τη συσχέτιση με μηδενική καθυστέρηση.

### 2.1.2 Ανάλυση συνοχής (Coherence Analysis)

Παρόλο που η ανάλυση πολλαπλής συσχέτισης έχει χρησιμοποιηθεί ευρέως για την ανάλυση fMRI δεδομένων (τόσο σε πρότυπη μορφή πειράματος με ρητό καθήκον του υποκειμένου όσο και σε κατάσταση ηρεμίας του υποκειμένου (resting state) ), η χρήση της πολλαπλής συσχέτισης για μηδενική καθυστέρηση σαν μέτρο διασυνδεσιμότητας (connectivity measure) είναι αμφιλεγόμενη (Cecchi and Guillermo et al., 2007). Από τη μία, η πολλαπλή συσχέτιση είναι ευαίσθητη στη μορφή της συνάρτησης αιμοδυναμικής απόκρισης, η οποία έχει αναφερθεί ότι διαφέρει όχι μόνο για διαφορετικά υποκείμενα αλλά και για διαφορετικές περιοχές του εγκεφάλου (Bandettini et al., 1999; Buckner et al., 1998; Lee et al., 2001; Miezin et al., 2000) . Από την άλλη, μεγάλη συσχέτιση μπορεί να εντοπιστεί μεταξύ περιοχών οι οποίες δεν έχουν στην πραγματικότητα διακυμάνσεις στη ροή του αίματος. Επιπλοκές από θόρυβο λόγω της καρδιακής δραστηριότητας αλλά και της δραστηριότητας των αιμοφόρων αγγείων μπορεί να οδηγήσουν στην ψευδαίσθηση μίας μεγάλης συσχέτισης (Friston et al., 1995).

Για να ξεπεράσουμε αυτά τα προβλήματα λοιπόν, ένα νέο είδος μετρικής που λέγεται συνοχή (Coherence) προτάθηκε από τους **Sun et al. (2004)**.

Η συνοχή είναι μία φασματική αναπαράσταση της συσχέτισης στον τομέα των συχνοτήτων. Όπως γνωρίζουμε από την ανάλυση Fourier, το σήμα μπορεί να αναλυθεί στο χώρο των συχνοτήτων σε έναν αριθμό από διακριτές συνιστώσες συχνοτήτων ή φάσμα συχνοτήτων. Για τις ίδιες λοιπόν χρονοσειρές που αναφέραμε παραπάνω, έστω  $F_x(k)$  και  $F_y(k)$ , η συνοχή για συχνότητα  $\lambda$  δίνεται από τη σχέση :

$$Coh_{x,y}(\lambda) = \frac{|F_{x,y}(\lambda)|^2}{F_{x,x}(\lambda)F_{y,y}(\lambda)} \quad (3)$$

Όπου  $F_{x,y}(\lambda)$  είναι το διασταυρωμένο φάσμα (cross spectrum) που στην ουσία περιγράφει τη συσχέτιση των χρονοσειρών στον τομέα των συχνοτήτων, ορισμένο από τον μετασχηματισμό Fourier της διασταυρωμένης συνδιακύμανσης που ακολουθεί :

$$F_{x,y}(\lambda) = \sum_u Cov_{x,y}(u) \times e^{-j\lambda u} \quad (4)$$

Όπου,  $F_{x,x}(\lambda)$  είναι το δυναμικό φάσμα (power spectrum) και το  $F_{y,y}(\lambda)$  αντίστοιχα. Ορίζονται από τις σχέσεις :

$$F_{x,x}(\lambda) = \sum_u Cov_{x,x}(u) \times e^{-j\lambda u} \quad (5)$$

$$F_{y,y}(\lambda) = \sum_u Cov_{y,y}(u) \times e^{-j\lambda u} \quad (6)$$

Η έκφραση της συσχέτισης στο τομέα συχνοτήτων βοηθάει τους ερευνητές να μελετήσουν τη σχέση μίας χρονοσειράς με ένα φυσικό και εγγενή τρόπο. Για παράδειγμα, οι διακυμάνσεις στη ροή του αίματος διαρκούν συνήθως χρονικές περιόδους περίπου 10 δευτερολέπτων. Έτσι η συνοχή ως μετρική σε χαμηλές συχνότητες κάτω από 0.1Hz είναι ιδιαίτερος συνδεδεμένη με την λειτουργική διασυνδεσιμότητα αφού η καρδιακή λειτουργία για παράδειγμα, συνήθως λειτουργεί σε συχνότητες περίπου στα 1.25 Hz. Συνοχή λοιπόν σε μία τέτοια ζώνη συχνοτήτων μπορεί να πρόκυπτε από την καρδιακή λειτουργία και όχι από μία πραγματική λειτουργική διασυνδεσιμότητα μεταξύ περιοχών του εγκεφάλου.

## 2.2 Γραμμικές μέθοδοι που βασίζονται στα δεδομένα και τη δομή τους (Data – driven)

Για να ξεπεράσουμε τους περιορισμούς των μεθόδων που βασίζονται σε κάποιο μοντέλο, αναπτύχθηκαν μέθοδοι οι οποίες είναι ανεξάρτητες από προηγούμενη πληροφορία ή κάποια υπόθεση μοντέλου. Υπάρχουν γενικά δύο βασικές κατηγορίες αυτών των μεθόδων για ανίχνευση λειτουργικής διασυνδεσιμότητας (Functional connectivity) αλλά και γενικότερα ανάλυσης των FMRI. Η πρώτη κατηγορία αφορά μεθόδους Ανάλυσης Πίνακα (Decomposition Methods) όπως για παράδειγμα, η Ανάλυση σε Κύριες Συνιστώσες (Principal Component Analysis), η Ανάλυση Πίνακα σε Ιδιάζουσες Τιμές (Singular Value Decomposition) αλλά και η Ανάλυση σε

Ανεξάρτητες Συνιστώσες (Independent Component Analysis). Αυτή η κατηγορία μεθόδων προσπαθεί να εκφράσει το αυθεντικό σετ δεδομένων FMRI σαν ένα γραμμικό συνδυασμό από ορθογώνια (ανά δύο) γραμμικά ανεξάρτητα διανύσματα (PCA/SVD) ή στατιστικά ανεξάρτητες συνιστώσες (ICA) κτλ. Η δεύτερη κατηγορία μεθόδων βασίζεται στην Ανάλυση σε Πυρήνες (Cluster Analysis) όπως για παράδειγμα είναι η Ασαφής Ανάλυση σε Πυρήνες (Fuzzy Clustering Analysis) ή η Ιεραρχική Ανάλυση σε Πυρήνες (Hierarchical Cluster Analysis). Αυτή η κατηγορία μεθόδων προσπαθεί να εφαρμόσει παραδοσιακές clustering τεχνικές σε δεδομένα FMRI. Οι δύο τύποι μεθόδων είναι διερευνητικές και έτσι βοηθούν στην αναζήτηση λειτουργικής διασυνδεσιμότητας του εγκεφάλου σαν ολότητα.

Στην παρούσα διπλωματική εργασία θα ασχοληθούμε και στη συνέχεια θα εφαρμόσουμε, μεθόδους που ανήκουν στην πρώτη από τις δύο κατηγορίες.

### 2.2.1 Ανάλυση σε κύριες συνιστώσες και ανάλυση πίνακα σε ιδιάζουσες τιμές (PCA/SVD)

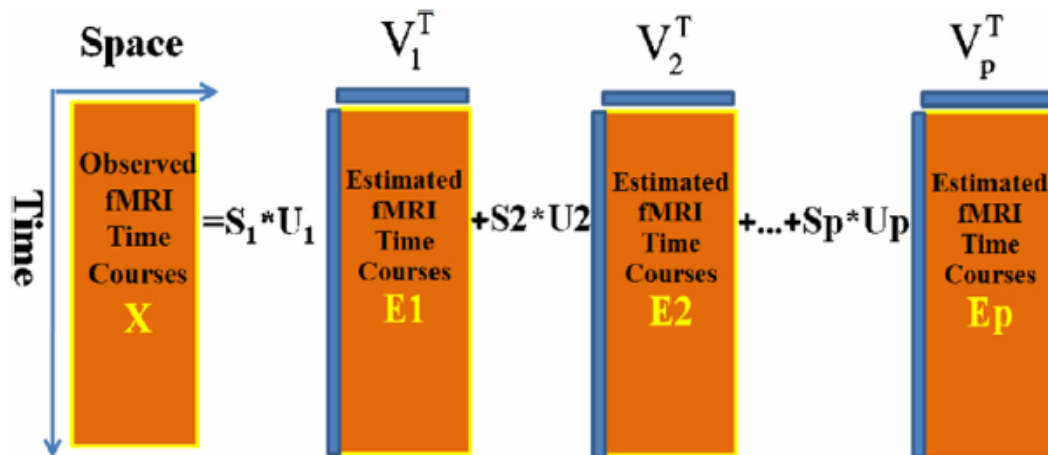
Οι δύο αυτές συγγενικές μέθοδοι έχουν ευρέως χρησιμοποιηθεί σαν βασικές μέθοδοι για την ανάλυση δεδομένων. Επειδή είναι πολύ συγγενικές και σε θεωρητικό επίπεδο θα τις εξετάσουμε μαζί.

Ο πυρήνας και η ουσία των PCA/SVD είναι η αναπαράσταση των παρατηρήσεων των FMRI χρονοσειρών  $X$  με ένα συνδυασμό από ορθογώνιες συνιστώσες. Κάθε συνιστώσα είναι φτιαγμένη από ένα χρονικό πρότυπο (μία κύρια συνιστώσα) πολλαπλασιασμένο επί ενός χωρικού προτύπου (ιδιοχάρτης (Eigen map)). Μαθηματικά, η ανάλυση πίνακα σε ιδιάζουσες τιμές του  $X$  ( $T$  χρονικά σημεία (time points)  $\times N$  εικονοστοιχεία (voxels)) είναι :

$$X = USV^T = \sum_{i=1}^p S_i U_i V_i^T \quad (7)$$

Όπου  $S_i$  είναι η ιδιάζουσα τιμή του  $X$ ,  $U_i$  είναι η  $i$ -οστή κύρια συνιστώσα,  $V_i$  είναι ο αντίστοιχος ιδιοχάρτης (Eigen map) και  $p$  είναι ο αριθμός των επιλεγμένων συνιστωσών. Οι γενικευμένοι ιδιοχάρτες αποκαλύπτουν τη διασυνδεσιμότητα των διάφορων εγκεφαλικών περιοχών : περιοχές με μεγάλες απόλυτες τιμές (είτε θετικές ή αρνητικές) θεωρούνται συσχετιζόμενες. Συνήθως, οι ερευνητές ξεσκαρτάρουν κάποια διανύσματα με πολύ μικρή συγκομιδή στη διασπορά των δεδομένων και έτσι κατά κάποιο τρόπο «καθαρίζουν» τα δεδομένα και διατηρούν την ουσία της πληροφορίας και το πραγματικό σήμα δυνατό.

Αρχικά, μέθοδοι σαν τις PCA/SVD (Εικόνα 5) προτάθηκαν για εφαρμογές σε δεδομένα PET (Friston et al., 1993), αργότερα εφαρμογές των μεθόδων έγιναν και σε fMRI δεδομένα (Baumgartner et al., 2000; Worsley et al., 2005). Παρά την καλή απόδοση στον εντοπισμό των εκτεταμένων περιοχών των συσχετιζόμενων voxels, οι PCA/SVD φαίνεται να αντιμετωπίζουν δυσκολίες που περιορίζουν τη χρήση τους στη μελέτη της λειτουργικής διασυνδεσιμότητας. Για παράδειγμα βρέθηκε ότι οι PCA/SVD αποτυγχάνουν να εντοπίσουν ενεργοποιήσεις σε χαμηλές αναλογίες αντίθεσης-θορύβου (Contrast-to-Noise Ratios), όταν άλλες πηγές διασποράς σήματος ενυπάρχουν όπως για παράδειγμα ο φυσιολογικός θόρυβος (Baumgartner et al., 2000). Αυτό είναι κάτι φυσικό αφού οι PCA/SVD αναλύουν τα δεδομένα σε κύριες συνιστώσες γραμμικά ανεξάρτητες με τρόπο τέτοιο ώστε να εγκλωβίσουν την μεγαλύτερη δυνατή διασπορά. Ακόμα δεν υπάρχει κανόνας για τον αριθμό συνιστωσών που είναι απαραίτητος (ή το πόσες συνιστώσες πρέπει να αγνοηθούν) και οι PCA / SVD μόνο ελαττώνουν δεύτερης τάξης εξαρτήσεις μεταξύ των συνιστωσών. Τελικά, οι μέθοδοι αυτές συχνά χρησιμοποιούνται σαν ένα βήμα προ-επεξεργασίας των δεδομένων. Σκοπός τους είναι η μείωση διάστασης των δεδομένων ώστε να ακολουθήσουν με μεγαλύτερη επιτυχία μέθοδοι όπως είναι η Ανάλυση σε Ανεξάρτητες Συνιστώσες (ICA) για την απομόνωση περιοχών ενδιαφέροντος (ROIs).



**Εικόνα 5.** Μία απεικόνιση της ανάλυσης πίνακα σε ιδιάζουσες τιμές (SVD) δεδομένων FMRI.  $S_i$  είναι ιδιάζουσα τιμή του  $X$ ,  $U_i$  είναι η  $i$ -οστή κύρια συνιστώσα και  $V_i$  είναι ο αντίστοιχος ιδιοχάρτης,  $p$  είναι ο αριθμός των επιλεγμένων συνιστωσών [Li et al., 2009].

## 2.2.2 Ανάλυση σε ανεξάρτητες συνιστώσες (ICA)

Η ανάλυση σε ανεξάρτητες συνιστώσες είναι μία πρόσφατα-αναπτυσσόμενη δημοφιλής μέθοδος για τον εντοπισμό λειτουργικής διασυνδεσιμότητας σε fMRI δεδομένα. Από τη στιγμή που η χρήση της δεν απαιτεί κάποια γνώση και πληροφορία για τα χώρο-χρονικά μοτίβα του πηγαίου σήματος, η ICA είναι μία μέθοδος που ταιριάζει σε μεγάλο βαθμό στην ανάλυση των fMRI σε κατάσταση ηρεμίας (resting state fMRI). Ως επακόλουθο, υπάρχει ένα αυξανόμενο ενδιαφέρον στην εφαρμογή της μεθόδου για τη μελέτη fMRI resting-state και την ανάλυση λειτουργικής διασυνδεσιμότητας.

Όπως έχει αναφερθεί, η μέθοδος Ανάλυσης σε Ανεξάρτητες Συνιστώσες (ICA) ψάχνει να βρει έναν γραμμικό συνδυασμό από συνιστώσες όπως και οι μέθοδοι Ανάλυσης σε Κύριες Συνιστώσες (PCA) και Πίνακα σε Ιδιάζουσες Τιμές (SVD). Η διαφορά των μεθόδων έγκειται στο ότι η ICA βρίσκει συνιστώσες που είναι (κατά το δυνατόν) στατιστικά ανεξάρτητες μεταξύ τους (Hyvarinen et al., 2000) την ώρα που οι μέθοδοι PCA/SVD βρίσκουν ορθογώνιες συνιστώσες. Μαθηματικά λοιπόν, για δεδομένα fMRI  $X$  ( $T$  (time points)  $\times$   $N$  (voxels) ) το μοντέλο ανάλυσης σε ανεξάρτητες συνιστώσες μπορεί να περιγραφεί ως :

$$X = AC = \sum_{i=1}^N A_i C_i \quad (8)$$

Όπου  $C_i$  είναι η  $i$ -οστή υποκείμενη ανεξάρτητη πηγή του σήματος (ανεξάρτητη συνιστώσα).  $A$  είναι ο πίνακας μίξης (mixing matrix) με διαστάσεις  $T \times N$ . Διαφορετικές πηγές σήματος είναι ανεξάρτητες η μία από την άλλη (Ανεξάρτητα ενδεχόμενα μεταξύ τους):

$$P(C_1, C_2, \dots, C_N) = \prod_{i=1}^N P(C_i) \quad (9)$$

Όπου  $P(C_i)$  είναι η πιθανότητα για την συγκεκριμένη  $i$ -οστή ανεξάρτητη πηγή. Δηλώνοντας τώρα τον  $W$  σαν ψεύδο-αντίστροφο του  $A$  (ο  $W$  καλείται επίσης και unmixing πίνακας), μπορούμε να υπολογίσουμε τις ανεξάρτητες συνιστώσες από τη σχέση :

$$C = WX \quad (10)$$

Όσο για την υλοποίηση της ICA (σε υπολογιστικό επίπεδο) που θα χρησιμοποιήσουμε και στο πραγματικό πρόβλημα που θα αναλυθεί εκτενώς στο κεφάλαιο 3, υπάρχουν δυο βασικοί και δημοφιλείς αλγόριθμοι. Ένας είναι ο Infomax (Bell and Sejnowski, 1995) και ο άλλος είναι ο αλγόριθμος ICA του σταθερού σημείου (Fixed-point) (Hyvarinen, 1998) εκ των οποίων και οι δύο εργάζονται πάνω στην ελαχιστοποίηση της κοινής πληροφορίας ανάμεσα στις ανεξάρτητες συνιστώσες  $C_i$ . Ο αλγόριθμος Infomax πετυχαίνει το σκοπό αυτό μεγιστοποιώντας με προσαρμοστικό τρόπο την εντροπία της εξόδου ενός νευρωνικού δικτύου με τέτοιο αριθμό εξόδων όσες είναι και οι ανεξάρτητες συνιστώσες που πρόκειται να υπολογιστούν (ο αριθμός των  $C$

δίνεται ως είσοδος στον αλγόριθμο). Από την άλλη ο αλγόριθμος σταθερού σημείου κάνει χρήση της έννοιας της αρνητικής εντροπίας (Negentropy) (Comon, 1994). Οι Esposito et al. (2002) συγκρίναν τους δύο αλγόριθμους ως προς το αποτέλεσμα και κατέληξαν στο ότι και οι δύο μπορούν να παράγουν αποτελέσματα πολύ υψηλής ακρίβειας. Ωστόσο, και οι δύο αλγόριθμοι έχουν τα πλεονεκτήματά τους: Όσον αφορά τον αλγόριθμο σταθερού σημείου (Fixed-Point) είναι πιο αποδοτικός σε όρους χρονικής και χωρικής ακρίβειας ωστόσο ο αλγόριθμος Infomax είναι καλύτερος σε ένα καθολικό μοντέλο υπολογισμού και μείωσης θορύβου (Esposito et al., 2002).

Αρκετά όμοια με τις PCA/SVD, η ICA αναλύει τα αρχικά δεδομένα σε ένα γραμμικό συνδυασμό από συνιστώσες οι οποίες όμως είναι στατιστικά ανεξάρτητες μεταξύ τους και αντιστοιχούν σε IC χάρτες που μετρούν την ενεργοποίηση των voxels σύμφωνα με την χρονοσειρά που αντιστοιχούν. Στη συνέχεια αν για τους χάρτες θέσουμε ένα σημείο αναφοράς πάνω από το οποίο θα θεωρούμε ότι δύο περιοχές συνδέονται μεταξύ τους και το αντίθετο (threshold map), μπορούμε να αποκτήσουμε τους πίνακες διασυνδεσιμότητας περιοχών του εγκεφάλου που αντιστοιχούν στις ανάλογες ανεξάρτητες πηγές, κάτι που θα δούμε καλύτερα αλλά και στην πράξη, στο επόμενο Κεφάλαιο.

Ωστόσο, κατά την εφαρμογή της μεθόδου ICA μπορεί να γίνουν δύο διαφορετικές θεωρήσεις οι οποίες θα δώσουν και διαφορετικά αποτελέσματα. Το πρώτο βήμα είναι να κατασκευάσουμε έναν 2D πίνακα από τον αρχικό πίνακα των δεδομένων που είναι 4D (Ο πίνακας αυτός των τεσσάρων διαστάσεων είναι αποτέλεσμα ένωσης στο χρόνο τιμών για τις τρεις διαστάσεις των σαρώσεων). Στη συνέχεια :

- Μπορούμε να θεωρήσουμε ότι τα δεδομένα μας αποτελούνται από  $T$  (τα χρονικά σημεία) τυχαίες μεταβλητές κάθε μία από τις οποίες μετρώνται σε  $N$  voxels. Με αυτή τη λογική σαν αποτέλεσμα θα έχουμε  $T$  3D χωρικούς χάρτες ενεργοποίησης. Τότε, κάθε 3D χάρτης «ξεδιπλώνεται» για να δώσει έναν αρχικό πίνακα  $X$  μεγέθους  $N \times T$ . Επομένως ο mixing matrix  $A$  θα είναι διαστάσεων  $T \times T$ .
- Αντίθετα, μπορούμε να θεωρήσουμε ότι τα δεδομένα μας αποτελούνται από  $N$  τυχαίες μεταβλητές κάθε μία από τις οποίες μετρώνται σε  $T$  χρονικά σημεία. Με αυτή τη λογική θα έχουμε  $N$  ανεξάρτητες χρονοσειρές κάθε μία από τις οποίες με μήκος  $T$ . Επομένως καταλήγουμε σε έναν αρχικό πίνακα  $X$  διαστάσεων  $T \times N$  και έναν mixing matrix  $A$  διαστάσεων  $N \times N$

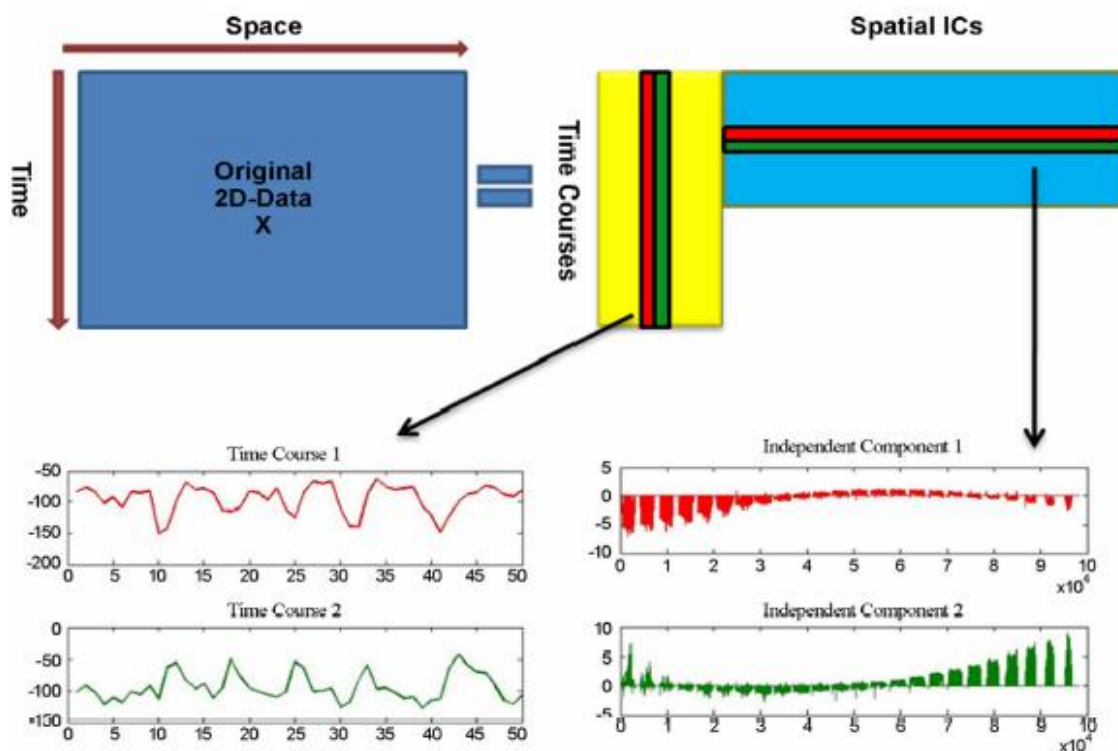
Μπορούμε λοιπόν να γράψουμε το μοντέλο της ICA και ως :

$$X^T = AC^T \quad (11)$$

Στην πρώτη περίπτωση (Εικόνα 6) η υλοποίηση της ICA είναι χωρική (spatial ICA) και οι γραμμές του πίνακα  $C$  περιέχουν χωρικά ανεξάρτητες πηγές (3D maps) μεγέθους  $N$  ενώ στη δεύτερη περίπτωση είναι χρονική



(temporal ICA) και οι γραμμές του πίνακα  $C$  περιέχουν χρονικά ανεξάρτητες πηγές (χρονοσειρές) μεγέθους  $T$ .



**Εικόνα 6.** Μία απεικόνιση για χωρικό ICA (spatial) σε FMRI δεδομένα. Εδώ δηλαδή οι ανεξάρτητες συνιστώσες είναι χωρικά ανεξάρτητες. Έτσι ως αποτέλεσμα έχουμε χωρικά ανεξάρτητους IC χάρτες και τις αντίστοιχες χρονοσειρές στον πίνακα μίξης [Li et al., 2009].

Όσον αφορά το πότε θα πρέπει να αναλύσουμε τα δεδομένα μας σε χωρικά ανεξάρτητες συνιστώσες και χωρικά ανεξάρτητη χρονοσειρά (Spatial ICA) ή χρονικά ανεξάρτητες συνιστώσες και χρονικά ανεξάρτητη χρονοσειρά (temporal ICA), θα διαχωρίσουμε τη μέθοδο ICA όντως σε δύο παραλλαγές, την χωρική (Spatial) και τη χρονική (Temporal) ICA. Όμως τότε ανακύπτει φυσιολογικά το ερώτημα : Ποια παραλλαγή της ICA θα έπρεπε να διαλέξουμε για να εντοπίσουμε λειτουργική διασυνδεσιμότητα σε FMRI;

Από την εισαγωγή της μεθόδου ICA στη μελέτη FMRI δεδομένων (McKeown et al., 1998), τόσο η spatial ICA όσο και η temporal ICA έχουν χρησιμοποιηθεί ευρέως. Ωστόσο το κριτήριο για το ποια από τις δύο παραλλαγές αυτές θα πρέπει να χρησιμοποιούμε φαίνεται να είναι η συγκεκριμένη δραστηριότητα την οποία πρόκειται να μελετήσουμε. Οι ερευνητές ανέφεραν ότι Spatial και Temporal ICA μπορούν να αποκλίνουν ως προς το αποτέλεσμα, ανάλογα με τα χαρακτηριστικά των επικείμενων σημάτων προς εκτίμηση (Calhoun et al., 2001). Συγκεκριμένα αν οι ανεξάρτητες

πηγές που θέλουμε να εντοπίσουμε είναι χωρικά συσχετισμένες αλλά όχι χρονικά, καλύτερη επιλογή θα αποτελεί η Temporal ICA ενώ σε αντίθετη περίπτωση που οι ανεξάρτητες πηγές είναι χρονικά αλλά όχι χωρικά συσχετισμένες η Spatial ICA. Αυτό συμβαίνει επειδή η ICA αναζητά ανεξάρτητες συνιστώσες, έτσι αν όντως υπάρχει συσχέτιση χωρική και όχι χρονική θα ήταν λάθος να αναζητάμε χωρικά ανεξάρτητες συνιστώσες από τη στιγμή που κάτι τέτοιο έρχεται σε αντίθεση με τη βασική υπόθεση της ICA για ανεξαρτησία των συνιστωσών.

Παρόλο βέβαια το αυξανόμενο ενδιαφέρον για την εφαρμογή της ICA σε FMRI resting-state, υπάρχουν κάποια σημεία «παγίδες» που θα πρέπει να αναφέρουμε.

Πρώτα από όλα, όπως αναφέραμε και παραπάνω η ICA βασίζεται στην υπόθεση ότι οι συνιστώσες (πηγές σήματος) είναι ανεξάρτητες, είτε χωρικά είτε χρονικά. Η παραβίαση αυτής της υπόθεσης θα μείωνε πολύ τις πιθανότητες για αποδοτικότητα της ICA (Calhoun et al., 2001).

Δεύτερον, προκύπτουν προβληματισμοί τόσο για τον αριθμό των ανεξάρτητων συνιστωσών που θα πρέπει να υπολογίσουμε όσο και το πως θα μπορούσαμε να κάνουμε threshold έναν IC χάρτη για να εκμεταλλευτούμε τις πληροφορίες του για τη μελέτη λειτουργικής διασυνδεσιμότητας.

Μετά από έρευνα πάνω σε αυτούς τους προβληματισμούς, οι ερευνητές κατέληξαν στο ότι όταν ο αριθμός των ανεξάρτητων συνιστωσών είναι μικρότερος από αυτόν των πραγματικών πηγών σημάτων, τα αποτελέσματα της μεθόδου συσχετίζονται σε μεγάλο βαθμό με τον αριθμό των ανεξάρτητων συνιστωσών (Ma et al., 2007). Όσον αφορά το δεύτερο προβληματισμό, στη πραγματικότητα είναι μία πολύ δύσκολη διαδικασία για να πραγματοποιηθεί άμεσα. Στην πράξη λοιπόν, είναι συνηθισμένο να μετατρέπουμε έναν ανεξάρτητο χάρτη με μία μη κανονική κατανομή σε έναν  $Z$ -χάρτη (στατιστική συνάρτηση  $Z$ ) με κανονική κατανομή (McKeown et al., 1998 ; Zhao et al., 1999). Ωστόσο αποτελέσματα ερευνών έδειξαν ότι η μετατροπή του ανεξάρτητου χάρτη σε  $Z$ -χάρτη τείνει να υπερεκτιμά το λόγο μεταξύ αρνητικών και θετικών τιμών σε κάθε voxel (False-Positive Rate) (Ma et al., 2007). Αυτή η υπερεκτίμηση βέβαια, πολλές φορές δεν είναι τόσο μεγάλη και γι' αυτό σε αρκετές περιπτώσεις γίνεται αποδεκτή.

Επιπλέον η ICA είναι ένα μοντέλο το οποίο παράγει αποτελέσματα χωρίς θόρυβο (noise-free generative model). Τα δεδομένα μας εξηγούνται πλήρως από τις ανεξάρτητες πηγές σημάτων που περιέχονται στον πίνακα  $C$  και τον πίνακα μίξης  $A$  (mixing matrix). Ως εκ τούτου αυτό δεν επιτρέπει την αξιολόγηση της στατιστικής ακρίβειας των πηγών μέσα στα όρια της μηδενικής υπόθεσης. Για να λύσουμε και αυτό το πρόβλημα πρόσφατα αναπτύχθηκε ένα νέο μοντέλο το οποίο ονομάζεται πιθανοτική ICA ή pICA το οποίο υποθέτει ότι παρατηρούμενες  $p$ -διάστατες χρονοσειρές κατασκευάζονται από ένα σετ  $q$  (με  $q < p$ ) στατιστικά ανεξάρτητων μη Γκαουσιανών πηγών (spatial maps) μέσω μιας γραμμικής και στιγμιαίας διαδικασίας «μίξης» που αλλοιώνεται από θόρυβο-κανονικής κατανομής  $\eta(t)$  :

$$X_i = AS_i + \mu + \eta_i \quad (12)$$

Εδώ το  $X_i$  αναφέρεται στο  $p$ -διάστατο διάνυσμα στήλη των ατομικών μετρήσεων στην τοποθεσία του εικονοστοιχείου (voxel)  $i$ ,  $A$  είναι ο πίνακας μίξης, η μεταβλητή  $S_i$  δηλώνει το  $q$ -διάστατο διάνυσμα στήλη των σημάτων της μη

Γκαουσιανής πηγής που περιέχεται στα δεδομένα, το  $\mu$  είναι το σταθερό μέρος (σταθερά) και το  $\eta_i$  δηλώνει τον Γκαουσιανό θόρυβο  $\eta_i \sim N(\mathbf{0}, \sigma^2 \Sigma_i)$  (Beckmann and Smith, 2004).

### 2.2.3 Ανάλυση ICA σε ομάδες (group ICA)

Σε αντίθεση με άλλες μονοπαραγοντικές μεθόδους όπως π.χ. η ανάλυση παλινδρόμησης η ICA δεν επεκτείνεται φυσιολογικά σε μία μέθοδο ικανή να αναδείξει κάποια συμπεράσματα για ομάδες υποκειμένων. Για παράδειγμα, όταν χρησιμοποιούμε ένα γενικό γραμμικό μοντέλο, ο ερευνητής προσδιορίζει τις μεταβλητές ενδιαφέροντος και έτσι η ανάδειξη συμπερασμάτων θα επεκταθεί φυσιολογικά αφού όλες οι ατομικότητες στην ομάδα μοιράζονται τις ίδιες αυτές μεταβλητές ενδιαφέροντος. Σε αντίθεση, για την ICA διαφορετικές ατομικότητες θα έχουν διαφορετικές χρονοσειρές και θα ταξινομηθούν διαφορετικά, έτσι δεν είναι προφανές το πως μπορούμε να επεκτείνουμε την ICA για ομάδες υποκειμένων.

Ωστόσο έχουν προταθεί αρκετές υλοποιήσεις της ICA για ομάδες υποκειμένων (Beckman and Smith, 2005; Calhoun et al., 2001a; Calhoun et al., 2001b; Calhoun et al., 2004a; Esposito et al., 2005; Lukic et al., 2002). Οι διάφορες προσεγγίσεις διαφέρουν σε όρους που αφορούν το πως θα οργανωθούν τα δεδομένα πριν την ICA ανάλυση, τι τύποι δεδομένων είναι διαθέσιμοι και το πως θα πετύχουμε τη στατιστική σημαντικότητα. Παρακάτω θα γίνει παρουσίαση των προσεγγίσεων αυτών καθώς και περιγραφή μίας τυπικής εφαρμογής της μεθόδου group ICA κατά τα πρότυπα των Calhoun et al. (2009)

Οι προσεγγίσεις μπορούν να ταξινομηθούν σε πέντε βασικές κατηγορίες όπως φαίνονται παρακάτω στην εικόνα 7. Η πρώτη προσέγγιση αφορά την διεξαγωγή ICA για κάθε υποκείμενο και την μετέπειτα προσπάθεια να συνδυαστούν τα αποτελέσματα των μελών της ομάδας μέσω διάφορων μεθόδων όπως π.χ. η χωρική συσχέτιση των συνιστωσών (Calhoun et al., 2001a; Esposito et al., 2005). Αυτή η μέθοδος έχει το πλεονέκτημα ότι επιτρέπει ένα μοναδικό χωρικό ή χρονικό χαρακτηριστικό (για κάθε υποκείμενο) όμως έχει το μειονέκτημα ότι αφού τα δεδομένα είναι θορυβώδη οι συνιστώσες δεν είναι απαραίτητο ότι ανακατασκευάζονται κατά τον ίδιο τρόπο για το κάθε υποκείμενο.

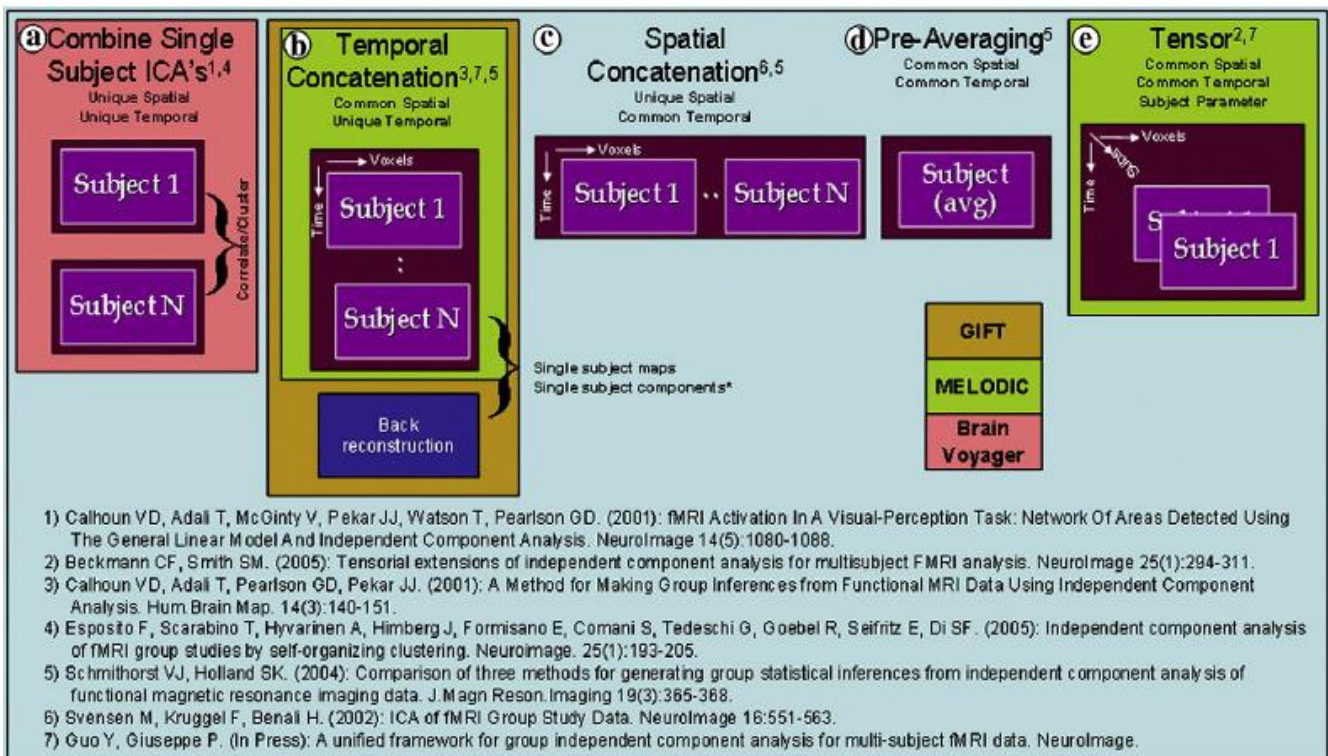
Οι άλλες τέσσερις προσεγγίσεις διεξάγουν την ICA απευθείας στην ομάδα, είναι άμεσες. Το πλεονέκτημα των προσεγγίσεων αυτών είναι ότι πραγματοποιούν μία ICA, η οποία μετά μπορεί να διαιρεθεί σε συγκεκριμένα μέρη του υποκειμένου, έτσι η σύγκριση των διαφορών των υποκειμένων σε μία συνιστώσα είναι άμεση. Δύο από τις τέσσερις είναι η χωρική (spatial concatenation) αλλά και η χρονική αλληλουχία (temporal concatenation) που έχουν εξεταστεί εκτενώς. Στην προσέγγιση με χρονική αλληλουχία επιτρέπεται μία μοναδική ατομική χρονοσειρά για κάθε υποκείμενο, αλλά υποτίθενται κοινός ομαδικός χάρτες. Από την άλλη, η χωρική αλληλουχία επιτρέπει μοναδικούς χάρτες για κάθε υποκείμενο αλλά υποτίθενται κοινές χρονοσειρές. Για τις δύο διαφορετικές προσεγγίσεις, έρευνες έχουν δείξει ότι η προσέγγιση με χρονική αλληλουχία φαίνεται να λειτουργεί καλύτερα για fMRI δεδομένα (Schmithorst and Holland, 2004) και η πιθανότερη εξήγηση σε αυτό είναι ότι οι χρονικές μεταβολές σε

αυτόν τον τύπο δεδομένων είναι αρκετά μεγαλύτερες από τις χωρικές μεταβολές. Για το λόγο αυτό, η χρονική προσέγγιση (χρονική αλληλουχία) των άμεσων μεθόδων για group ICA έχουν χρησιμοποιηθεί ευρέως για fMRI δεδομένα.

Άλλη μία άμεση μέθοδος για group ICA είναι η προ-κανονικοποίηση (Pre-Averaging) η οποία κανονικοποιεί τα δεδομένα πριν τη διεξαγωγή της ICA. Έτσι, η προσέγγιση αυτή είναι υπολογιστικά πιο ελαφριά. Παρ' όλα αυτά όμως, κάνει μία πολύ αυστηρή υπόθεση που απαιτεί μία κοινή χρονοσειρά και έναν κοινό χωρικό χάρτη.

Τέλος, η εναπομείνασα άμεση μέθοδος είναι η μέθοδος του ταυστή (tensor) που ενέχει την εκτίμηση μίας κοινής χρονοσειράς και μίας κοινής εικόνας για κάθε συνιστώσα αλλά επιτρέπει τον υπολογισμό μίας ειδικής παραμέτρου για το κάθε υποκείμενο

Παρακάτω θα περιγράψουμε μία τυπική εφαρμογή της μεθόδου group ICA καθώς και την περαιτέρω ανάλυση που κάνουμε για τις ομαδικές συγκρίσεις και την εξαγωγή συμπερασμάτων.



**Εικόνα 7.** Αναπαράσταση βασικών κατηγοριών group ICA για ανάλυση fMRI [Calhoun et al., 2009]

Τυπική εφαρμογή της μεθόδου group ICA γίνεται με πολλαπλά βήματα μείωσης διάστασης δεδομένων με εφαρμογή π.χ. PCA/SVD (κυρίως PCA λόγω της ευκολίας στον υπολογισμό). Μαθηματικά λοιπόν, έχουμε  $X_i = F_i^{-1} Y_i$  μεγέθους  $L \times V$  μειωμένης διάστασης από το υποκείμενο  $i$ , όπου  $Y_i$  είναι  $K \times V$  πίνακας δεδομένων (περιέχοντας δεδομένα που έχουν προ-επεξεργαστεί και κανονικοποιηθεί),  $F_i^{-1}$  είναι ο  $L \times K$  μειωμένης διάστασης πίνακας (που μειώσαμε λόγω του PCA),  $V$  αριθμός

εικονοστοιχείων,  $K$  αριθμός των χρονικών σημείων των FMRI δεδομένων και  $L$  το μέγεθος της χρονικής διάστασης. Τα μειωμένα δεδομένα από όλα τα υποκείμενα τοποθετούνται σε ένα πίνακα και μειώνονται πάλι σε διάσταση μέσω του PCA (κατά προτίμηση) σε  $N$  διαστάσεις (τον αριθμό των συνιστωσών που θα υπολογίσουμε). Ο πίνακας λοιπόν  $LM \times V$  μειωμένος και με  $M$  υποκείμενα είναι :

$$X = G^{-1} \begin{bmatrix} F_1^{-1} Y_1 \\ \vdots \\ F_M^{-1} Y_M \end{bmatrix} \quad (13)$$

Όπου,  $G^{-1}$  είναι ένας  $N \times LM$  μειωμένος πίνακας και είναι πολλαπλασιασμένος από δεξιά με τον  $LM \times V$  αναμεμειγμένο πίνακα για τα  $M$  υποκείμενα.

Μετά την εφαρμογή της ICA μπορούμε να γράψουμε  $X = A S$  όπου  $A$  είναι ο  $N \times N$  πίνακας μίξης και ο  $S$  είναι  $N \times V$  χάρτης συνιστωσών. Αντικαθιστώντας αυτήν την έκφραση για  $X$  στην (13) και πολλαπλασιάζοντας αριστερά και δεξιά με  $G$ , έχουμε:

$$GAS = \begin{bmatrix} F_1^{-1} Y_1 \\ \vdots \\ F_M^{-1} Y_M \end{bmatrix} \quad (14)$$

Ανοίγοντας τον πίνακα  $G$  για όλα τα υποκείμενα έχω :

$$\begin{bmatrix} G_1 \\ \vdots \\ G_M \end{bmatrix} AS = \begin{bmatrix} F_1^{-1} Y_1 \\ \vdots \\ F_M^{-1} Y_M \end{bmatrix} \quad (15)$$

Μετά γράφουμε την εξίσωση για το  $i$ -οστό υποκείμενο δουλεύοντας μόνο για τα στοιχεία αυτού :

$$G_i AS_i = F_i^{-1} Y_i \quad (16)$$

Ο πίνακας  $S_i$  στην εξίσωση (16) περιέχει χάρτες ατομικούς για κάθε υποκείμενο  $i$  και υπολογίζεται από τη σχέση :

$$S_i = (G_i A)^{-1} F_i^{-1} Y_i \quad (17)$$

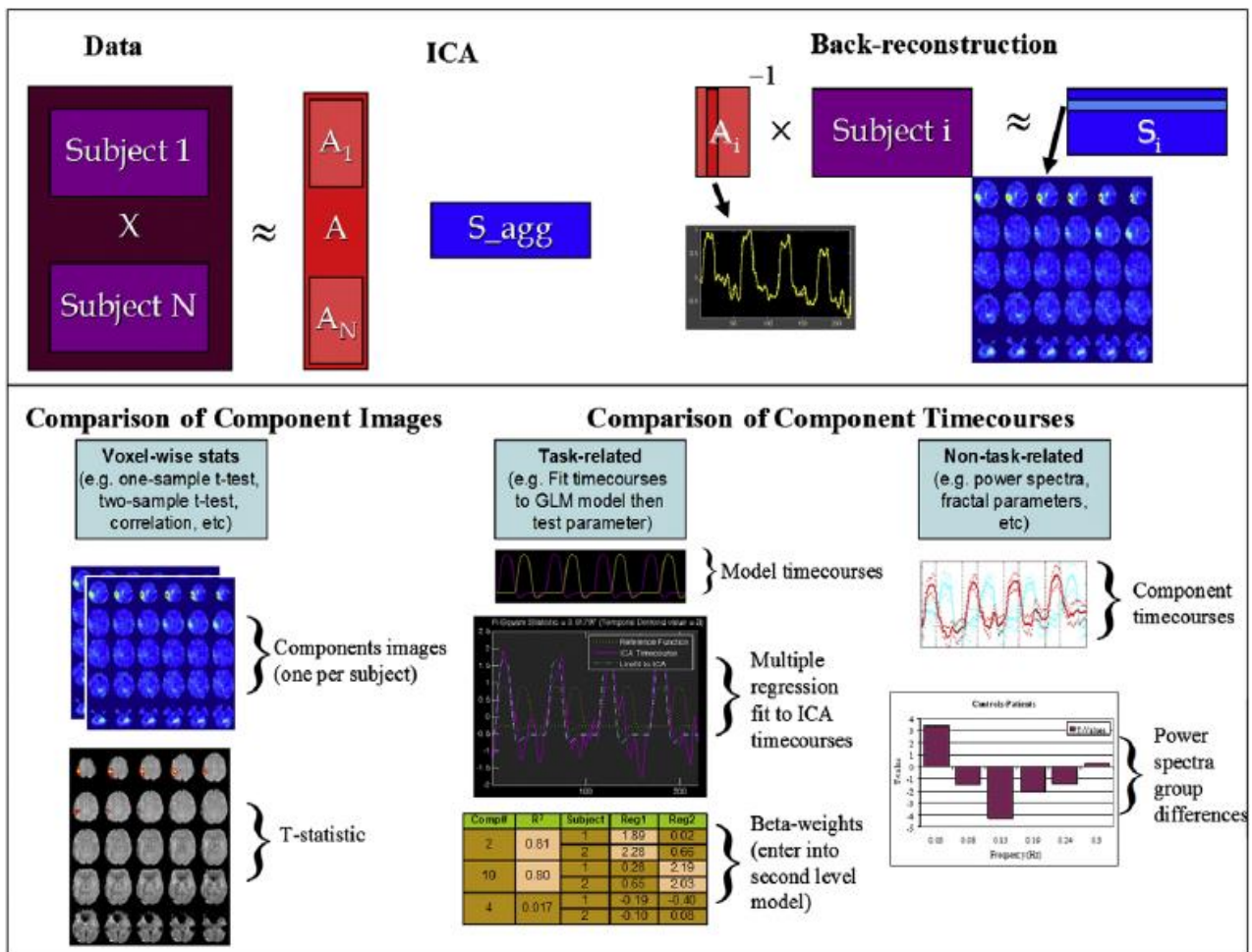
Τελικά πολλαπλασιάζουμε δεξιά και αριστερά με  $F_i$  και έχουμε :

$$Y_i \approx F_i G_i A S_i \quad (18)$$

Η τελευταία σχέση (18) μας παρέχει την ανάλυση ICA για τα δεδομένα του  $i$ -οστού αντικειμένου και περιέχεται στον πίνακα  $Y_i$ . Ο  $N \times V$  πίνακας  $S_i$  περιέχει  $N$  χάρτες πηγής και ο  $K \times N$  πίνακας  $F_i G_i A$  είναι ο πίνακας μίξης ενός συγκεκριμένου υποκειμένου και περιέχει την χρονοσειρά για κάθε μία από τις  $N$  συνιστώσες.

Στη συνέχεια, μετά τα βήματα που περιγράφηκαν παραπάνω για την υλοποίηση (Εικόνα 8), ομαδικές συγκρίσεις και συμπεράσματα μπορούν να εξαχθούν αναλύοντας τις χρονοσειρές των υποκειμένων και τους χωρικούς χάρτες τους (Εικόνα 8). Η ανάλυση αυτή μπορεί να διαχωριστεί σε τρεις κατηγορίες. Η πρώτη αφορά την

εκτίμηση των χωρικών ιδιοτήτων στατιστικά. Για παράδειγμα, με διεξαγωγή t-στατιστικού ελέγχου για τα εικονοστοιχεία. Η δεύτερη αφορά την ανάλυση χρονοσειρών με προσαρμογή ενός γενικού γραμμικού μοντέλου (General Linear Model) π.χ. πολλαπλή παλινδρόμηση, απλώς εδώ οι χρονοσειρές θα αποτελούν εξαρτημένη μεταβλητή. Οι εκτιμώμενοι παράμετροι μπορούν μετά να αναλυθούν σε υψηλότερης τάξης στατιστική ανάλυση για να βγάλουμε κάποια συμπεράσματα μεταξύ των υποκειμένων της ομάδας όπως το κατά πόσο μία συνιστώσα διαμορφώνεται από ένα συγκεκριμένο ερέθισμα, το αν μία συνιστώσα διαμορφώνεται περισσότερο από ένα συγκεκριμένο ερέθισμα, το αν μία ομάδα δείχνει μία πιο δυνατή διαμόρφωση σε μία συγκεκριμένη σωματική διεργασία κτλ. (Stevens et al., 2007). Η ανάλυση αυτή, μας παρέχει έναν πολύ αποδοτικό τρόπο για να βγάζουμε συμπεράσματα για τις συνιστώσες (ανεξάρτητες πηγές σημάτων). Τελευταία κατηγορία είναι η ανάλυση συνιστωσών που δε σχετίζεται με κάποια διεργασία (non-task-related) για δεδομένα σε ήρεμη κατάσταση (resting state). Σε αυτήν την περίπτωση, κάποιος θα μπορούσε να εκτιμήσει διαφορές σε φασματική δύναμη (spectral power) μεταξύ των ομάδων (Garrity et al., 2007) ή να υπολογίσει πρόσθετες παραμέτρους όπως μορφοκλασματική (fractal) διάσταση των χρονοσειρών των υποκειμένων.



**Εικόνα 8.** Πάνω βλέπουμε μία αναπαράσταση των βημάτων που εκτελεί ένας αλγόριθμος Group ICA ενώ κάτω μία αναπαράσταση των συγκρίσεων που μπορούμε να κάνουμε ανάμεσα σε χωρικούς χάρτες των υποκειμένων και τις χρονοσειρές τους [Calhoun et al., 2009].

## 2.3 Μη γραμμικές μέθοδοι ανάλυσης που βασίζονται στα δεδομένα και τη δομή τους

Οι μη γραμμικές μέθοδοι υπερτερούν των κλασικών γραμμικών μεθόδων στη διάκριση των περιοχών του ενδιαφέροντος (ROIs). Επιπλέον, μπορεί επίσης να χρησιμοποιηθούν και σε ένα πιο περιορισμένο πλαίσιο, που επιτρέπει την εξερεύνηση του προτύπου των αιμοδυναμικών αποκρίσεων των περιοχών ενδιαφέροντος. Ωστόσο ανοικτό πρόβλημα παραμένει η επιλογή διάστασης, η οποία δεν είναι ακόμη πλήρως ικανοποιητική. Ωστόσο, έρευνες έχουν αναδείξει τη δύναμη των μη γραμμικών μεθόδων για την εξερεύνηση των δεδομένων, την απεικόνιση και την κατανόηση.

Στα πλαίσια της εργασίας αυτής θα αναλύσουμε την μέθοδο ISOMAP (Tenenbaum et al., 2000) την οποία και θα εφαρμόσουμε για να αντιμετωπίσουμε και το κυρίως πρόβλημα ανάλυσης λειτουργικής διασυνδεσιμότητας στο κεφάλαιο 3.

### 2.3.1 Ισομετρική απεικόνιση χαρακτηριστικών (Isometric feature mapping - ISOMAP)

Η μέθοδος ISOMAP (Tenenbaum et al., 2000) χρησιμοποιήθηκε πρόσφατα από τους Anderson και Cohen (2013) στην προσπάθεια τους για την ανάλυση λειτουργικής διασυνδεσιμότητας σε δεδομένα fMRI resting-state. Η μέθοδος αυτή βασίζεται στην κλασική γραμμική μέθοδο Multi-Dimensional Scaling (MDS) αλλά στοχεύει στη διατήρηση της εγγενούς γεωμετρικής δομής των δεδομένων αφού συλλέγει την γεωδαισιακή απόσταση μεταξύ τους. Στην ουσία υπολογίζει γεωδαισιακές αποστάσεις μεταξύ μακρινών σημείων έχοντας μόνο ως δεδομένη την απόσταση πριν τον μετασχηματισμό στον αρχικό χώρο (input space). Για τα γειτονικά σημεία που βρίσκονται πολύ κοντά μεταξύ τους η απόσταση που γνωρίζουμε είναι μία καλή εκτίμηση της γεωδαισιακής απόστασης, όμως για τα μακρινά σημεία, η γεωδαισιακή απόσταση μπορεί να υπολογιστεί προσθέτοντας στην ακολουθία «μικρά άλματα» από γειτονικά σημεία. Αυτές οι εκτιμήσεις υπολογίζονται ελαφριά προσδιορίζοντας το πιο σύντομο μονοπάτι μεταξύ δύο οποιονδήποτε σημείων σε έναν γράφο με κόμβους τα σημεία και ακμές τις αποστάσεις μεταξύ τους.

Ο αλγόριθμος του ISOMAP λοιπόν έχει τρία βασικά βήματα.

Στο πρώτο βήμα απλώνουμε τα δεδομένα στον αρχικό χώρο (input space) και επιλέγουμε τη συνθήκη ή τη μέθοδο που θα αποφασίζει αν και πότε δύο σημεία θα θεωρούνται γειτονικά (δηλαδή επιτρέπεται η μετακίνηση από το ένα στο άλλο), βασιζόμενοι σε αποστάσεις  $d_X(i, j)$  μεταξύ κάθε ζευγαριού σημείων στον αρχικό χώρο  $X$ . Δύο απλές μέθοδοι για την εκπλήρωση του βήματος αυτού είναι πρώτον η ένωση κάθε σημείου με οποιοδήποτε άλλο σημείο που βρίσκεται σε σταθερή ακτίνα  $\epsilon$  ή απλώς με τη μέθοδο των  $k$  κοντινότερων γειτόνων, βασιζόμενοι στον κανόνα  $k$ -NN ( $k$  - Nearest Neighbours rule). Οι σχέσεις γειτνίασης μεταξύ των σημείων εκφράζονται μέσω ενός γράφου  $G$ . Οι κόμβοι του γράφου  $G$  προσδιορίζουν τα σημεία και οι ακμές

να προσδιορίζουν τα σημεία που γειτνιάζουν με βάρη  $d_X(i, j)$  τις ευκλείδειες αποστάσεις μεταξύ τους. Κατά το δεύτερο βήμα του αλγορίθμου εκτιμώνται οι γεωδαισιακές αποστάσεις  $d_M(i, j)$  (πίνακας γειτνίασης  $\Rightarrow$  συμμετρικός) μεταξύ των σημείων του προτύπου  $M$  και υπολογίζονται οι κοντινότερες αποστάσεις σημείου από σημείο  $d_G(i, j)$  στον γράφο  $G$ . Ένας απλός αλγόριθμος που βρίσκει τις ελάχιστες αποστάσεις μεταξύ των κόμβων ενός γράφου είναι για παράδειγμα ο αλγόριθμος του Dijkstra (αν έχουμε αρνητικά βάρη μπορούμε να χρησιμοποιήσουμε τον αλγόριθμο Floyd-Warshall  $O(V^3)$ ) με υπολογιστική πολυπλοκότητα  $O(V^2)$  όπου  $V$  αριθμός των κόμβων του γράφου  $G$ . Περισσότεροι αλγόριθμοι μπορούν να βρεθούν από **Grams et al. (1994)**. Στο τρίτο και τελευταίο βήμα, ο αλγόριθμος της μεθόδου ISOMAP εφαρμόζει MDS για τον πίνακα με τις ελάχιστες αποστάσεις σημείου από σημείο του γραφήματος  $G$ . Κατασκευάζεται δηλαδή μία ενσωμάτωση του  $D_G = \{d_G(i, j)\}$  για τα δεδομένα σε έναν  $d$ -διάστατο Ευκλείδειο χώρο  $Y$  που διατηρεί βέλτιστα (κατά το δυνατό) την εγγενή δομή της μορφής-προτύπου των δεδομένων. Οι συντεταγμένες  $y_i$  για τα σημεία στον χώρο  $Y$  επιλέγονται με βάση την ελαχιστοποίηση της συνάρτησης κόστους :

$$E = \|\tau(D_G) - \tau(D_Y)\|_{L^2} \quad (19)$$

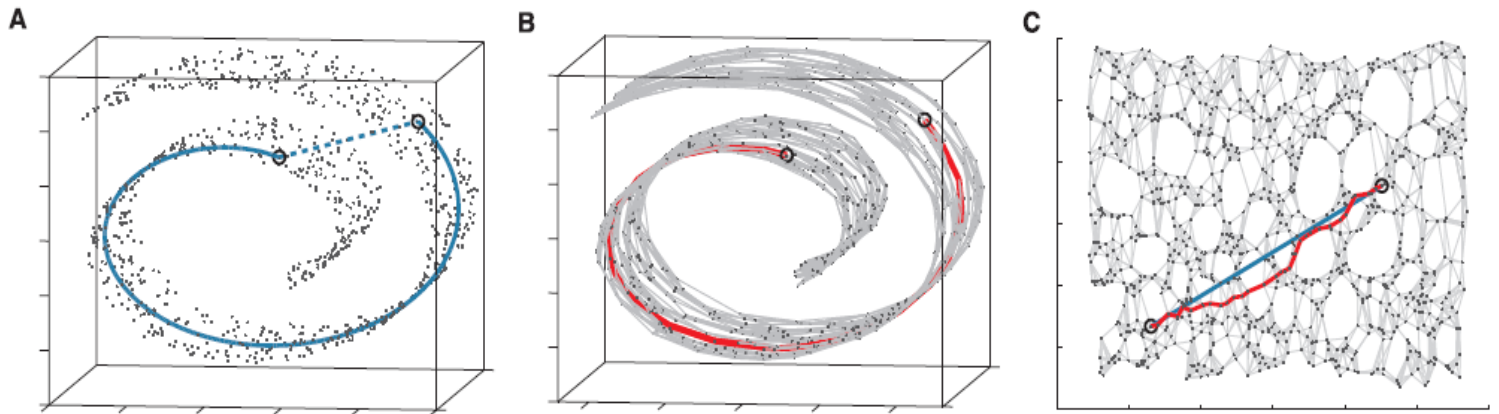
Όπου,  $D_Y$  είναι ο πίνακας των Ευκλείδειων αποστάσεων  $d_Y(i, j) = \{\|y_i - y_j\|\}$ , και  $\|A\|_{L^2}$  η  $L^2$  νόρμα πίνακα  $\sqrt{\sum_{i,j} A_{i,j}^2}$ . Ο τελεστής  $\tau$  μετατρέπει τις αποστάσεις σε εσωτερικά γινόμενα. Πιο συγκεκριμένα,  $\tau(D) = -HS/2$  όπου  $S$  είναι ο πίνακας τετραγωνισμένων αποστάσεων  $\{S_{i,j} = D_{i,j}^2\}$  και ο  $H$  είναι ο πίνακας κεντροποίησης (centering matrix)  $\{H_{ij} = d_{ij} - 1/N\}$  (**Mardia et al., 1979**).

Με τον τρόπο αυτό η MDS προσαρμόζει την γεωμετρία των δεδομένων σε χαμηλότερη διάσταση με μορφή τέτοια ώστε να διατηρεί αποδοτικά την εγγενή γεωμετρία τους. Το ολικό ελάχιστο της συνάρτησης κόστους επιτυγχάνεται θέτοντας συντεταγμένες  $y_i$  για τα σημεία με βάση τα  $d$  επιλεγμένα ιδιοδιανύσματα του πίνακα  $\tau(D_G)$  που αντιστοιχούν στις υψηλότερες ιδιοτιμές (**Mardia et al., 1979**).

Όπως σε PCA/MDS, έτσι και στην μέθοδο ISOMAP η πραγματική διάσταση των δεδομένων μετά την εφαρμογή της μεθόδου μπορεί να εκτιμηθεί με μία μείωση του λάθους καθώς η διάσταση αυξάνεται. Ένα παράδειγμα-εφαρμογή της μεθόδου είναι η ελβετική «κουλούρα» γνωστή και ως «Swiss roll» (Εικόνα 9), όπου όπως θα περιμέναμε οι μέθοδοι MDS και PCA αποτυγχάνουν να αναγνωρίσουν.

Όπως ακριβώς λοιπόν οι PCA, MDS και άλλες κλασσικές γραμμικές μέθοδοι υπόσχονται να μειώσουν τη διάσταση και να βρουν την αληθινή δομή των δεδομένων με γραμμική δομή, έτσι και η ISOMAP και άλλες μη-γραμμικές μέθοδοι υπόσχονται ασυμπτωτικά να βρουν την εγγενή γεωμετρική δομή μη γραμμικών δομών. Όπως η Swiss roll, αυτές οι δομές δεδομένων, των οποίων η γεωμετρία είναι αυτή των κυρτών περιοχών του Ευκλείδειου χώρου αλλά και των οποίων η περιβάλλουσα γεωμετρία ανήκει σε υψηλής διάστασης χώρο, μπορεί να είναι ιδιαίτερα διπλωμένη, στριμμένη ή κυρτή. Για τις μη-Ευκλείδειες δομές, όπως ένα ημισφαίριο ή η επιφάνεια ενός ντόνατ, η ISOMAP παράγει μια βέλτιστη και χαμηλή σε διαστάσεις Ευκλείδεια αναπαράσταση.





**Εικόνα 9.** Οπτική αναπαράσταση στην εφαρμογή του ISOMAP για το Swiss roll. Βλέπουμε αρχικά (A) ποια είναι η γεωδαισιακή απόσταση (με συνεχόμενη γραμμή-καμπύλη) και ποια είναι η ευκλείδεια απόσταση (με συγκεκριμένη γραμμή). Δίπλα ακριβώς (B) βλέπουμε μια αναπαράσταση του γραφήματος αποστάσεων με ακμές μεταξύ σημείων που γειτνιάζουν μεταξύ τους. Με κόκκινη καμπύλη φαίνεται η γεωδαισιακή απόσταση δύο σημείων. Στην τελευταία αναπαράσταση (C) βλέπουμε τη διαφορά που έχουν γεωδαισιακή (κόκκινο) και ευκλείδεια (μπλε) απόσταση μεταξύ τους σε μία απόδοση δύο διαστάσεων του Swiss roll [Tenenbaum et al., 2000]

Η απόδειξη για την ασυμπτωτική σύγκλιση του ISOMAP πηγάζει από το γεγονός ότι όσο ο αριθμός των σημείων των δεδομένων αυξάνεται, οι ελάχιστες αποστάσεις μεταξύ των σημείων  $d_G(i, j)$  προσφέρουν μία όλο και καλύτερη εκτίμηση της γεωδαισιακής απόστασης  $d_M(i, j)$ , φτάνοντας ασυμπτωτικά την τέλεια ακρίβεια για θεωρητικά άπειρα δεδομένα (Tenenbaum et al., 2000). Το πόσο γρήγορα θα συγκλίνουν οι  $d_G(i, j)$  με τις  $d_M(i, j)$  εξαρτάται από συγκεκριμένες παραμέτρους της δομής στον χώρο υψηλής διάστασης (ακτίνα καμπυλότητας και διαχωρισμός εσωτερικών κορυφών) και την πυκνότητα των σημείων της. Στο βαθμό που ένα σύνολο δεδομένων παρουσιάζει ακραίες τιμές αυτών των παραμέτρων ή αποκλίνει από μια ομοιόμορφη πυκνότητα, η ασυμπτωτική σύγκλιση εξακολουθεί να ισχύει, αλλά το μέγεθος του δείγματος που απαιτείται για την εκτίμηση γεωδαισιακής απόστασης με ακρίβεια μπορεί να είναι εξωπραγματικά μεγάλο.

Η ISOMAP δηλαδή χρησιμοποιεί έναν απλό τρόπο να αναλύσει και να χειριστεί δεδομένα υψηλής διάστασης διατηρώντας την φυσική-εγγενή γεωμετρία των δεδομένων. Ας δούμε λοιπόν μερικά παραδείγματα εφαρμογών της μεθόδου σε διαφορετικής φύσεως δεδομένα από Tenenbaum et al. (2000).

Στην εικόνα 10 βλέπουμε ένα παράδειγμα εφαρμογής της μεθόδου σε δεδομένα που αφορούν συνθετικές φωτογραφίες προσώπων που γνωρίζουμε ότι έχουν 3 βαθμούς ελευθερίας. Η μέθοδος ISOMAP μειώνει αποδοτικά τη διάσταση των δεδομένων και διαχωρίζει τα δεδομένα βάση εγγενών-πραγματικών χαρακτηριστικών που είναι καλά κρυμμένα στα δεδομένα (στην περίπτωση αυτή η στάση του προσώπου της φωτογραφίας στο χώρο).

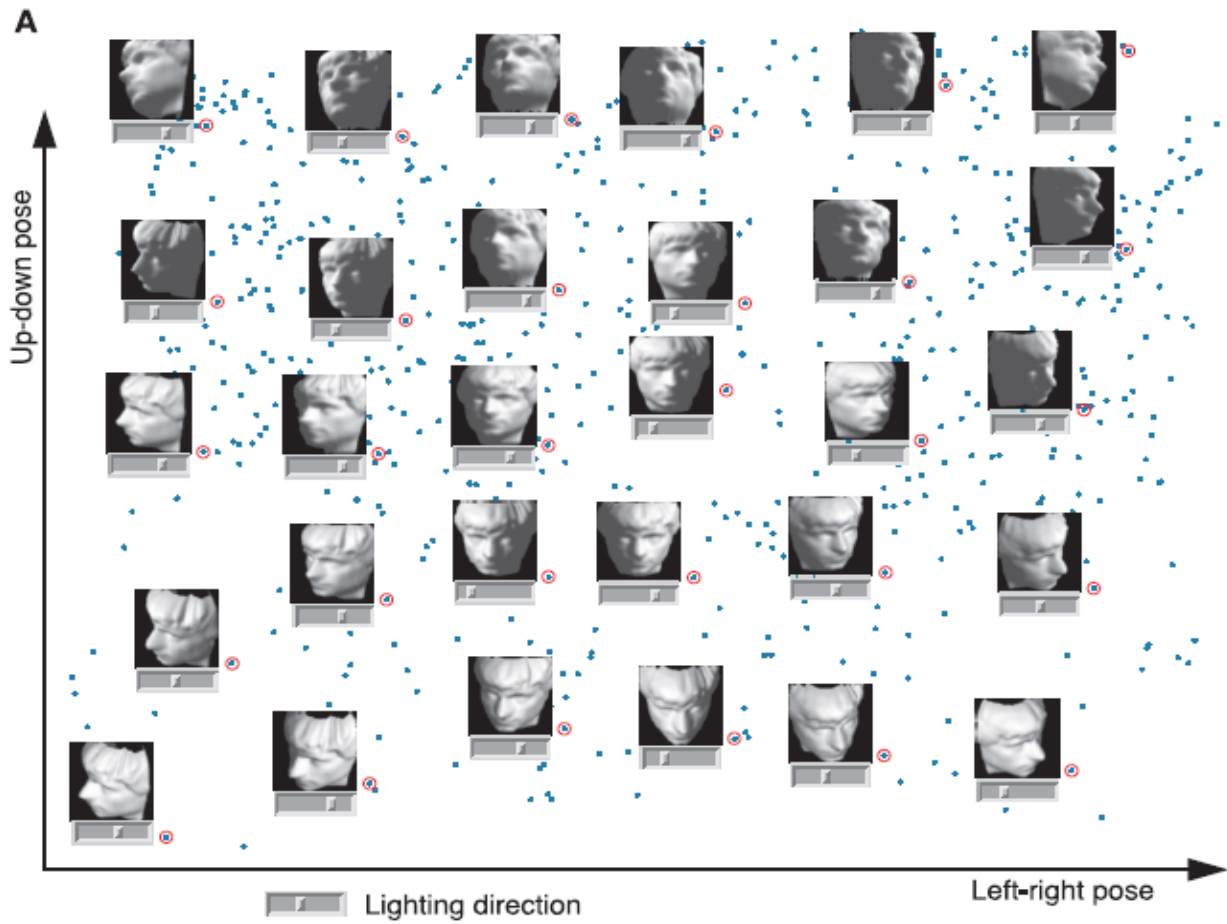
Στην εικόνα 11 παρατηρούμε ότι η μέθοδος ανταπεξέρχεται ικανοποιητικά και σε περισσότερο «δύσκολα» δεδομένα, από ψηφία σχεδιασμένα από ανθρώπινο χέρι, τα οποία δεν έχουν ξεκάθαρη μορφή. Ωστόσο, η μέθοδος χειρίζεται πολύ καλά και αυτά

τα δεδομένα και εξάγει σημαντικά χαρακτηριστικά που εξηγούν την εγγενή δομή τους (στην περίπτωση αυτή ανάλογα με την καμπυλότητα στο σχεδιασμό του άνω και κάτω μέρους του ψηφίου), πράγμα που το PCA αλλά και το MDS σαν γραμμικές μέθοδοι δεν θα μπορούσαν να πετύχουν.

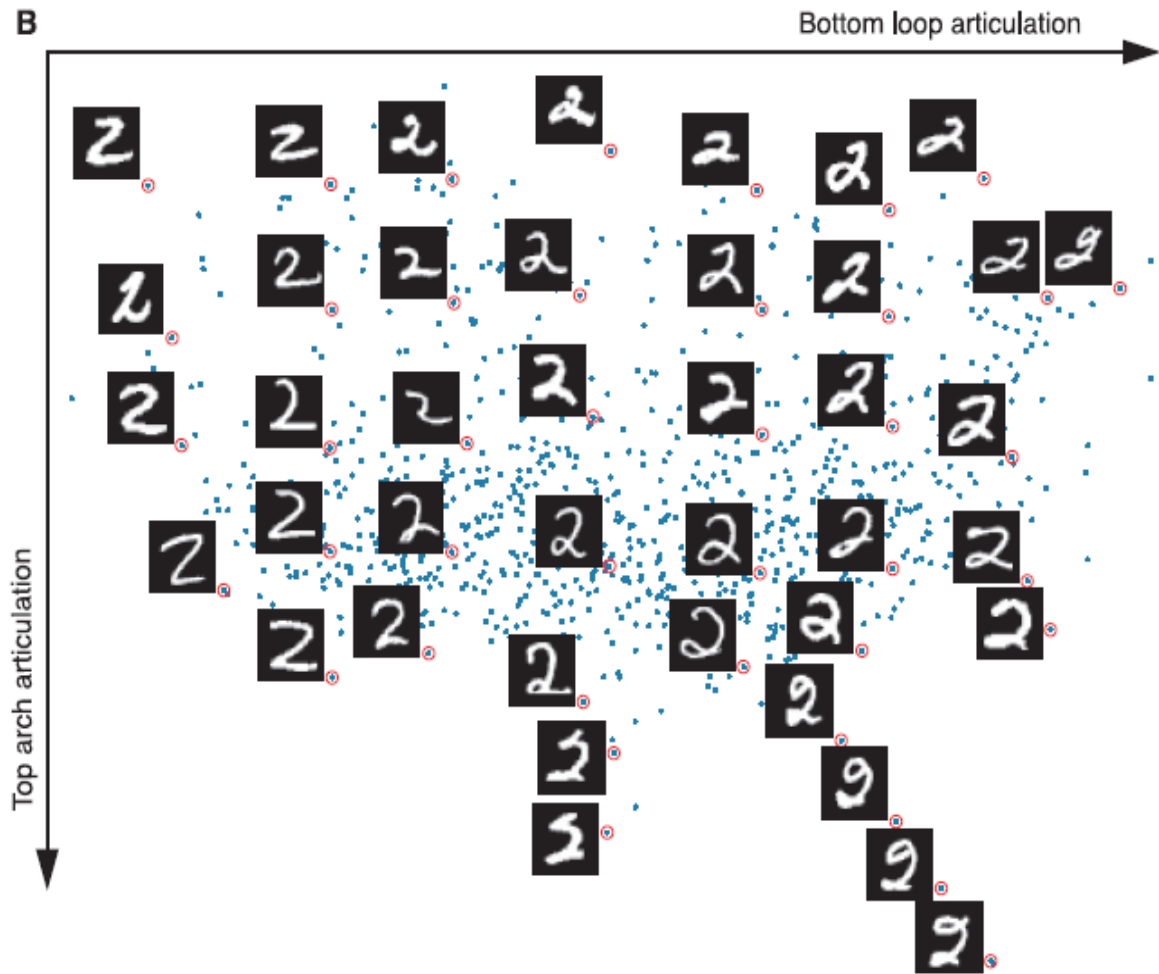
Τέλος, στην εικόνα 12 βλέπουμε συγκεντρωτικά τα προαναφερθέντα σετ δεδομένων και ένα πρόσθετο σετ δεδομένων που αφορά «θορυβώδεις» πραγματικές φωτογραφίες που απεικονίζουν ένα ανθρώπινο χέρι που αλλάζει στάσεις ως προς την έκταση των δακτύλων και την περιστροφή του βραχίονα. Η εικόνα αυτή απεικονίζει τη γεωδαισιακή απόσταση τεσσάρων σημείων της εκάστοτε εφαρμογής (που αναφέραμε παραπάνω) εντόπισε η ISOMAP σαν αυτή να ήταν μία ευθεία γραμμή. Η διαδρομή δηλαδή που κάνει κάποιος για να φτάσει από ένα σημείο αφετηρία σε ένα τερματικό σημείο. Η διαδρομή που βλέπουμε είναι ουσιαστικά η γεωδαισιακή διαδρομή.

Συμπερασματικά, υπάρχουν πολλές μη γραμμικές μέθοδοι μείωσης διάστασης για την ανάδειξη σημαντικών χαρακτηριστικών - παραγόντων για δεδομένα υψηλής διάστασης που έχουν ευρεθεί και όλες έχουν θετικά και αρνητικά χαρακτηριστικά ανάλογα με το πρόβλημα που αντιμετωπίζεται. Ωστόσο, Όλες οι μέθοδοι ιδίως κιάλας η εφαρμογή τους σε ογκώδη δεδομένα (big data) είναι προσεγγιστικές και ενέχουν κινδύνους για λανθασμένα συμπεράσματα. Βέβαια, με την εμπειρία αρκετών εφαρμογών-πειραμάτων έχουμε καταλήξει στο γεγονός ότι κάποιες μέθοδοι είναι καταλληλότερες από κάποιες άλλες ανάλογα με τον τύπο των δεδομένων με τον οποίο δουλεύουμε, το υπολογιστικό κόστος που διατιθέμεθα να πληρώσουμε, το πόσο αραιά ή θορυβώδη είναι τα δεδομένα κτλ.

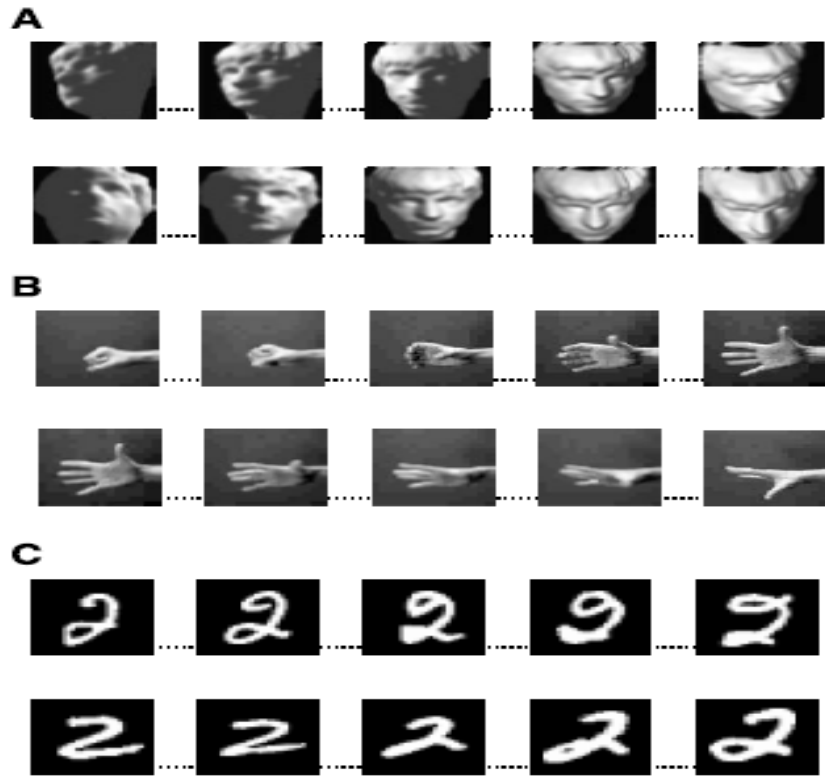
Στο επόμενο κεφάλαιο, θα χρησιμοποιήσουμε κάποιες από τις μεθόδους που αναφέραμε στο κεφάλαιο αυτό για την αντιμετώπιση ενός προβλήματος ανίχνευσης λειτουργικής διασυνδεσιμότητας εγκεφαλικών σημάτων σε πραγματικά δεδομένα FMRI resting state.



**Εικόνα 10.** Συνθετικές φωτογραφίες από πρόσωπα, παρατηρούμε ότι η ISOMAP ταξινομεί τα σημεία με βάση πραγματικά χαρακτηριστικά, τα οποία στην προκειμένη είναι η στάση του προσώπου αλλά η κατεύθυνση του φωτός [Tenenbaum et al., 2000].



**Εικόνα 11.** Βλέπουμε το ψηφίο 2 σχεδιασμένο από ανθρώπινο χέρι, παρατηρούμε ότι ουσιαστικά δεν υπάρχει ξεκάθαρο πρότυπο για το δύο, όμως και πάλι η ISOMAP αναδεικνύει δύο παράγοντες που εξηγούν τα δεδομένα. Εδώ ανάλογα με τη καμπυλότητα στο σχεδιασμό του άνω και κάτω μέρους του ψηφίου [Tenenbaum et al., 2000].



**Εικόνα 12.** Εδώ έχουμε συγκεντρωτικά 3 σετ δεδομένων τα οποία αναφέραμε πιο πάνω. Βλέπουμε τη γεωδαισιακή απόσταση τεσσάρων σημείων της εκάστοτε εφαρμογής που εντόπισε η ISOMAP σαν αυτή να ήταν μία ευθεία γραμμή. Η διαδρομή δηλαδή που κάνει κάποιος για να φτάσει από ένα σημείο σε ένα άλλο [Tenenbaum et al., 2000].

### 3. Ανάλυση λειτουργικής διασυνδεσιμότητας σε πραγματικά δεδομένα Λειτουργικής Απεικόνισης Μαγνητικού Συντονισμού

---

Στο κεφάλαιο αυτό θα αναλύσουμε το κύριο πρόβλημα της εργασίας που είναι η ανάλυση λειτουργικής διασυνδεσιμότητας σε πραγματικά δεδομένα FMRI. Αρχικά θα παρουσιαστούν τα δημογραφικά χαρακτηριστικά των δεδομένων μας και στη συνέχεια θα γίνει μία αναδρομή στις μέχρι τώρα ερευνητικές προσπάθειες και κεκτημένα πολύ ειδικά στο πεδίο ανάλυσης που μας ενδιαφέρει. Μετά την αναδρομή αυτή θα ακολουθήσουν τα βήματα και οι μέθοδοι που ακολουθήθηκαν για την διεκπεραίωση της ανάλυσης μας σε υποκεφάλαια.

#### 3.1 Παρουσίαση των δεδομένων FMRI του προβλήματος

Τα δεδομένα οποία θα αναλύσουμε αποτελούν πειραματικά δεδομένα λειτουργικής απεικόνισης μαγνητικού συντονισμού (FMRI) σε ήρεμη κατάσταση (resting state) και είναι ανοικτά για το κοινό και την επιστημονική κοινότητα στην διεύθυνση :

[http://fcon\\_1000.projects.nitrc.org/indi/retro/cobre.html](http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html).

Τα δεδομένα αυτά παράχθηκαν και εν συνεχεία δόθηκαν στο κοινό και την επιστημονική κοινότητα από ένα Κέντρο Βιοϊατρικής Έρευνας Αριστείας (Center of biomedical research excellence). Συνοπτικά, τα Κέντρα Ιατροβιολογικών Ερευνών Αριστείας ( COBRE ) είναι κέντρα υπό την ηγεσία και χρηματοδότηση του Εθνικού Ιδρύματος Υγείας των Η.Π.Α . Χρηματοδοτούν ερευνητές με εμπειρία και ειδικευση στο θέμα της πρότασης επιχορήγησης και υποστηρίζουν θεματικά, διεπιστημονικά κέντρα που στοχεύουν στην ενίσχυση της θεσμικής ικανότητας της βιοϊατρικής έρευνας.

Το σετ των δεδομένων μας αποτελείται από FMRI σε κατάσταση χαλάρωσης (resting state) ανάμεσα σε 72 ασθενείς διαγνωσμένους με κάποιο τύπο σχιζοφρένειας και 74 υγιείς ανθρώπους που συμμετείχαν στο πείραμα. Όλοι οι συμμετέχοντες στην πειραματική διαδικασία είχαν ηλικία από 18-65 χρονών.

Παρακάτω και συγκεκριμένα στην εικόνα 13 βλέπουμε τον δημογραφικό πίνακα με στοιχεία του δείγματος.

	N	Age (SD)	% Female	% Right-Handed
Schizophrenia	72	38.16 (13.89)	0.19	0.83
Patients	74	35.82 (11.58)	0.31	0.96

**Εικόνα 13.** Δημογραφικός πίνακας που μας παρουσιάζει το πλήθος  $N$  των υγιών και σχιζοφρενών του δείγματος, τη μέση ηλικία (Age) και τυπική απόκλιση των δειγμάτων εντός της παρένθεσης (SD) καθώς και ποιο ποσοστό αυτών είναι αντίστοιχα γυναίκες και δεξιόχειρες [Anderson and Cohen, 2013].

Όλοι οι συμμετέχοντες που έλαβαν μέρος στο πείραμα για την εξαγωγή των δεδομένων πέρασαν από μαγνητικό τομογράφο. Αντίθετα, εξαιρέθηκαν αυτοί οι οποίοι είχαν κάποιο ιστορικό νευρολογικής πάθησης, νοητικής καθυστέρησης, σοβαρό εγκεφαλικό τραύμα με περισσότερο από πέντε λεπτά απώλεια των αισθήσεων, χρήση ναρκωτικών ουσιών ή εξάρτηση έως και 12 μήνες πριν από τη συμμετοχή τους στο πείραμα. Οι πληροφορίες για την διάγνωση των συμμετεχόντων έγινε με χρήση της δομημένης κλινικής συνέντευξης για την ταξινόμηση σύμφωνα με το Διαγνωστικό και Στατιστικό Εγχειρίδιο των Ψυχικών Διαταραχών (DSM).

Τα δεδομένα ήρεμης κατάστασης (resting state) συλλέχθηκαν με ενιαία λήψη με τη μέθοδο  $k$ -space echo-planar imaging (EPI), με διόρθωση αναβάθμισης δειγματοληψίας χρησιμοποιώντας διασυνδυκτική γραμμή (intercommissural line) (AC - PC). Επίσης για τα χαρακτηριστικά του πειράματος έχουμε τις παραμέτρους TR: 2 s, TE : 29 ms, το μέγεθος πίνακα : 64x64 σε 32 φέτες του εγκεφάλου και το μέγεθος του εικονοστοιχείου (voxel): 3x3x4 mm<sup>3</sup>.

Η παράμετρος TE (echo time) αντιπροσωπεύει το χρονικό διάστημα που μεσολαβεί ανάμεσα στην εκπομπή των RF (radio frequency) παλμών και τη μέτρηση του σήματος που εκπέμπουν οι πυρήνες πίσω λόγω του φαινομένου του πυρηνικού συντονισμού (βλέπε Εισαγωγή).

Η παράμετρος TR (repetition time) αφορά το χρονικό διάστημα που μεσολαβεί από την εκπομπή ενός RF παλμού μέχρι τον επόμενη εκπομπή.

Η ανάλυση των δεδομένων και το υπολογιστικό κομμάτι της εργασίας έγινε με χρήση του στατιστικού πακέτου FSL για την προ-επεξεργασία των δεδομένων και την ανάλυση τους σε ανεξάρτητες συνιστώσες, καθώς και της γλώσσας R, που είναι μία γλώσσα κυρίως με στατιστικές εφαρμογές και βιβλιοθήκες που εξυπηρετούν τους σκοπούς της ανάλυσης μας. Η ανάλυση μας έλαβε χώρα σε περιβάλλον Linux και στο τέλος της παρούσας εργασίας (Παράρτημα) παρατίθεται ο κώδικας που χρησιμοποιήθηκε για την ανάλυση με σκοπό την εύκολη επαναπαραγωγή των αποτελεσμάτων και τη μετέπειτα βελτίωση και αξιοποίηση του κώδικα αυτού από τον αναγνώστη.

### 3.2 Μια επισκόπηση για τα FMRI δεδομένα και αναφορές για τις μέχρι σήμερα προσπάθειες ανάλυσης λειτουργικής διασυνδεσιμότητας

Η λειτουργική απεικόνιση μαγνητικού συντονισμού είναι μία τυπική διαδικασία που μετράει την εγκεφαλική λειτουργία στο χρόνο όπως αυτή αναλύθηκε και στην εισαγωγή της εργασίας. Το σήμα του επιπέδου εξάρτησης οξυγόνωσης αίματος (Blood Oxygen Level Dependent) είναι μία έμμεση αντανάκλαση της δραστηριότητας των νευρώνων που ανιχνεύεται κατά την διάρκεια μίας FMRI σάρωσης και η ανάλυση του διεξάγεται υπό την υπόθεση ότι η δραστηριότητα των νευρώνων συμπίπτει με την αυξημένη ροή αίματος. Η αύξηση στη ροή του αίματος κατά την δραστηριοποίηση των νευρώνων είναι γνωστή ως αιμοδυναμική απόκριση (**Kim et al., 1999**). Όταν υπάρξει ενεργοποίηση σε μία περιοχή του εγκεφάλου, οξυγονωμένη αιμοσφαιρίνη ρέει στην περιοχή για να αυξήσει τοπικά την συγκέντρωση οξυγόνου. Η δεοξυαιμοσφαιρίνη (αιμοσφαιρίνη που δεν έχει δεσμευμένο οξυγόνο) έχει ένα γρηγορότερα εξασθενούμενο MR σήμα (μικρότερο  $T_2^*$ ) από ότι η οξυαιμοσφαιρίνη (**Cohen and Bookheimer, 1994**), έτσι το σήμα από τις καλά οξυγονωμένες περιοχές έχει ως αποτέλεσμα ένα πιο δυνατό MR σήμα από περιοχές του εγκεφάλου που δεν έχουν αυξημένη ροή αίματος. Συμπερασματικά λοιπόν, περιοχές με αυξημένη νευρωνική δραστηριότητα δίνουν ένα δυνατότερο MR σήμα που αντικατοπτρίζει έμμεσα την δυναμική της νευρωνικής δραστηριότητας.

Η 4-διαστάσεων FMRI σάρωση μπορεί να χρησιμοποιηθεί για ανακάλυψη ανατομικών περιοχών που είναι υπαίτιες για συγκεκριμένες διαδικασίες όπως η αντίληψη της γλώσσας (language processing) (**Bookheimer, 2002**), η αναγνώριση προσώπου (**Gauthier et al., 1999**) ή ακόμα και για διάγνωση ελλατωματικής περιοχής που σχετίζεται με γνωστικές διαταραχές όπως η νόσος Alzheimer, τραυματική εγκεφαλική βλάβη ή σχιζοφρένεια (**Ford et al., 2003; Anderson et al., 2010**). Τέτοιες μελέτες τυπικά αναλύουν ροή αίματος σε περιοχές του εγκεφάλου για να ανακαλύψουν περιοχές που ενεργοποιούνται κατά τη διάρκεια μίας διαδικασίας ή συγκρίνουν τα σήματα BOLD δύο ομάδων, όπως ασθενείς με Alzheimer και υγιείς ανθρώπους, αναδεικνύοντας διαφορές που θα μπορούσαν να αποτελούν αιτίες για τη γνωστική βλάβη.

Η μέθοδος της λειτουργικής διασυνδεσιμότητας δικτύου (Functional Connectivity Network) που θα εφαρμόσουμε και στην ανάλυση μας παρακάτω χρησιμοποιείται για να εξεταστεί αν ο συγχρονισμός ανάμεσα σε ανατομικά-οριζόμενες εγκεφαλικές περιοχές ή λειτουργικά - υποτιθέμενα δίκτυα διαφέρει λόγω ηλικίας, ασθένειας κτλ. ή αν η διαφορά αυτή σχετίζεται με μία διαδικασία που διεξάγεται. Διαφορές στη διασυνδεσιμότητα μεταξύ εγκεφαλικών δικτύων θεωρείται από την επιστημονική κοινότητα ότι μπορεί να είναι υπεύθυνες για αρκετές διαταραχές-ανωμαλίες όπως είναι ο αυτισμός (**Koshino et al., 2005**) και η σχιζοφρένεια (**Liang et al., 2006**). Πιο γενικά, η χρονική διασυνδεσιμότητα μεταξύ εγκεφαλικών περιοχών έχει χρησιμοποιηθεί ευρέως σε δεδομένα FMRI στην προσπάθεια ανακάλυψης σχέσεων επιρροής ανάμεσα σε νευρωνικούς πληθυσμούς (**Roebroeck et al., 2005**) κάνοντας χρήση της αιτιότητας Granger αλλά και για να εντοπιστούν οι διαφορές ανάμεσα σε σχιζοφρενείς και υγιείς ανθρώπους (**Garrity et**



al., 2007; Jafri et al., 2008; Anderson et al., 2010; Yu et al., 2011), χρησιμοποιώντας μέτρα πολλαπλής συσχέτισης για την ανάλυση των χρονοσειρών.

Συγκεκριμένα για τη σχιζοφρένεια, αλλοιωμένες ιδιότητες small-world δικτύων (η τυπική απόσταση  $L$  μεταξύ δύο τυχαίων κόμβων μεγαλώνει αναλογικά με το λογάριθμο του αριθμού των κόμβων  $N$  του δικτύου) βρέθηκαν μετά από σύγκριση σχιζοφρενών-υγιών ανάμεσα σε 90 φλοιώδεις και υποφλοιώδεις περιοχές του εγκεφάλου (Liu et al., 2008). Ενισχυμένη τοπική λειτουργική διασυνδεσιμότητα σε διάστημα 0.06-0.125 Hz βρέθηκε στη σχιζοφρένεια, μαζί με μειωμένη ένταση από τους Basset et al. (2008). Σε ένα default mode network (ένα δίκτυο αλληλεπιδρώντων περιοχών του εγκεφάλου με δραστηριότητα σε μεγάλο βαθμό συσχετιζόμενη μεταξύ αυτών και διακριτή από άλλα δίκτυα του εγκεφάλου), βρέθηκαν ασυνήθιστα μεγάλη λειτουργική διασυνδεσιμότητα και αλλοιωμένη χρονική συχνότητα (Garrity et al., 2007; Whitfield-Gabrieli et al., 2009). Οι ασθενείς με σχιζοφρένεια είχαν μεγαλύτερη συσχέτιση ανάμεσα σε επτά επιλεγμένα δίκτυα ήρεμης κατάστασης (resting state) από ότι είχαν οι υγιείς (Jafri et al., 2008) και διαφορετικά τοπολογικά μέτρα βρέθηκαν ανάμεσα σε δίκτυα ήρεμης κατάστασης που διαγνώστηκαν από τη μέθοδο group-ICA (Yu et al., 2011). Συλλογικά, οι μελέτες αυτές αλλά και πολλές άλλες, προτείνουν ότι όσον αφορά τη σχιζοφρένεια, μέτρα λειτουργικής διασυνδεσιμότητας μπορούν να χρησιμοποιηθούν για τον εντοπισμό γνωρισμάτων που είναι χαρακτηριστικά της ίδιας της ασθένειας.

Επειδή η διασυνδεσιμότητα εξαρτάται από το πως ορίζονται δίκτυα ή περιοχές και το πως οι γραφικές ιδιότητες των περιοχών αυτών μπορούν να μετρηθούν (Toppi et al., 2012; Zalesky et al., 2012), είναι ζωτικής σημασίας για τους ερευνητές να μπορούν να προσαρμόσουν την ανάλυση διασυνδεσιμότητας στα δικά τους δεδομένα έτσι ώστε να επιτρέπουν στην επιστημονική κοινότητα να κρίνει τόσο την προ-υπάρχουσα γνώση τους όσο και το αποτέλεσμα της ανάλυσης αυτής καθ' αυτής. Για παράδειγμα, οι Sato et al. (2010), εφάρμοσαν ανάλυση λειτουργικής διασυνδεσιμότητας ανάμεσα σε περιοχές ενδιαφέροντος (ROIs), τη στιγμή που οι Chu et al. (2011) ανέλυσαν τη διασυνδεσιμότητα ανάμεσα στα εικονοστοιχεία. Όμοια, οι Yu et al. (2011) εφάρμοσαν τη μέθοδο λειτουργικής διασυνδεσιμότητας δικτύου ανάμεσα σε ανεξάρτητες συνιστώσες ενός υποκειμένου εφαρμόζοντας ICA (Anderson et al., 2010).

Η μεγαλύτερη πρόκληση στην ταξινόμηση των FMRI είναι η αφθονία των παρατηρήσεων σε μία σάρωση, οι περισσότερες από τις οποίες είναι συσχετισμένες σε μεγάλο βαθμό τόσο ως προς το χρόνο όσο και το χώρο. Παρόλο που πολλά από τα εικονοστοιχεία θα είναι άδεια, αυτά δεν θα είναι συστηματικά άδεια για όλα τα υποκείμενα σαν αποτέλεσμα των διαφορών στο μέγεθος και το σχήμα του εγκεφάλου. Επειδή πολλά από τα σημεία στα δεδομένα είναι πλεονάζοντα, χρησιμοποιούνται τεχνικές μείωσης διάστασης δημιουργώντας στατιστικά δελτία ατομικά για τα εικονοστοιχεία ( $t$ -έλεγχοι, έλεγχοι συσχέτισης), απομονώνοντας «περιοχές του ενδιαφέροντος» (ROI) ή εφαρμόζοντας κλασσικές μεθόδους μείωσης διάστασης όπως το PCA για να αναλυθεί ολόκληρη η σάρωση σε ορθογώνιες πηγές σήματος στο χρόνο. Νεότερες μέθοδοι όπως αυτή του ICA (Hyvarinen and Oja, 2000) μιμούνται την προσέγγιση του PCA με τη διαφορά ότι επιβάλλουν πρόσθετες υποθέσεις όπως είναι η στατιστική ανεξαρτησία στην περίπτωση αυτή (ICA).

Μόλις ολοκληρωθούν τα βήματα της εξαγωγής των κυρίων χαρακτηριστικών και της μείωσης διάστασης, τα μειωμένα σε διάσταση δεδομένα μπορούν να οδηγηθούν σε ταξινομητές όπως για παράδειγμα είναι ο SVM, τα τυχαία δάση (random forests) και οι αλγόριθμοι ώθησης (boosting algorithms). Αυτές οι τεχνικές ταξινόμησης έχουν χρησιμοποιηθεί πρόσφατα για να διαχωριστούν σαρώσεις PET (Positron Emission Tomography) με ασθενείς που είναι θετικοί στον HIV από υγιείς ανθρώπους (Liow et al., 2000), για να διαχωριστούν σαρώσεις fMRI με ασθενείς εθισμένους σε ναρκωτικές ουσίες από υγιείς (Zhang and Samaras, 2005) καθώς και για να διαχωριστούν ασθενείς με Alzheimer, σχιζοφρένεια, ψυχιατρικές ασθένειες και εγκεφαλικά τραύματα από υγιείς ανθρώπους του εκάστοτε δείγματος (Ford et al., 2003).

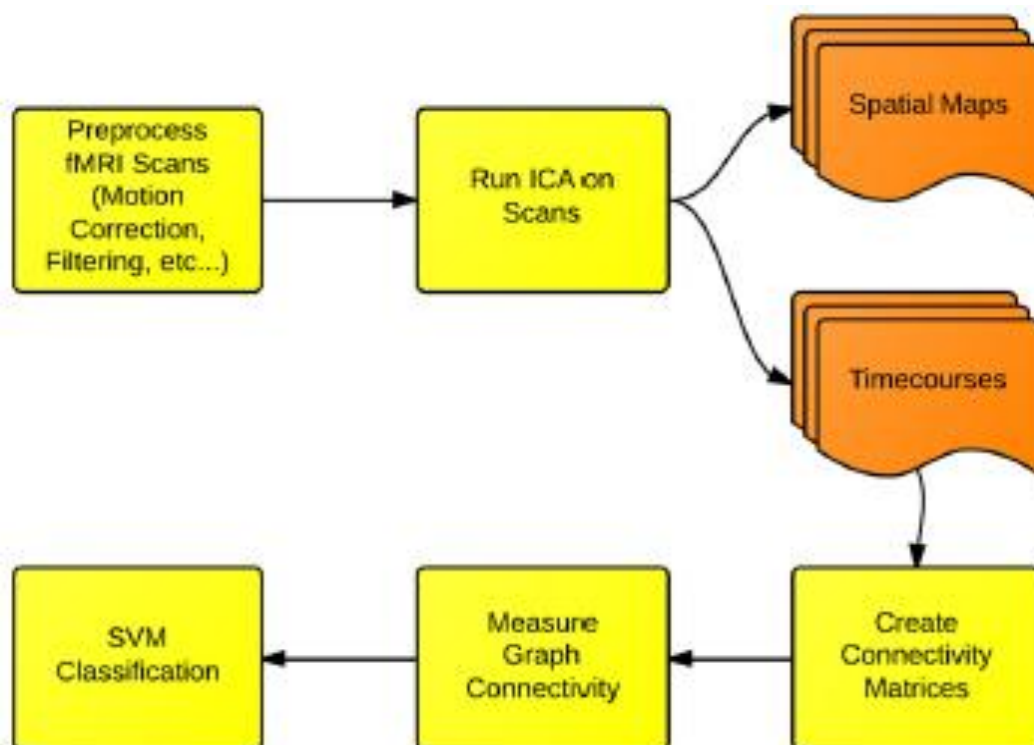
Ωστόσο αλγόριθμοι σαν τους παραπάνω κατασκευάζονται και προσαρμόζονται έτσι ώστε να ταιριάζουν σε κάθε σετ δεδομένων το οποίο καλούνται να διαχωρίσουν, επομένως είναι δύσκολο κανείς να επικυρώσει τα αποτελέσματα σε διαφορετικές ομάδες ασθενών ή ακόμα και σε ίδια ομάδα ασθενών αλλά με διαφορετικό πληθυσμό. Επιπρόσθετα, οι μελέτες αυτές αφορούν πολύ μικρά δείγματα με περίπου 20 υποκείμενα. Η επαναπαραγωγή τέτοιων αποτελεσμάτων είναι συχνά μη επιβεβαιώσιμη, αφήνοντας ανοικτό το ενδεχόμενο κριτικής σε ότι αφορά μία καλή επίδοση στην ταξινόμηση που μπορεί να οφείλεται κυρίως στο μοντέλο και όχι σε πραγματικές λειτουργικές και ανατομικές διαφορές μεταξύ των ομάδων του δείγματος. Επίσης είναι δύσκολο να χειριστεί κανείς τα δεδομένα που παράγονται στα εργαστήρια αφού οι παράμετροι σάρωσης, η ανάλυση εικόνας και η αλληλουχία των εικόνων πρέπει να είναι περίπου η ίδια. Για το λόγο αυτό, η ικανότητα αξιολόγησης των μοντέλων που χρησιμοποιούνται για την ανάλυση σε διαφορετικά σετ δεδομένων θα αύξανε την εγκυρότητα των αποτελεσμάτων, μιας και τα μοντέλα που χρησιμοποιούνται για την ανάλυση όμοιων ομάδων ασθενών θα πρέπει να παράγουν παρόμοια αποτελέσματα, κρατώντας πάντα κατά το δυνατόν σταθερές τις παραμέτρους που αναφέραμε παραπάνω.

Παρακάτω θα αναλυθούν εκτενώς τα βήματα και οι μέθοδοι που χρησιμοποιήθηκαν για την ανάλυση λειτουργικής διασυνδεσιμότητας συμπεριλαμβανομένου και της προσπάθειας ταξινόμησης των δύο ομάδων των δεδομένων μας, δηλαδή το διαχωρισμό ασθενών διαγνωσμένων με σχιζοφρένεια από υγιείς ανθρώπους του δείγματος.

### 3.3 Μέθοδοι και βήματα που έγιναν για την ανάλυση των δεδομένων μας

Στο υποκεφάλαιο αυτό της εργασίας, θα αναλυθούν επ' ακριβώς οι μέθοδοι και τα βήματα που έγιναν για την ανάλυση λειτουργικής διασυνδεσιμότητας ανάμεσα στις δύο ομάδες του δείγματος. Σκοπός μας είναι η εξαγωγή σημαντικών χαρακτηριστικών για την ανάδειξη της διαφορετικότητας και τελικά την προσπάθεια ταξινόμησης των υποκειμένων του δείγματος.

Αρχικά λοιπόν καθαρίσαμε και προ-επεξεργαστήκαμε τα δεδομένα (διόρθωση κίνησης, PCA κτλ.) και στη συνέχεια κάναμε χρήση της γραμμικής μεθόδου μείωσης διάστασης ICA (spatial) για να αναλύσουμε κάθε σάρωση σε ένα σετ από χωρικά ανεξάρτητα εγκεφαλικά δίκτυα που μεταβάλλονται στο χρόνο σύμφωνα με τις αντίστοιχες χρονοσειρές. Στη συνέχεια φτιάξαμε πίνακες λειτουργικής διασυνδεσιμότητας για κάθε υποκείμενο ξεχωριστά μετρώντας τη πολλαπλή συσχέτιση των χρονοσειρών για κάθε δίκτυο υποκειμένου. Στη συνέχεια κάθε πίνακας μετετράπη σε ένα γράφο με τη βοήθεια της μη γραμμικής μεθόδου ISOMAP για να βρεθούν οι ιδιότητες της διασυνδεσιμότητας του (Λειτουργική διασυνδεσιμότητα δικτύου). Τελικά οι ιδιότητες διασυνδεσιμότητας του γράφου χρησιμοποιήθηκαν σαν χαρακτηριστικά για έναν SVM ταξινομητή. Επιπρόσθετα κάναμε και κάποιους τ-ελέγχους για να εκτιμήσουμε αν η διασυνδεσιμότητα «μικρού-κόσμου» (small-world connectivity) για τα ICA δίκτυα είναι διαφορετική (σημαντικά διαφορετική) ανάμεσα σε ασθενείς με σχιζοφρένεια και υγιείς ανθρώπους του δείγματος. Σχηματικά, η διαδικασία που ακολουθήθηκε είναι η εξής (Εικόνα 14):



**Εικόνα 14.** Αναπαράσταση της διαδικασίας που ακολουθήθηκε για την ανάλυση των δεδομένων μας σε διάγραμμα ροής (flow chart) [Anderson and Cohen, 2013].

### 3.3.1 Προ-επεξεργασία των δεδομένων

Το πρώτο μας βήμα για την ανάλυση ήταν η προ-επεξεργασία (pre-process) των δεδομένων μας.

Αυτή η διαδικασία περιλαμβάνει αρχικά διόρθωση κίνησης η οποία οφείλεται στο γεγονός ότι το υποκείμενο πρακτικά δεν μπορεί να μένει ακίνητο σε όλο το διάστημα των σαρώσεων. Αυτό έχει σαν αποτέλεσμα τη δημιουργία μεγάλων προβλημάτων στην μετέπειτα ανάλυση μιας και υποτίθεται ότι κάθε εικονοστοιχείο (voxel) αφορά μία μοναδική περιοχή του εγκεφάλου. Στη συνέχεια θα πρέπει τα δεδομένα να καθαριστούν από ίχνη που αφήνει το κρανίο και οποιοδήποτε άλλο υλικό (π.χ. μη εγκεφαλικοί ιστοί) πέρα από τον εγκέφαλο που είναι το αντικείμενο μελέτης, αφού αυτά προσθέτουν θόρυβο περιπλέκοντας τα δεδομένα και αποπροσανατολίζοντας τελικά την ανάλυση. Ακόμα γίνεται λείανση (smoothing) των εικόνων - μία πρακτική συνηθισμένη για την προ-επεξεργασία FMRI - έτσι ώστε να ενισχυθεί η συσχέτιση των περιοχών του εγκεφάλου και αυτό προκύπτει από το γεγονός ότι γειτονικές περιοχές του εγκεφάλου παρουσιάζουν εγγενείς ομοιότητες μεταξύ τους. Με άλλα λόγια πολύ δυνατά σήματα αποδυναμώνονται την ώρα που αρκετά χαμηλά σήματα ενδυναμώνονται προκειμένου να αναδειχθεί καλύτερα η συσχέτιση μεταξύ των περιοχών του εγκεφάλου. Ένα ακόμη βήμα που χρησιμοποιήθηκε για την προ - επεξεργασία είναι το φιλτράρισμα υψηλής συχνότητας στα 100 Hz το οποίο στοχεύει στο να καθαρίσει τα δεδομένα από το θόρυβο των χαμηλών συχνοτήτων (Smith et al., 1999). Στη συνέχεια εφαρμόζεται η κλασική μέθοδος μείωσης διάστασης δεδομένων PCA η οποία μειώνει αποδοτικά τα δεδομένα και κρατάει ένα πολύ υψηλό ποσοστό επεξήγησης των δεδομένων μέσω της διασποράς (πάνω από 95%).

Για τη παραπάνω διαδικασία όπως έχει προαναφερθεί χρησιμοποιήθηκε το πακέτο FSL που είναι ένα λογισμικό ειδικευμένο για την ανάλυση δεδομένων από MRI απεικονίσεις. Χρησιμοποιώντας λοιπόν τη συνάρτηση FEAT του πακέτου σε γραφικό περιβάλλον κάναμε την απαραίτητη προ - επεξεργασία που αναφέρθηκε παραπάνω με κώδικα που ο αναγνώστης μπορεί να βρει στο Παράρτημα που βρίσκεται στο τέλος της εργασίας.

### 3.3.2 Ανάλυση ανεξάρτητων συνιστωσών για τα υποκείμενα

Μετά την προ - επεξεργασία των δεδομένων μας, πρέπει να βρούμε σημαντικά χαρακτηριστικά για τα δεδομένα προκειμένου να πετύχουμε αργότερα τόσο την ταξινόμηση τους με ένα αξιοσημείωτο ποσοστό επιτυχίας, όσο και να αναδείξουμε διαφορές στις δύο ομάδες των δεδομένων μας μέσω της ανάλυσης λειτουργικής διασυνδεσιμότητας. Αυτό θα μας βοηθήσει για δυο λόγους. Αρχικά, γιατί η εξαγωγή των σημαντικών χαρακτηριστικών θα μειώσει το θόρυβο ξεσκαρτάροντας στην ουσία σήματα τα οποία δεν έχουν κάποια σημαντική ομοιότητα με άλλα. Το γεγονός αυτό βασίζεται στην υπόθεση ότι ο θόρυβος είναι ανεξάρτητος στο εύρος των παρατηρήσεων ή με άλλα λόγια ένα σήμα που εντοπίζεται μόνο σε μία περιοχή είναι

πιθανότερο να αποτελεί θόρυβο από ένα σήμα το οποίο εντοπίζεται συνεχώς στον εγκέφαλο. Ο δεύτερος λόγος είναι ότι μειώνοντας τα δεδομένα της σάρωσης σε έναν αριθμό πιο χειροπιαστό από συνεχή σήματα, μειώνεται και η τάση της υπέρ προσαρμογής (over fitting) κατά τη διαδικασία της ταξινόμησης. Όπως είναι γνωστό η πολυπλοκότητα της ταξινόμησης είναι συνάρτηση του αριθμού των διαστάσεων-χαρακτηριστικών.

Για την εξαγωγή των σημαντικών χαρακτηριστικών χρησιμοποιήσαμε την μέθοδο ICA η οποία όπως είδαμε και στο προηγούμενο κεφάλαιο της εργασίας έχει αποκτήσει ιδιαίτερη φήμη για την εφαρμογή της σε δεδομένα FMRI. Αυτή η μέθοδος, μπορεί να απομονώσει δίκτυα ανάλογα με την δραστηριότητα των νευρώνων, γεγονός που έχει επικυρωθεί με την ανίχνευση χαρακτηριστικών που υπάρχουν στο εύρος των υποκειμένων και των σαρώσεων που αντιστοιχούν σε λειτουργικά ανιχνεύσιμα εγκεφαλικά δίκτυα (McKeown et al., 2003; Anderson et al., 2011).

Στην εφαρμογή μας στα πλαίσια της παρούσας εργασίας, θα γίνει χρήση της μεθόδου ICA σε κάθε υποκείμενο ξεχωριστά (ICA within subject) και δεν θα πραγματοποιηθεί η μέθοδος ομαδικής ICA (group ICA) που θα μας έδινε κοινά δίκτυα στο εύρος όλων των υποκειμένων. Αυτό βασίζεται στην υπόθεση ότι υπάρχουν διαφορετικά σετ από εγκεφαλικά δίκτυα στην σχιζοφρένεια και υποθέτοντας ότι τα ίδια ακριβώς δίκτυα λειτουργούν και στις δύο ομάδες του δείγματος θα αποδυνάμωνα πολύ τις μεταξύ τους διαφορές - που είναι ακριβώς αυτό που εμείς θέλουμε να αναδείξουμε

Υπό την υπόθεση ότι η δραστηριότητα του εγκεφάλου είναι κατασκευασμένη σε ανατομικά δίκτυα που αλληλοεπιδρούν μαζί για να παράγουν σημαντικές ψυχοσυμπεριφορικές καταστάσεις, η συνολική δραστηριότητα αναλύεται σε υποσυνιστώσες με τη βοήθεια της ICA. Εκτός από αυτό, ο χώρος «ξετυλίγεται», εκεί όπου οι 4-διαστάσεων σάρωση (3 διαστάσεις χώρου και 1 χρόνου) μετασχηματίζεται σε έναν πίνακα με διαστάσεις χώρο επί το χρόνο, έτσι ώστε μία σάρωση διάστασης  $(X, Y, Z, T)$  να γίνει ένας πίνακας διάστασης  $(T \text{ (χρόνος)}, X \times Y \times Z \text{ (χώρος)})$ . Έτσι μία σάρωση σε χρόνο  $T$  και χωρικής διάστασης  $S$  μπορεί να εκφραστεί σαν γραμμικός συνδυασμός από  $M < T$  συνιστώσες και αντίστοιχες χρονοσειρές :

$$X_{ts} = \sum_{\mu=1}^M A_{t\mu} C_{\mu s} \quad (20)$$

Όπου με  $X_{ts}$  συμβολίζεται η ένταση της σάρωσης τη χρονική στιγμή  $t < T$  και χωρικής τοποθεσίας  $s < S$ ,  $A_{t\mu}$  είναι το εύρος της συνιστώσας  $\mu$  σε χρόνο  $t$ , και  $C_{\mu s}$  είναι το χωρικό μέγεθος για τη συνιστώσα  $\mu$  στη χωρική τοποθεσία  $s$ .

Οι συνιστώσες  $C$  εκτιμώνται να είναι όσο το δυνατόν στατιστικά (χωρικά) ανεξάρτητες λύνοντας ουσιαστικά το αντίστροφο πρόβλημα -στην περίπτωση μας- με τον αλγόριθμο FAST-ICA. Για την εκτίμηση των ανεξάρτητων πηγών  $c$  στο γενικευμένο πρόβλημα  $x = Ac$  λύνεται το αντίστροφο πρόβλημα  $y = w'x$  όπου  $w$  είναι μία γραμμή του  $A^{-1}$  (αντίστροφος του πίνακα μίξης). Στη συνέχεια,

$$y = w'x \Rightarrow y = w'Ac \text{ και αντικαθιστώντας το } z = A'w \text{ έχουμε } y = z'c.$$

Το  $w$  βελτιστοποιείται έτσι ώστε  $y = w'x = z'c$  να είναι όσο το δυνατόν περισσότερο μη Γκαουσιανή πηγή, πράγμα που συνεπάγεται και οδηγεί σε ακόμα λιγότερο Γκαουσιανές πηγές  $c$  λόγω του Κεντρικού Οριακού Θεωρήματος. Μεγιστοποίηση της κύρτωσης, ελαχιστοποίηση της εντροπίας και μεγιστοποίηση της

νεγεντροπίας για το  $w$  είναι όλες μέθοδοι για να βρει κάποιος τις λιγότερο Γκαουσιανές  $y = \hat{w} x = \hat{z} c$ . Ωστόσο, εμείς θα ασχοληθούμε με την νεγεντροπία, την οποία υπολογίζει η FAST-ICA ώστε να αναλύσει τα δεδομένα σε ανεξάρτητες συνιστώσες. Έχουμε λοιπόν :

$$\text{Nentropy} = J(y) = H(y_{Gauss}) - H(y) \quad (21)$$

Όπου,

$$H(y) = -\sum_i P(y = a_i) \log[P(y = a_i)] \quad (22)$$

Ενώ για τη συνεχή περίπτωση η σχέση γίνεται,

$$H(y) = -\int f(y) \log(f(y)) dy \quad (23)$$

Η προεπιλογή παραμέτρου στην συνάρτηση MELODIC (για το ICA) που χρησιμοποιήσαμε στο στατιστικό πακέτο FSL, ήταν η παράλληλη εξαγωγή ενώ επίσης περιλάμβανε χρονική κανονικοποίηση και διεξαγωγή 1000 το πολύ επαναλήψεων για τον αλγόριθμο FAST-ICA χρησιμοποιώντας την νεγεντροπία :

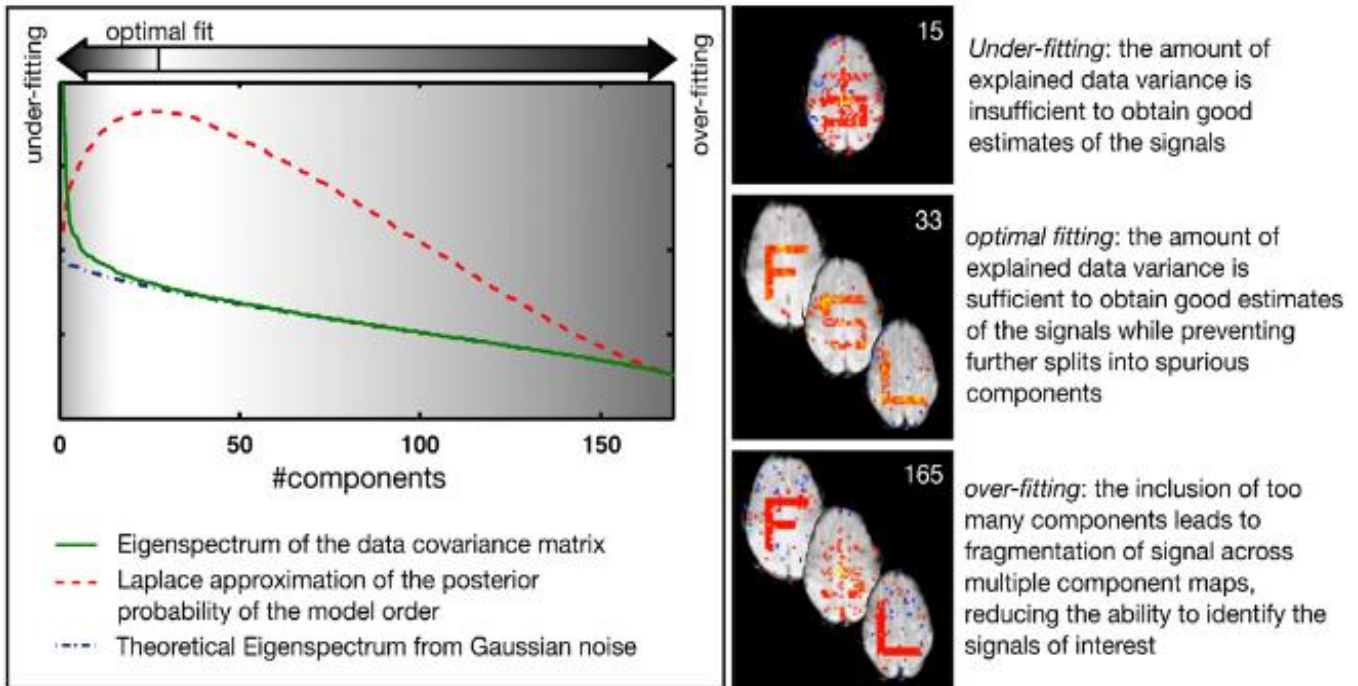
$$G(u) = \frac{1}{a} \log[\cosh(au)] \quad (25)$$

Όπου,  $a \in [1,2]$  και είναι η σταθερά που χρησιμοποιείται για να εκτιμηθεί η νεγεντροπία.

Αξίζει επίσης να σημειωθεί ότι σαν προεπιλογή η συνάρτηση MELODIC για το ICA κάνει έναν αυτόματο υπολογισμό για τον αριθμό των ανεξάρτητων συνιστωσών τις οποίες και θα αναζητήσει στη συνέχεια. Το θέμα αυτό είναι αρκετά επίμαχο όπως είδαμε και στο προηγούμενο κεφάλαιο. Πολλές φορές αυτός ο αριθμός μπορεί να επηρεάσει τα συνολικά αποτελέσματα της μεθόδου (Ma et al., 2007). Ως επακόλουθο καταλαβαίνουμε ότι κάθε υποκείμενο θα έχει διαφορετικό αριθμό ανεξάρτητων συνιστωσών για τη μετέπειτα ανάλυση. Το γεγονός αυτό είναι λογικό αν σκεφτεί κανείς ότι τα δεδομένα είναι θορυβώδη και ανάμεσα στα υποκείμενα υπάρχουν ανατομικές διαφορές κτλ.

Ο τρόπος με τον οποίο η συνάρτηση MELODIC του FSL κάνει αυτόν τον αυτόματο υπολογισμό για τον αριθμό των ανεξάρτητων συνιστωσών που πρόκειται να υπολογίσει, είναι αρχικά η εύρεση των ιδιοτιμών του πίνακα διασποράς για το κάθε υποκείμενο. Ανάλογα με το μέγεθος των ιδιοτιμών η επεξήγηση των δεδομένων είναι μεγαλύτερη ή μικρότερη. Εδώ όμως θα πρέπει να προσέξουμε ότι δεν μας συμφέρει να πάρουμε ούτε υπερβολικά μεγάλο αριθμό ανεξάρτητων συνιστωσών (over-fitting) αλλά ούτε και υπερβολικά μικρό (under-fitting). Συγκεκριμένα, η συνάρτηση MELODIC συγκρίνει το φάσμα των ιδιοτιμών του πίνακα διασποράς των δεδομένων, το φάσμα ιδιοτιμών λόγω του Γκαουσιανού θορύβου και χρησιμοποιεί την προσέγγιση Laplace για να υπολογίσει την δεσμευμένη πιθανότητα (posterior) της τάξης του μοντέλου σαν τη διάσταση με τη μεγαλύτερη πιθανότητα (Beckman and Smith, 2004). Όταν αυτή μεγιστοποιηθεί βρίσκει τον αριθμό των συνιστωσών που πρόκειται να υπολογίσει (εικόνα 15).

Τελικά, για την ανάλυση ICA των δεδομένων μας χρησιμοποιήσαμε χωρική (spatial) ICA (που είναι και η μοναδική επιλογή της συνάρτησης του στατιστικού πακέτου FSL) που στην ουσία ψάχνει στατιστική ανεξαρτησία για τους χωρικούς χάρτες (spatial maps). Εναλλακτικά, θα μπορούσαμε να εφαρμόσουμε χρονική (temporal) ICA γεγονός που θα σήμαινε τη βελτιστοποίηση στην χρονική ανεξαρτησία των συνιστωσών.



**Εικόνα 15.** Επιλογή του αριθμού ανεξάρτητων συνιστωσών στο πακέτο FSL. Το παρατηρούμενο φάσμα ιδιοτιμών από τον πίνακα διασποράς (πράσινο) συγκρίνεται με το αναμενόμενο φάσμα ιδιοτιμών λόγω Γκαουσιανού θορύβου (μπλε). Χρησιμοποιώντας την προσέγγιση Laplace για τη δεσμευμένη πιθανότητα (posterior) της τάξης του μοντέλου, μπορούμε να υπολογίσουμε την τάξη αυτή σαν τη διάσταση με τη μεγαλύτερη πιθανότητα. Από το σημείο μεγιστοποίησης και έπειτα το φάσμα των ιδιοτιμών δεν είναι διαφορετικό από την εκτίμηση του φάσματος του Γκαουσιανού θορύβου. Ρίχνοντας επιπλέον τις διαστάσεις κάτω από το μέγιστη τιμή της προσέγγισης Laplace οδηγεί σε μία υποτίμηση σημάτων (underfitting) που θα μας έδιναν χρήσιμες πληροφορίες ενώ το αντίθετο θα οδηγούσε σε υπέρ – προσαρμογή (overfitting) όπου κάποια σήματα διασπώνται και εμφανίζονται ως θόρυβος σε πολλούς χωρικούς χάρτες για το ίδιο υποκείμενο [Woolrich et al., 2008].

### 3.3.3 Κατασκευή πινάκων λειτουργικής διασυνδεσιμότητας

Για την κατασκευή πινάκων λειτουργικής διασυνδεσιμότητας (ένας πίνακας για κάθε υποκείμενο) χρησιμοποιήθηκε μία μετρική που βασίζεται στη χρονική συσχέτιση χρονοσειρών που αντιστοιχούν σε χωρικούς χάρτες του εκάστοτε υποκειμένου. Η συνάρτηση απόστασης  $d(A_{\mu_i}, A_{\mu_j})$  που θα παρουσιαστεί παρακάτω είναι ένας

μετασχηματισμός από μέγιστες - σε απόλυτη τιμή πολλαπλές συσχετίσεις (για όλα τα lags) - μεταξύ δύο χρονοσειρών. Ο υπολογισμός αυτός γίνεται για κάθε πιθανό ζευγάρι χρονοσειρών ενός υποκειμένου και έτσι μετασχηματίζουμε τα δεδομένα FMRI σε έναν πίνακα. Αυτό είναι ένα μέτρο για τη λειτουργική διασυνδεσιμότητα ανάμεσα σε δύο συνιστώσες αλλά είναι απλώς ένα μέτρο από τα πολλά πιθανά που θα μπορούσαμε να χρησιμοποιήσουμε (π.χ. με τη βοήθεια της συνοχής (coherence) ).

Η συνάρτηση πολλαπλής συσχέτισης (Cross correlation function) μεταξύ των χρονοσειρών υπολογίζεται στο εύρος των χρονικών καθυστερήσεων (temporal lags) όπως αυτή αναλύθηκε στο 2<sup>ο</sup> κεφάλαιο. Αφαιρούμε λοιπόν τη μέγιστη απόλυτη τιμή της πολλαπλής συσχέτισης από τη μονάδα για να φτιάξουμε τη μετρική,  $d(A_{\mu_i}, A_{\mu_j})$  που δίνεται από τη σχέση :

$$d(A_{\mu_i}, A_{\mu_j}) = 1 - \max[ |CCF(A_{\mu_i}, A_{\mu_j}, l)| ] \quad (26)$$

Η συνάρτηση πολλαπλής συσχέτισης CCF δίνεται από τη σχέση (27) :

$$CCF(A_{\mu_i}, A_{\mu_j}, l) = \frac{E[(a_{\mu_i, t+l} - \bar{A}_{\mu_i})(a_{\mu_j, t} - \bar{A}_{\mu_j})]}{\sqrt{E[(a_{\mu_i, t} - \bar{A}_{\mu_i})^2] E[(a_{\mu_j, t} - \bar{A}_{\mu_j})^2]}} \quad (27)$$

Όπου  $l$  είναι η χρονική καθυστέρηση που χωρίζει δύο χρονοσειρές,  $A_{\mu_i}$  και  $A_{\mu_j}$ , ο δείκτης  $\mu$  αναφέρεται στην ανεξάρτητη συνιστώσα (χωρικός χάρτης που αντιστοιχεί η χρονοσειρά) με την οποία δουλεύουμε, και  $\bar{A}_{\mu_i}$  είναι ο μέσος όρος όλων των χρονοσειρών  $A_{\mu_i} = (a_{\mu_i, 1}, a_{\mu_i, 2}, \dots, a_{\mu_i, T})$ . Οι χρονοσειρές υπολογίζονται σε χρονικές καθυστερήσεις από 0 μέχρι 3 χρονικά σημεία που αντιστοιχούν σε περίπου 6 δευτερόλεπτα (για τα δεδομένα μας), αφού μεγαλύτερες χρονικές καθυστερήσεις έχουν ως αποτέλεσμα λιγότερα χρονικά σημεία να υπολογίζονται τη συσχέτιση δίνοντας μία πιο θορυβώδη εκτίμηση ενώ επίσης αυτό δεν θα ήταν λογικό και από μία βιολογική άποψη για τις νευρολογικές συζεύξεις.

Οι πίνακες αυτοί είναι δύσκολο να ερμηνευθούν αφού αποτελούν απλώς αναπαραστάσεις από ένα σετ συνδεδεμένων αντικειμένων. Ακόμα, οι γραμμές και οι στήλες των πινάκων αυτών αναπαριστούν μοναδικές ανεξάρτητες συνιστώσες ανάμεσα στα υποκείμενα, οι οποίες δεν είναι συγκρίσιμες βέβαια στο εύρος των υποκειμένων. Ο πρωταρχικός μας στόχος είναι να μετρήσουμε διασυνδεσιμότητα, όχι μόνο το πόσο στενά συνδεδεμένα είναι τα αντικείμενα αλλά επίσης το πως αυτή η διασυνδεσιμότητα αλλάζει αντίστοιχα με την ομάδα που μελετάμε (σχιζοφρενείς-υγιείς). Για παράδειγμα, αλληλεπιδρά κάθε δίκτυο με όλα τα υπόλοιπα; Υπάρχουν υπογράφοι που είναι αποκομμένοι από τον αρχικό γράφο; Διαφέρει ο αριθμός των βημάτων για να πάμε από έναν κόμβο σε έναν άλλο; Είναι μερικά γραφήματα πιο πυκνά από άλλα; Για να απαντηθούν αυτές οι ερωτήσεις θα πρέπει να μετατρέψουμε τους πίνακες διασυνδεσιμότητας σε γράφους και μετά να συνεχίσουμε με ανάλυση διασυνδεσιμότητας δικτύου-γράφοι.



### 3.3.4 Κατασκευή γράφου με τη μέθοδο ISOMAP

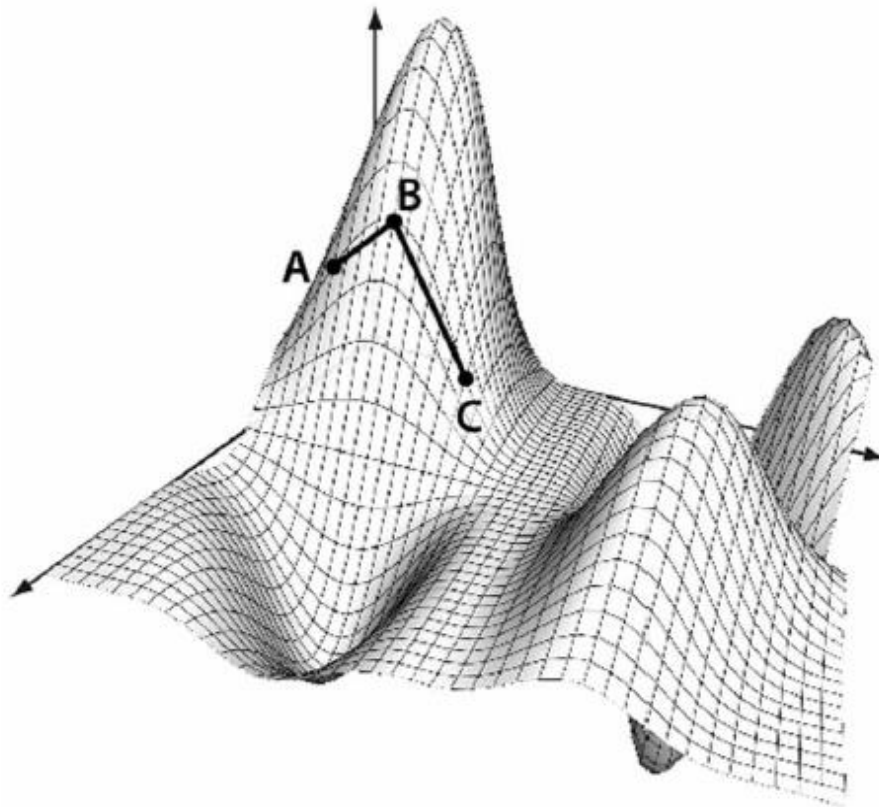
Αφού σχηματίσαμε τους πίνακες λειτουργικής διασυνδεσιμότητας για κάθε υποκείμενο προχωρήσαμε στην κατασκευή γράφου με τη βοήθεια του κάθε πίνακα.

Κάθε πίνακας είναι μία δομή από συνδεδεμένα σημεία σε μια υψηλής διάστασης πολλαπλότητα (manifold), όπου κάθε σημείο συμβολίζει ουσιαστικά μία ανεξάρτητη χρονοσειρά που αντιστοιχεί σε έναν ανεξάρτητο χωρικό χάρτη και η απόσταση ανάμεσα σε 2 σημεία μετρά την ομοιότητα όσον αφορά το συγχρονισμό των δύο χρονοσειρών. Κάθε σημείο συνδέεται με ένα άλλο ανεξάρτητα την ομοιότητα. Για να κατασκευάσουμε γράφους από τους πίνακες λειτουργικής διασυνδεσιμότητας, ξεσκαρτάρουμε τις αδύναμες συνδέσεις στο εύρος των σημείων και μετά προσαρμόζουμε την απλούστερη δομή σε έναν χαμηλότερης διάστασης χώρο χρησιμοποιώντας τη μη γραμμική μέθοδο μείωσης διάστασης ISOMAP που αναλύσαμε στο κεφάλαιο 2. Στη συνέχεια υπολογίζουμε τα γραφοθεωρητικά στοιχεία διασυνδεσιμότητας γράφου για να εντοπίσουμε τις συνδέσεις μεταξύ των εγκεφαλικών δικτύων ήρεμης κατάστασης (resting state networks).

Εννοιολογικά, κάθε σεντ σημείων που περιέχεται στον πίνακα αποστάσεων διάστασης  $d$  που υπολογίσαμε, μπορεί να προσαρμοστεί σε ένα χώρο διάστασης  $d-1$  χωρίς να έχουμε απώλεια πληροφορίας (πρακτικά όλες οι αποστάσεις μπορούν να διατηρηθούν). Συνήθως ένας τέτοιος μετασχηματισμός υποθέτει ότι ο χώρος στον οποίο βρίσκονται τα δεδομένα είναι γραμμικός. Ωστόσο μία τέτοια υπόθεση μπορεί να είναι λανθασμένη. Για παράδειγμα, αν προσπαθούσαμε να μετρήσουμε την απόσταση από το Σακραμέντο στην Αμερική, μέχρι την Σανγκάη στην Κίνα - χρησιμοποιώντας τις συντεταγμένες της κάθε πόλης στο τρις ορθογώνιο σύστημα αξόνων  $(x,y,z)$  - η γραμμική απόσταση μεταξύ των πόλεων ενώ θα ήταν υπολογίσιμη θα θεωρούσε ότι το σωστό και συντομότερο μονοπάτι περνάει από τον πυρήνα της γης, κάτι που είναι πρακτικά αδύνατο. Αντί αυτού λοιπόν, ένας πιο ορθολογικός τρόπος να μετρήσουμε την απόσταση θα ήταν μέσω διασυνδετικών σημείων όπου θα μπορούσαμε να πετάξουμε με αεροπλάνο και να φτάσουμε π.χ. από το Σακραμέντο στο Λος Άντζελες, από το Λος Άντζελες στο Τόκιο και τελικά από το Τόκιο στη Σανγκάη. Αυτή η μέθοδος μέτρησης της απόστασης είναι γνωστή ως γεωδαισιακή απόσταση ή απόσταση μεταξύ συνδετικών σημείων. Υποθέτει δηλαδή, ότι το ταξίδι ανάμεσα σε μακρινά σημεία, συχνά απαιτεί πορεία μέσω ενδιάμεσων κόμβων όπως φαίνεται παρακάτω στην εικόνα 16 .

Αυτή λοιπόν την προσέγγιση θα χρησιμοποιήσουμε κι εμείς για να «εξαφανίσουμε» τις πολύ αδύναμες συνδέσεις, αντικαθιστώντας αυτές με διασυνδετικά μονοπάτια δημιουργώντας γράφους από τους πίνακες διασυνδεσιμότητας που κατασκευάσαμε. Μετατρέπουμε λοιπόν κάθε πίνακα σε μία δομή γράφου χρησιμοποιώντας αρχικά μία από τις δύο βασικές προσεγγίσεις για την εφαρμογή του ISOMAP (Tenenbaum et al., 2000). Οι προσεγγίσεις αυτές έχουν να κάνουν - όπως προαναφέρθηκε και στο δεύτερο κεφάλαιο - με τη συνθήκη γειτνίασης μεταξύ των κόμβων μας. Είτε ορίζουμε μία ακτίνα  $\epsilon$ , βάση της οποίας κάθε σημείο που βρίσκεται σε απόσταση ίση με την ακτίνα αυτή θεωρείται ότι γειτνιάζει με το σημείο αυτό, είτε εφαρμόζουμε την μέθοδο κοντινότερων γειτόνων ( $k$ -nearest neighbours,  $k$ -NN) που αναλύσαμε στο κεφάλαιο 2. Με τον τρόπο αυτό καταφέρνουμε να ξεσκαρτάρουμε τις πολύ αδύναμες συνδέσεις αντικαθιστώντας τις αποστάσεις αυτές

(μεταξύ μη συνδεδεμένων κόμβων) με την ελάχιστη απόσταση μεταξύ συνδεδεμένων κόμβων. Οι ελάχιστες αποστάσεις από κάθε σημείο σε οποιοδήποτε άλλο σημείο του γράφου υπολογίζονται μέσω αλγορίθμων όπως του Dijkstra ή Floyd - Warshal κτλ. Συνεχίζοντας την εφαρμογή του ISOMAP σαν τελευταίο βήμα έχουμε την εφαρμογή της γραμμικής μεθόδου MDS που ψάχνει μία βέλτιστη προσαρμογή σε χάρτη χαμηλότερης διάστασης, των συνδεδεμένων κόμβων - ανεξάρτητων συνιστωσών του κάθε υποκειμένου.



**Εικόνα 16.** Η απόσταση ανάμεσα στο A και το C υπολογίζεται σαν το μονοπάτι από το A στο B και τελικά στο C, αντί της γραμμικής απόστασης απευθείας από το A στο C. Το γεγονός αυτό παραβιάζει την υπόθεση ότι τα σημεία βρίσκονται σε ένα γραμμικό χώρο χρησιμοποιώντας ευκλείδεια απόσταση [Anderson and Cohen, 2013].

Πρέπει να σημειωθεί ότι οι δύο ορισμοί για τη διασυνδεσιμότητα που αναφέραμε πριν, ανάμεσα στην ακτίνα  $\epsilon$  και την εφαρμογή της μεθόδου  $k$ -κοντινότερων γειτόνων μπορούν να οδηγήσουν σε διαφορετικά αποτελέσματα. Ορίζοντας διασυνδεσιμότητα μεταξύ των κόμβων χρησιμοποιώντας την  $\epsilon$ -απόσταση μπορεί να οδηγηθούμε σε έναν κατακερματισμένο γράφο, με περιοχές που δεν συνδέονται με κανένα άλλο (άμεσα ή έμμεσα) υπογράφο. Κάτι τέτοιο θα μπορούσε να προκληθεί από κάποιο σημείο (ή σημεία) που είναι τόσο μακριά από άλλα, ώστε να μην επιτυγχάνεται μία σύνδεση με τον κυρίως γράφο. Αυτό, είναι ένα παράδειγμα όπου μία απριόρι γνώση για την ασθένεια αυτή καθ' αυτή μπορεί να μας πληροφορεί για την επιλογή παραμέτρων που σχετίζονται με τις μεθόδους που πρόκειται να

χρησιμοποιηθούν. Σε ασθένειες όπως η σχιζοφρένεια ή ο αυτισμός, η υπόθεση της από-διασυνδεσιμότητας (disconnectivity) μπορεί να δοκιμαστεί άμεσα με υπολογισμό σε ένα στατιστικό πακέτο, για παράδειγμα, το ποσοστό κατακερματισμού των γράφων ανάμεσα στους ασθενείς και τους υγιείς.

Στην περίπτωση μας χρησιμοποιήσαμε τη μέθοδο  $k$ -NN με κάποια παραλλαγή, ως συνθήκη γειτνίασης - διασυνδεσιμότητας. Επιλέγουμε λοιπόν τους  $k$  πιο κοντινούς γείτονες όλων των κόμβων για να συνδεθούν μεταξύ τους, ορίζοντας τιμή του  $k$  το 10% των συνολικών κόμβων που έχει ο γράφος, ή εναλλακτικά  $k = 2$ , όποια από τις δύο τιμές αυτές είναι μεγαλύτερη. Αυτό, ωθεί το γράφο να είναι απολύτως συνδεδεμένος, αντίθετα με τη προσέγγιση της  $\varepsilon$ -απόστασης. Επιλέγουμε αυτήν την επιλογή παραμέτρου γιατί στη περίπτωσή μας έχουμε να κάνουμε με σταθμισμένα (weighted) γραφήματα. Οι προσεγγίσεις  $\varepsilon$ -απόστασης τυπικά μετατρέπουν τον πίνακα γειτνίασης σε διωνυμικό αναθέτοντας βάρη σε δύο κόμβους που έχουν απόσταση πάνω από μία οριακή τιμή  $\varepsilon$  (threshold) και μηδενίζοντας τα βάρη όταν η απόσταση είναι μικρότερη από αυτήν (**Rubinov and Sporns, 2010**). Από τη στιγμή που χρησιμοποιούμε σταθμισμένα γραφήματα με την επιλογή της παραλλαγής αυτής της  $k$ -NN, επιτρέπουμε στα επαρκώς κοντινά σημεία να διατηρήσουν τα βάρη που τους έχουν δοθεί (αποστάσεις δηλαδή, από τους πίνακες λειτουργικής διασυνδεσιμότητας), ενώ κόβουμε τη σύνδεση εκείνων των σημείων που δεν είναι γειτονικά σύμφωνα με τη μέθοδο.

Όσον αφορά το υπολογιστικό κομμάτι στο σημείο αυτό χρησιμοποιήθηκε το πακέτο `vegan` της γλώσσας R για την εφαρμογή της μεθόδου ISOMAP (Παράρτημα).

### 3.3.5 Μέτρηση γραφοθεωρητικών στοιχείων

Στο σημείο αυτό, κάθε εγκεφαλική σάρωση κάθε υποκειμένου μετασχηματίστηκε σε μία γραφική δομή, όπου κάθε κόμβος αναπαριστά ένα εγκεφαλικό δίκτυο και η διασυνδεσιμότητα μεταξύ των κόμβων αναπαριστά την ομοιότητα στην δραστηριότητα των δικτύων. Κάθε γράφος μπορεί τώρα να συνοψιστεί από τις ιδιότητες που αφορούν την διασυνδεσιμότητα του. Υπάρχουν πολλά γραφοθεωρητικά μέτρα που εξηγούν τα χαρακτηριστικά και τη διασυνδεσιμότητα ενός γράφου στη γλώσσα R. Παρακάτω θα δούμε κάποια από αυτά και συγκεκριμένα αυτά που υπολογίσαμε για τα γραφήματα των υποκειμένων μας.

Όπως είδαμε και παραπάνω φτιάχνοντας γράφους από κάθε πίνακα χρησιμοποιώντας μη γραμμική μετρική-απόσταση (γεωδαισιακή απόσταση), όχι μόνο επιτρέπει μία πιο αποδοτική χαμηλής διάστασης προβολή του πίνακα, αλλά ακόμα βοηθάει το γράφημα να συνδεθεί πιο αποδοτικά διακόπτοντας συνδέσεις κόμβων με χαμηλή διασυνδεσιμότητα και διατηρώντας εκείνες με την υψηλή. Το γεγονός αυτό λοιπόν μας βοηθάει να αποφασίσουμε πόσες είναι οι δυνατές συνδέσεις ανάμεσα σε κόμβους ενός υποκειμένου, πόσα υποδίκτυα (υπογράφοι) υπάρχουν, ποιο είναι το μέγεθος αυτών των υποδικτύων καθώς και πόσο αποδοτικά συνδέονται οι κόμβοι στο σύνολο (overall). Οι ιδιότητες αυτές των γραφημάτων, που είναι όλες συγγενικές μεταξύ τους, δίνουν ποσοτικές μετρήσεις για τη διασυνδεσιμότητα και μπορούν να χρησιμοποιηθούν για να αναδειχθούν διαφορές ανάμεσα σε εγκεφαλικά δίκτυα που συνδέονται με διάφορες ασθένειες - ανωμαλίες. Τα παρακάτω ατομικά μέτρα για κάθε

γράφημα μπορούν να συγκριθούν άμεσα μεταξύ των ομάδων (σχιζοφρενείς-υγιείς) αν οι πολλαπλές συγκρίσεις προσαρμοστούν κατάλληλα.

Μερικά από τα χαρακτηριστικά που μπορούμε να υπολογίσουμε για την ανάλυση λειτουργικής διασυνδεσιμότητας στα γραφήματα μας και περιέχονται στο πακέτο `igraph` της R (το οποίο και χρησιμοποιήσαμε) είναι τα εξής :

- Μέσο μήκος μονοπατιού (Average path length) : Είναι ο υπολογισμός του μέσου μήκους των ελαχίστων διαδρομών από έναν κόμβο σε έναν άλλο.
- Αριθμός κλίκα (Clique number) : Υπολογισμός μεγέθους της μεγαλύτερης κλίκα στο γράφο. Η κλίκα, στην περιοχή της θεωρίας γραφημάτων, είναι ένα υποσύνολο από κόμβους σε ένα μη-κατευθυνόμενο (undirected) γράφημα με την ιδιότητα ότι κάθε δύο διακριτοί κόμβοι γειτνιάζουν (adjacent).
- Πυκνότητα γράφου (Graph density) : Είναι αποτέλεσμα του λόγου των κόμβων με τον αριθμό των πιθανών ακμών (αριθμός κόμβων διαιρούμενος με τον αριθμό όλων των δυνατών ακμών στο γράφημα) .
- Διασυνδεσιμότητα ακμών (Edge connectivity) : Ο ελάχιστος αριθμός των ακμών που χρειάζεται να αφαιρέσουμε από τον εν λόγω γράφο, για να έχουμε αποσυνδεδεμένο γράφο (disconnected graph)
- Διάμεσος εγγύτητας (Median closeness) : Η διάμεσος των βημάτων για την προσπέλαση κάθε κόμβου δοθέντος ενός κόμβου.
- Διάμεσος βαθμού (Median degree) : Η διάμεσος του αριθμού των ακμών που οδηγούν σε ένα κόμβο, με τους βρόχους να μετρώνται διπλά.
- Αριθμός των κόμβων (Vertex count) : Αριθμός των κόμβων σε ένα γράφο.
- Αριθμός των ακμών (Edges count) : Αριθμός των ακμών σε ένα γράφο.
- Μέγιστος βαθμός (Maximum Degree) : μέγιστος αριθμός ακμών που οδηγούν σε ένα κόμβο με τους βρόχους να μετρώνται διπλά.
- Μεταβατικότητα (transitivity) : πιθανότητα δύο κόμβοι να συνδέονται μεταξύ τους. Αυτό το μέτρο ονομάζεται επίσης συντελεστής ομαδοποίησης (clustering coefficient).

Από τα παραπάνω μέτρα χρησιμοποιούμε δυο από αυτά για να υπολογίσουμε την ιδιότητα «μικρού-κόσμου» (small world) που παράχθηκε από το παιχνίδι Erdős-Renyi (**Erdős and Renyi, 1961**). Εναλλακτικές μέθοδοι υπολογισμού της ιδιότητας αυτής προτάθηκαν και πρόσφατα από τους **Rubinon και Sporns (2010)**. Η ιδιότητα small-world υπολογίζεται από τη σχέση :

$$\sigma = \frac{\gamma}{\lambda} \quad (28)$$

Όπου  $\gamma$  είναι ο λόγος του συντελεστή ομαδοποίησης του πραγματικού δικτύου με τη μέση τιμή του συντελεστή ομαδοποίησης από  $n$  τυχαία δίκτυα με ισοδύναμο αριθμό ακμών και βαρών όπως το πραγματικό (εξαγόμενο από τα δεδομένα) δίκτυο αλλά τυχαίο συνδεδεμένο (randomly rewired). Το  $\lambda$  ορίζεται όμοια με το  $\gamma$  αλλά χρησιμοποιεί το μέσο μήκος μονοπατιού. Η μεταβλητή  $n$  που προαναφέραμε βρίσκεται περίπου ανάμεσα στο 500 και το 5000.

Τυπικά, σε βιολογικά δίκτυα έχουμε :

- ✓  $\gamma \gg 1$ , για παράδειγμα, μεγαλύτερη τοπική ομαδοποίηση από ότι σε ένα τυχαίο γράφο
- ✓  $\lambda \approx 1$ , για παράδειγμα, παρόμοιο χαρακτηριστικό μήκος μονοπατιού (διάμεσος των μέσων των ελαχίστων μονοπατιών κάθε κόμβου σε οποιονδήποτε άλλο κόμβο, characteristic path length) με ένα τυχαίο δίκτυο

Κάθε δίκτυο με  $\sigma \gg 1$  θεωρείται ότι είναι «μικρός κόσμος» («small world») (Humphries and Gurney, 2008)

### 3.3.6 Ταξινόμηση χρησιμοποιώντας SVM ταξινομητή

Αφού μετασχηματίστηκε κάθε fMRI σάρωση για το κάθε υποκείμενο σε γράφημα -όπου οι κόμβοι του αναπαριστούν λειτουργικά δίκτυα και οι αποστάσεις μεταξύ τους αναπαριστούν την ομοιότητα στη δραστηριότητα των κόμβων αυτών-, αναλύθηκε και η λειτουργική διασυνδεσιμότητα των γραφημάτων (Functional Network Connectivity). Στο σημείο αυτό χρησιμοποιήσαμε αυτά τα χαρακτηριστικά διασυνδεσιμότητας γραφήματος που προαναφέραμε, για την ταξινόμηση των δύο ομάδων (από σχιζοφρενείς και υγιείς). Η γλώσσα R έχει έναν τεράστιο αριθμό από βιβλιοθήκες που είναι διαθέσιμες για ταξινόμηση. Τελικά, στην παρούσα εργασία χρησιμοποιήθηκε ένας απλός SVM (Support Vector Machine) ταξινομητής για τον οποίο χρειάστηκε το πακέτο e1071 της γλώσσας R.

Ο αλγόριθμος του SVM προσπαθεί να βρει ένα υπερεπίπεδο (hyperplane) που διαχωρίζει με βέλτιστο τρόπο διαφορετικές ομάδες, χρησιμοποιώντας μόνο σημεία που περιέχονται στο περιθώριο (margin) ή περιοχή επικάλυψης (region of overlap). Για ένα σετ σημείων  $(x_i, y_i)$  όπου  $(x_i) \in \mathbb{R}^n$  είναι το σετ των γραφοθεωρητικών στοιχείων για το γράφο  $G_i$  που αντιστοιχεί στο υποκείμενο  $i$ , ένα μέλος της κλάσης  $y_i \in (-1,1)$  (π.χ. ασθενής ή υγιής), ο SVM θα προσαρμόσει ένα υπερεπίπεδο έτσι ώστε να διαιρεί βέλτιστα τις κλάσεις στο  $(-1,1)$ . Αν ένα υπερεπίπεδο περιγράφεται σαν  $w * x_i - b$  όπου  $w$  είναι τα κανονικά διανύσματα του υπερεπιπέδου τότε παράλληλα υπερεπίπεδα που διαχωρίζουν τις παρατηρήσεις μπορούν να ορισθούν σαν  $w * x_i - b \geq -1$  για  $y_i = 1$  και  $w * x_i - b \leq -1$  για  $y_i = -1$ . Πρόβλημα βελτιστοποίησης γίνεται τώρα η μεγιστοποίηση της απόστασης ανάμεσα στα επίπεδα,  $\frac{2}{\|w\|}$ , τέτοια ώστε  $y_i(w * x_i - b) \geq -1$ .

Χρησιμοποιώντας λοιπόν της ιδιότητες των γράφων (μέσο μήκος μονοπατιού, αριθμό κλικών κτλ.) σαν χαρακτηριστικά, μπορούμε να διεξάγουμε ταξινόμηση ανάμεσα στις δύο ομάδες των δεδομένων μας.

### 3.3.7 Έλεγχοι υποθέσεων για τα γραφοθεωρητικά στοιχεία

Σαν τελευταίο βήμα της ανάλυσης, στην προσπάθεια επικύρωσης και ανάδειξης των διαφορών στα γραφοθεωρητικά στοιχεία των δύο ομάδων (ασθενών-υγιών) πραγματοποιήθηκαν τυπικοί έλεγχοι υπόθεσης (formal hypothesis testing).

Γενικά, μπορούμε να συγκρίνουμε ατομικά ένα-ένα τα χαρακτηριστικά για τις δύο ομάδες ή να συγκρίνουμε και τα 13 χαρακτηριστικά μαζί, άμεσα. Ωστόσο, από τη στιγμή που τα χαρακτηριστικά αυτά είναι συσχετιζόμενα και συγγενικά μεταξύ τους, ένα γράφημα με μικρή σχετικά διάμεσο εγγύτητας (median closeness) σημαίνει ότι παρουσιάζει μία επίσης μικρή απόσταση μεταξύ δύο κόμβων του γράφου και άρα και μία αυξημένη μεταβατικότητα (transitivity). Για να συγκρίνουμε λοιπόν απευθείας και τα 13 χαρακτηριστικά για τις δύο ομάδες θα έπρεπε να προσαρμόσουμε τον έλεγχο μας για πολλαπλές συγκρίσεις χρησιμοποιώντας την διόρθωση Bonferonni (Bonferonni correction) όπου θα θεωρούσαμε σημαντικές μόνο p-τιμές ελέγχου κάτω από  $0.05/13$ .

## 4. Παρουσίαση αποτελεσμάτων

---

Στο κεφάλαιο αυτό θα παραθέσουμε τα αποτελέσματα της ανάλυσης που έγινε στο προηγούμενο κεφάλαιο. Ο αναγνώστης υπενθυμίζεται ότι μπορεί να χρησιμοποιήσει τον κώδικα που παράχθηκε στην ανάλυση μας καθώς και να αναπαράγει τα αποτελέσματα που θα παρατεθούν παρακάτω. Ο κώδικας θα αναρτηθεί στο παράρτημα της παρούσας εργασίας και αφορά την υλοποίηση της ανάλυσης που κάναμε στο κεφάλαιο 3 χρησιμοποιώντας το στατιστικό πακέτο FSL (για την προ-επεξεργασία των δεδομένων και την εφαρμογή της ICA) και τη γλώσσα R για στατιστικές εφαρμογές (όσον αφορά τα υπόλοιπα βήματα της ανάλυσης).

Παρακάτω βλέπουμε δύο FMRI σαρώσεις από ανθρώπους με σχιζοφρένεια (Εικόνες 17 και 19) και δύο σαρώσεις υγιών ανθρώπων του δείγματος (εικόνα 18 και 20), μετασχηματισμένες σε γράφους σύμφωνα με τη διαδικασία που αναλύσαμε στο προηγούμενο κεφάλαιο. Οι κόμβοι τους αναπαριστούν λειτουργικά δίκτυα του εγκεφάλου και οι αποστάσεις μεταξύ τους αναπαριστούν την ομοιότητα στη χρονική δραστηριότητα των κόμβων αυτών ή αλλιώς το μέγεθος της λειτουργικής διασυνδεσιμότητας μεταξύ τους.

Στο σημείο αυτό πρέπει να αναφέρουμε για τις παρακάτω εικόνες ότι η αναπαράσταση αυτή δεν αντιστοιχεί με πραγματικές αποστάσεις τομέων του εγκεφάλου ή την πραγματική τοποθεσία στη δομή του εγκεφάλου που εντοπίζονται πραγματικά εγκεφαλικά δίκτυα. Είναι μία αναπαράσταση 2-διαστάσεων που έδωσε το τελευταίο βήμα της μεθόδου ISOMAP, η εφαρμογή δηλαδή της μεθόδου MDS εφοδιασμένη όμως, με τις γεωδαισιακές αποστάσεις των ανεξάρτητων συνιστωσών (εκάστοτε υποκειμένου) μεταξύ τους.

Όπως προαναφέραμε οι αναπαραστάσεις αυτές είναι 2 - διαστάσεων και είναι αποτέλεσμα προσαρμογής του ISOMAP. Στις εικόνες 17 και 19 έχω δύο τυχαία γραφήματα σχιζοφρενών ενώ στις εικόνες 18 και 20 έχω δύο τυχαία γραφήματα υγιών του δείγματος. Οι άξονες Dim1 και Dim2 είναι σταθμίσεις (weightings) των δύο κύριων διαστάσεων της δομής που προσάρμοσε η ISOMAP προκειμένου να απεικονίσει πιστά τις ανεξάρτητες συνιστώσες σε χάρτη (ελαχιστοποιώντας μία συνάρτηση απόκλισης (stress function) ). Οι διαστάσεις αυτές έχουν να κάνουν με κύριες διευθύνσεις του συγχρονισμού των χρονοσειρών των δικτύων των υποκειμένων, όμως στο σημείο αυτό είναι πολύ δύσκολο να ερμηνευθούν με ακρίβεια τα συγκεκριμένα χαρακτηριστικά τα οποία εκφράζουν. Οι συντεταγμένες των κόμβων αποφασίστηκαν από τον πίνακα ελάχιστων αποστάσεων του ISOMAP και ουσιαστικά αντιστοιχούν στα 2 πρώτα ιδιοδιανύσματα του πίνακα .

Κάθε γράφημα δείχνει τη διασυνδεσιμότητα μεταξύ των κόμβων ενός υποκειμένου που ουσιαστικά κάθε κόμβος είναι και μία χρονοσειρά που αντιστοιχεί σε έναν ανεξάρτητο χωρικό χάρτη που έδωσε η ICA (spatial). Πρακτικά, αυτό σημαίνει ότι όσο πιο κοντά μεταξύ τους βρίσκονται δύο κόμβοι τόσο μεγαλύτερη η ομοιότητα στη δραστηριότητα. Αφού οι χρονοσειρές τους είναι κοντά, σύμφωνα με τη μετρική απόσταση που προσδιορίσαμε στο κεφάλαιο 3, θα είναι και μεγαλύτερη η λειτουργική

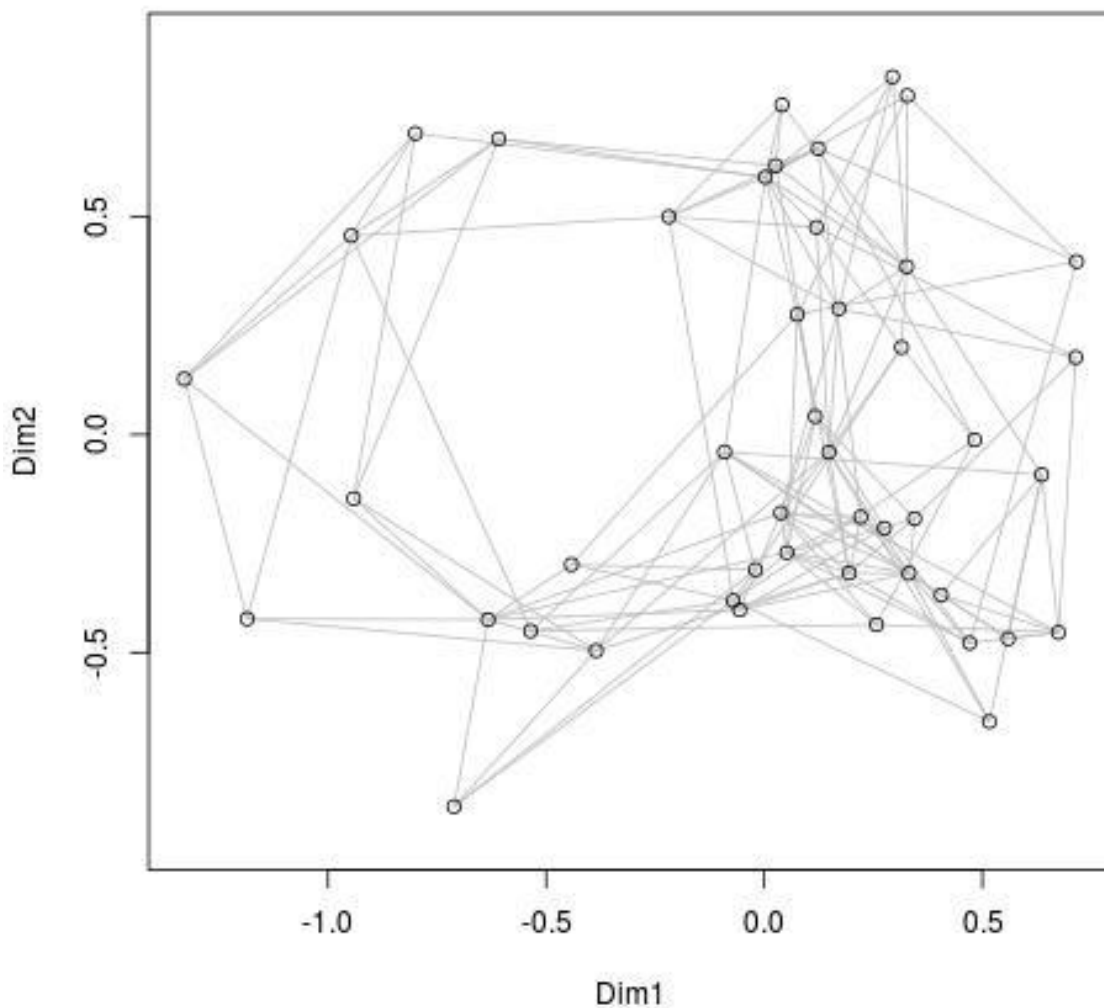
διασυνδεσιμότητα (συγχρονισμός) και η ομοιότητα στη συμπεριφορά μεταξύ των εγκεφαλικών δικτύων (κόμβοι) αυτών.

Από τα γραφήματα στο τέλος του κεφαλαίου (υγιείς, στις εικόνες 18,20 και σχιζοφρενείς, στις 17,19) μπορούμε να παρατηρήσουμε ότι για τα διαγράμματα των σχιζοφρενών οι αποστάσεις μεταξύ των κόμβων είναι μεγαλύτερες από των υγιών ενώ επίσης οι κόμβοι μεταξύ τους στην ομάδα των σχιζοφρενών είναι λιγότερο ομαδοποιημένοι-συγκροτημένοι από εκείνους των υγιών. Άρα συνολικά παρατηρούμε αμυδρά από τα γραφήματα ότι υπάρχει μία μεγαλύτερη διασυνδεσιμότητα στους κόμβους των υγιών από ότι στους κόμβους των σχιζοφρενών.

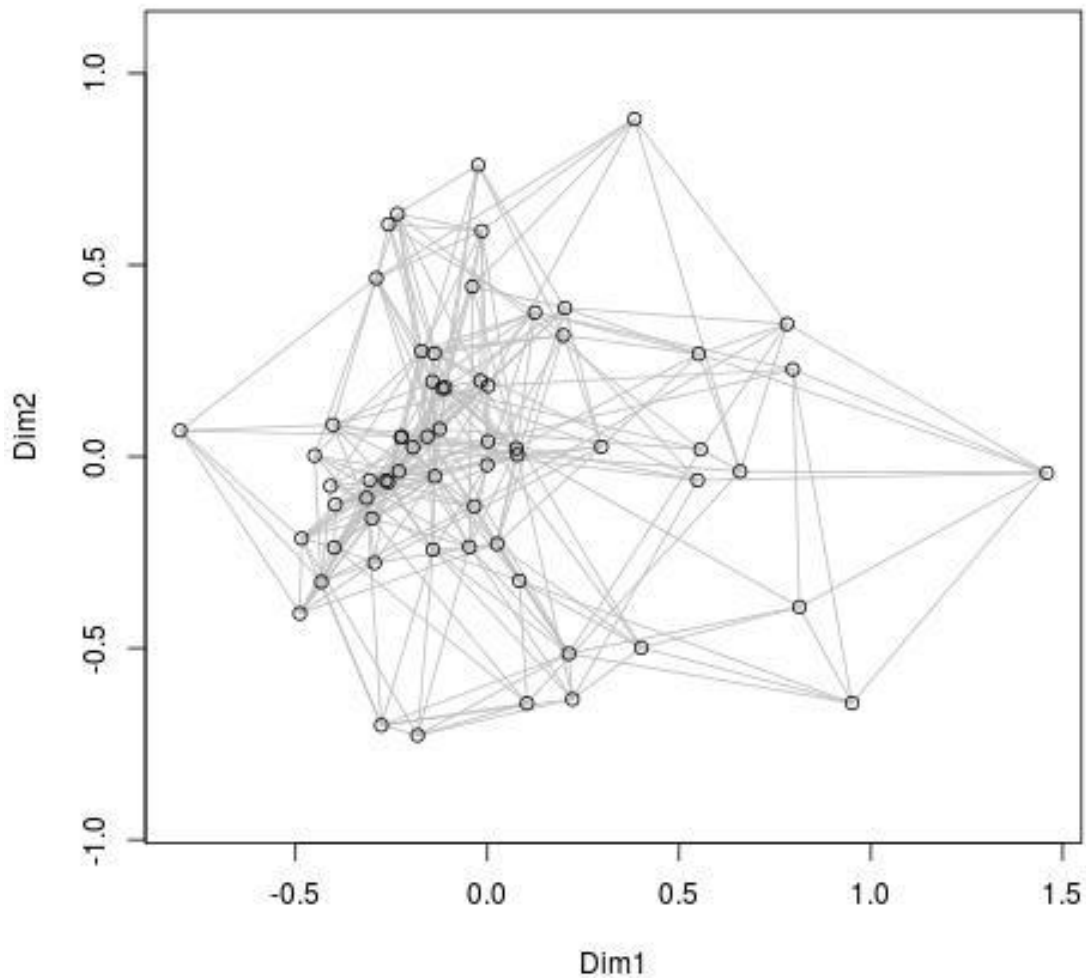
Στο αποτέλεσμα αυτό καταλήξαμε και από τους ελέγχους υποθέσεων που κάναμε για τα χαρακτηριστικά των γραφημάτων στις δύο ομάδες, αφού τα χαρακτηριστικά των οποίων η  $p$ -τιμή ελέγχου ήταν στατιστικά σημαντική ( $p < 0.05$ ) ήταν ο συντελεστής ομαδοποίησης-συγκρότησης (clustering coefficient), ο αριθμός των κλικών στο γράφημα και η μέση απόσταση μονοπατιού. Έχουμε λοιπόν ότι σύμφωνα με τους ελέγχους υποθέσεων οι σχιζοφρενείς έχουν μικρότερο συντελεστή συγκρότησης-ομαδοποίησης και μικρότερο αριθμό κλικών από ότι οι υγιείς του δείγματος. Χαμηλότερος συντελεστής ομαδοποίησης συγκρότησης σημαίνει εκ των πραγμάτων ότι οι κόμβοι (εγκεφαλικά δίκτυα) του γραφήματος για κάθε σχιζοφρενή είναι λιγότερο πιθανό να συνδεθούν ο ένας με τον άλλο, υποδεικνύοντας έτσι ότι τα δίκτυα είναι λιγότερο συγχρονισμένα και περισσότερο ανεξάρτητα μεταξύ τους. Ο αριθμός κλικών που δεν είναι ανεξάρτητος του συντελεστή συγκρότησης - ομαδοποίησης σημαίνει επίσης μεγαλύτερη διασυνδεσιμότητα διασυνδεσιμότητα των κόμβων μεταξύ τους. Επιπλέον, οι σχιζοφρενείς παρουσίασαν μικρότερη μέση απόσταση μονοπατιού (με μικρή όμως διαφορά των δειγματικών μέσων) ανάμεσα στους συνδεδεμένους κόμβους από ότι οι υγιείς. Τέλος και οι δύο ομάδες εμφάνισαν small-worldness χωρίς οριακά να πετύχουμε μια στατιστικά σημαντική διαφοροποίηση ( $p = 0.07 > 0.05$ ) ανάμεσα στους δείκτες των δύο ομάδων. Ωστόσο οι σχιζοφρενείς έδειξαν να έχουν μικρότερη (αν και δεν μπορέσαμε να απορρίψουμε την μηδενική υπόθεση) διασυνδεσιμότητα μικρού-κόσμου από ότι οι υγιείς.

Τέλος όσον αφορά την ταξινόμηση με SVM ταξινομητή, με χρήση της τεχνικής πολλαπλής επικύρωσης (Cross-validation), η οποία αξιολογεί το μοντέλο πρόβλεψης χωρίζοντας το αρχικό δείγμα σε ένα μέρος εκπαίδευσης (training set) και ένα μέρος για τη δοκιμή του αποτελέσματος (test set), πετύχαμε 65% επιτυχία έναντι του 50.7% που θα έδινε μία τυχαία ταξινόμηση αφού έχουμε στο δείγμα μας 74 υγιείς και 72 ασθενείς με σχιζοφρένεια.

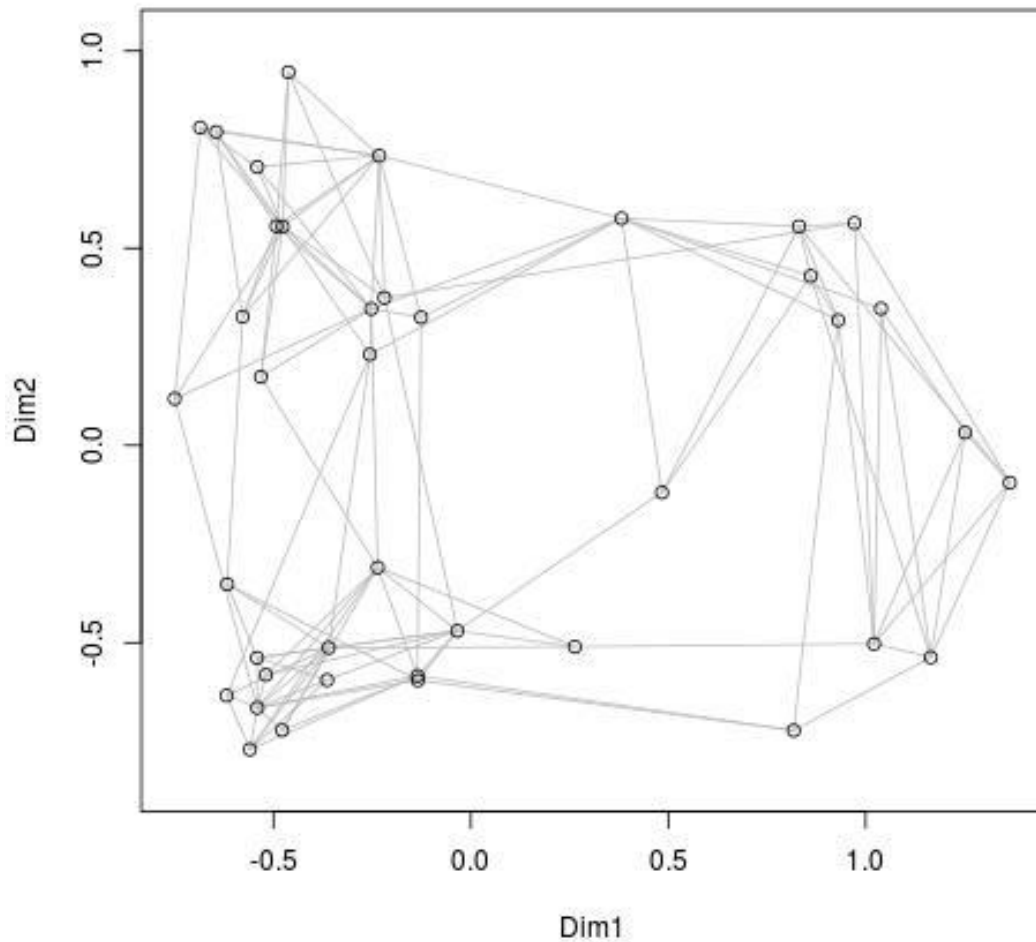




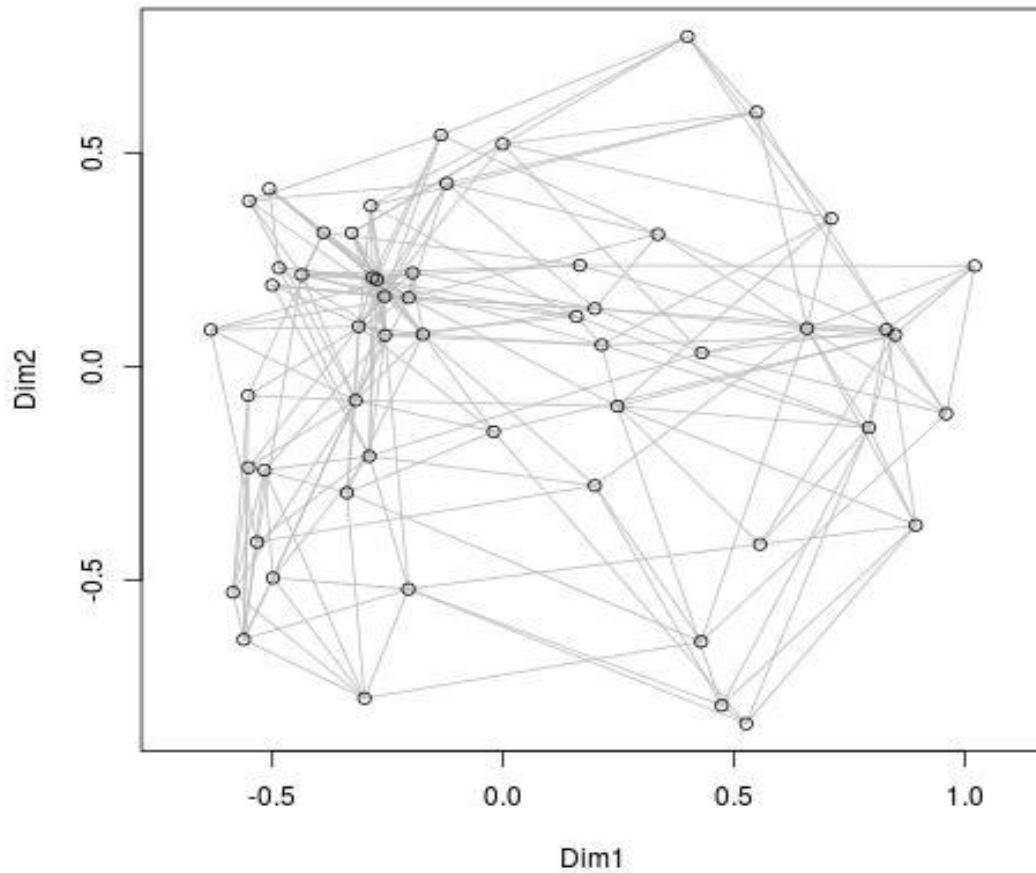
**Εικόνα 17.** Γράφημα ασθενούς διαγνωσμένου με σχιζοφρένεια που αναπαριστά την διασυνδεσιμότητα του. Εξάχθηκε από τη μετατροπή του πίνακα λειτουργικής διασυνδεσιμότητας σε μία δομή όπου ο κάθε κόμβος αναπαριστά μία συνιστώσα και η απόσταση μεταξύ των κόμβων αναπαριστά το βαθμό διασυνδεσιμότητας και την ομοιότητα στη συμπεριφορά. Οι διαστάσεις Dim 1 και Dim 2 είναι οι σταθμίσεις (weightings) των δύο κύριων διαστάσεων της προσαρμογής. Οι συντεταγμένες των κόμβων αντιστοιχούν στα δύο πρώτα ιδιοδιανύσματα του πίνακα ελαχίστων αποστάσεων του ISOMAP. Το παραπάνω γράφημα αντιστοιχεί στο υποκείμενο 0040106 του δείγματος .



**Εικόνα 18.** Γράφημα υγιούς ανθρώπου του δείγματος που αναπαριστά την διασυνδεσιμότητα του. Εξάχθηκε από τη μετατροπή του πίνακα λειτουργικής διασυνδεσιμότητας σε μία δομή όπου ο κάθε κόμβος αναπαριστά μία συνιστώσα και η απόσταση μεταξύ των κόμβων αναπαριστά το βαθμό διασυνδεσιμότητας και την ομοιότητα στη συμπεριφορά. Οι διαστάσεις Dim 1 και Dim 2 είναι οι σταθμίσεις (weightings) των δύο κύριων διαστάσεων της προσαρμογής. Οι συντεταγμένες των κόμβων αντιστοιχούν στα δύο πρώτα ιδιοδιανύσματα του πίνακα ελαχίστων αποστάσεων του ISOMAP. Το παραπάνω γράφημα αντιστοιχεί στο υποκείμενο 0040056 του δείγματος .



**Εικόνα 19.** Γράφημα ασθενούς διαγνωσμένου με σχιζοφρένεια που αναπαριστά την διασυνδεσιμότητα του. Εξάχθηκε από τη μετατροπή του πίνακα λειτουργικής διασυνδεσιμότητας σε μία δομή όπου ο κάθε κόμβος αναπαριστά μία συνιστώσα και η απόσταση μεταξύ των κόμβων αναπαριστά το βαθμό διασυνδεσιμότητας και την ομοιότητα στη συμπεριφορά. Οι διαστάσεις Dim 1 και Dim 2 είναι οι σταθμίσεις (weightings) των δύο κύριων διαστάσεων της προσαρμογής. Οι συντεταγμένες των κόμβων αντιστοιχούν στα δύο πρώτα ιδιοδιανύσματα του πίνακα ελαχίστων αποστάσεων του ISOMAP. Το παραπάνω γράφημα αντιστοιχεί στο υποκείμενο 0040032 του δείγματος .



**Εικόνα 20.** Γράφημα υγιούς ανθρώπου του δείγματος που αναπαριστά την διασυνδεσιμότητα του. Εξάχθηκε από τη μετατροπή του πίνακα λειτουργικής διασυνδεσιμότητας σε μία δομή όπου ο κάθε κόμβος αναπαριστά μία συνιστώσα και η απόσταση μεταξύ των κόμβων αναπαριστά το βαθμό διασυνδεσιμότητας και την ομοιότητα στη συμπεριφορά. Οι διαστάσεις Dim 1 και Dim 2 είναι οι σταθμίσεις (weightings) των δύο κύριων διαστάσεων της προσαρμογής. Οι συντεταγμένες των κόμβων αντιστοιχούν στα δύο πρώτα ιδιοδιανύσματα του πίνακα ελαχίστων αποστάσεων του ISOMAP. Το παραπάνω γράφημα αντιστοιχεί στο υποκείμενο 0040013 του δείγματος .

## 5. Συμπεράσματα

---

Στην παρούσα διπλωματική εργασία, παρουσιάσαμε μεθόδους για την ανάλυση λειτουργικής διασυνδεσιμότητας εγκεφαλικών δικτύων προκειμένου να αναδείξουμε τις υπάρχουσες διαφορές μεταξύ των ομάδων που ερευνήσαμε (υγιείς-ασθενείς με σχιζοφρένεια). Στόχος μας επίσης, ήταν να δείξουμε ότι τα δεδομένα FMRI resting-state μπορεί να είναι χρήσιμα για την αυτόματη διάγνωση των ασθενών. Μπορούμε λοιπόν να χρησιμοποιήσουμε μέτρα λειτουργικής διασυνδεσιμότητας για να διαχωρίσουμε ασθενείς από υγιείς, αφού όπως είδαμε οι σχιζοφρενείς έδειξαν να έχουν έναν στατιστικά μικρότερο δείκτη ομαδοποίησης-συγκρότησης (clustering coefficient) από ότι οι υγιείς, κάτι που πρακτικά σημαίνει ότι τα εγκεφαλικά δίκτυα των σχιζοφρενών είναι περισσότερο αποσυνδεδεμένα και λιγότερο συγχρονισμένα (**Liu et al., 2008**).

Επίσης χρησιμοποιώντας σχετικά απλές μεθόδους και έναν απλό ταξινομητή (με τον βασικό αλγόριθμο SVM) αναδείξαμε μία ταξινόμηση με επιτυχία 65% , ένα ποσοστό που ενώ φαίνεται να έχει χαμηλή αξιοπιστία είναι κάτι σημαντικό αν σκεφτεί κανείς ότι το μοντέλο μας ήταν πολύ απλό αφού δεν λάβαμε υπόψη πληροφορίες που θα μπορούσαν να αυξήσουν το ποσοστό επιτυχίας όπως δημογραφικά χαρακτηριστικά των υποκειμένων π.χ. ηλικία, φύλο κτλ. Επίσης δεν κάναμε έλεγχο ποιότητας των δεδομένων για να ξεσκεπάρουμε κάποια δεδομένα υποκειμένων που κινήθηκαν υπερβολικά την ώρα του πειράματος (excessive motion) ή ήταν αποπλανητικά λόγω κακών ρυθμίσεων ή τεχνικού ελαττώματος του σαρωτή MRI . Επίσης δεν έγινε κάποια ειδική μελέτη των ανεξάρτητων συνιστωσών του κάθε υποκειμένου που μπορεί να οφείλονται σε εξωγενείς παράγοντες όπως η κίνηση του υποκειμένου φυσιολογικές ανωμαλίες κτλ. Ακόμα θα μπορούσαμε να έχουμε ένα πιο ισορροπημένο δείγμα όσον αφορά το φύλο την ηλικία κτλ. των υποκειμένων. Για παράδειγμα στο δείγμα το οποίο αναλύσαμε οι άνδρες του δείγματος ήταν περισσότερο πιθανό να είναι σχιζοφρενείς από ότι οι γυναίκες. Αν σκεφτεί κανείς λοιπόν την απλοϊκότητα του μοντέλου το αποτέλεσμα μας είναι κάτι αξιοσημείωτο. Ωστόσο η ανάλυση αυτή, μπορεί να χρησιμοποιηθεί γενικά και σε άλλες περιπτώσεις με κάποιες παραλλαγές που θα βελτιώσουν δραματικά την απόδοση της.

Συμπερασματικά λοιπόν, δείξαμε πως η λειτουργική διασυνδεσιμότητα μπορεί να μετρηθεί από μία γραφοθεωρητική σκοπιά, χρησιμοποιώντας τις ιδιότητες μικρού κόσμου (small-world properties) εφαρμόζοντας γραμμικές και μη γραμμικές μεθόδους για τον μετασχηματισμό των 4-διαστάσεων δεδομένων μας (FMRI scans) σε γραφήματα με κόμβους τις ανεξάρτητες συνιστώσες των υποκειμένων και αποστάσεις τις γεωδαισιακές αποστάσεις μεταξύ τους. Ωστόσο υπάρχουν κι άλλες μέθοδοι για την εφαρμογή της ανάλυσης λειτουργικής διασυνδεσιμότητας όπως για παράδειγμα με τη μέθοδο της Granger αιτιότητας (Granger causality) (**Roebroeck et al., 2005**) ανάμεσα στις περιοχές του ενδιαφέροντος (ROIs) όπως παρουσίασαν οι **Sato et al. (2010)**, όπου αυτή η παρουσίαση εφαρμόζει λειτουργική διασυνδεσιμότητα μέσω συσχέτισης σε λειτουργικά δίκτυα που εξάγονται από την ICA. Ακόμα υπάρχουν και άλλες μετρικές αποστάσεις που θα μπορούσαμε να χρησιμοποιήσουμε αντί της πολλαπλής συσχέτισης (Cross-correlation) αφού σύμφωνα με πρόσφατα ευρήματα από τους **Zalesky et al.**

(2012) η χρήση μέτρων συσχέτισης για τον υπολογισμό απόστασης αυτόματα οδηγεί σε μη-τυχαίες γραφικές δομές.

Τέλος παρότι η ανάλυση αυτή έγινε για FMRI δεδομένα πιο γενικά μπορεί να εφαρμοστεί για προβλήματα όπου η σχέση στις πηγές των σημάτων μπορεί να αποφασίσει την κατηγορία στην οποία ανήκει το υποκείμενο μελέτης. Η κοινή συμπεριφορά των πηγών σημάτων (ανεξάρτητες συνιστώσες) παρατηρήθηκε σαν γράφημα όπου αποστάσεις ανάμεσα στις πηγές ανέδειξαν ομοιότητα στη συμπεριφορά. Παρόλο που στην παρούσα διπλωματική εργασία χρησιμοποιήσαμε μέτρα στατιστικής δεύτερης τάξης για να αναλύσουμε τη λειτουργική διασυνδεσιμότητα (συσχετίσεις), είναι πιθανό ότι τόση και περισσότερη δύναμη έχουν και στατιστικά εργαλεία υψηλότερης τάξης που λαμβάνουν υπόψη τη συνοχή (cohesiveness) τριών ή και περισσότερων συνιστωσών. Η λειτουργική διασυνδεσιμότητα λοιπόν, είναι μία τεχνική που θα πρέπει να εκτιμηθεί από διάφορες σκοπιές.

## 6. Βιβλιογραφία

---

Almási, G., Heidelberger, P., Archer, C. J., Martorell, X., Erway, C. C., Moreira, J. E., & Zheng, Y. (2005, June). Optimization of MPI collective communication on BlueGene/L systems. In *Proceedings of the 19th annual international conference on Supercomputing* (pp. 253-262). ACM.

Anderson, A., Dinov, I. D., Sherin, J. E., Quintana, J., Yuille, A. L., Cohen, M. S. (2010). Classification of spatially unaligned fMRI scans. *Neuroimage*, 49(3), 2509-2519.

Anderson, A., Bramen, J., Douglas, P. K., Lenartowicz, A., Cho, A., Culbertson, C. Cohen, M. S. (2011). Large sample group independent component analysis of functional magnetic resonance imaging using anatomical atlas-based reduction and bootstrapped clustering. *International Journal of Imaging Systems and Technology*, 21(2), 223-231.

Anderson, A., Cohen, M. S. (2013). Decreased small-world functional network connectivity and clustering across resting state networks in schizophrenia: an fMRI classification tutorial. *Frontiers in Human Neuroscience*, 7, 520.

Bandettini, P. A., Moonen, C. T. (1999). *Functional MRI*. Springer.

Bassett, D. S., Bullmore, E., Verchinski, B. A., Mattay, V. S., Weinberger, D. R., Meyer-Lindenberg, A. (2008). Hierarchical organization of human cortical networks in health and schizophrenia. *The Journal of Neuroscience*, 28(37), 9239-9248.

Baumgartner, R., Ryner, L., Richter, W., Summers, R., Jarmasz, M., Somorjai, R. (2000). Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis. *Magnetic Resonance Imaging*, 18(1), 89-94.

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 23(2), 137-152.

Beckmann, C. F., & Smith, S. M. (2005). Tensorial extensions of independent component analysis for multisubject fMRI analysis. *Neuroimage*, 25(1), 294-311.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6), 1129-1159.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34(4), 537-541.

Blamire, A. M., Ogawa, S., Ugurbil, K., Rothman, D., McCarthy, G., Ellermann, J. M. Shulman, R. G. (1992). Dynamic mapping of the human visual cortex by high-speed

*Magnetic Resonance Imaging. Proceedings of the National Academy of Sciences, 89(22), 11069-11073.*

*Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annual review of Neuroscience, 25(1), 151-188.*

*Buckner, R. L., Koutstaal, W., Schacter, D. L., Dale, A. M., Rotte, M., & Rosen, B. R. (1998). Functional-anatomic study of episodic retrieval: II. Selective averaging of event-related FMRI trials to test the retrieval success hypothesis. Neuroimage, 7(3), 163-175.*

*Calhoun, V. D., Adali, T., McGinty, V. B., Pekar, J. J., Watson, T. D., & Pearlson, G. D. (2001a). FMRI activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis. NeuroImage, 14(5), 1080-1088.*

*Calhoun, V. D., Adali, T., Pearlson, G. D., Pekar, J. J. (2001b). A method for making group inferences from functional MRI data using independent component analysis. Human Brain Mapping, 14(3), 140-151.*

*Calhoun, V. D., Adali, T., Pearlson, G. D., Pekar, J. J. (2001). Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Human Brain Mapping, 13(1), 43-53.*

*Calhoun, V. D., Adali, T., & Pekar, J. J. (2004). A method for testing conjunctive and subtractive hypotheses on group FMRI data using independent component analysis. Magnetic Resonance Imaging, 22(9), 1181-1191.*

*Calhoun, V. D., Liu, J., & Adali, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. Neuroimage, 45(1), S163-S172.*

*Cao, J., Worsley, K. (1999). The geometry of correlation fields with an application to functional connectivity of the brain. The Annals of Applied Probability, 9(4), 1021-1057.*

*Cecchi, G. A., Rao, A. R., Centeno, M. V., Baliki, M., Apkarian, A. V., Chialvo, D. R. (2007). Identifying directed links in large scale functional networks: application to brain FMRI. BMC Cell Biology, 8(Suppl 1), S5.*

*Chu, C., Handwerker, D. A., Bandettini, P. A., Ashburner, J. (2011, May). Measuring the consistency of global functional connectivity using kernel regression methods. In Pattern Recognition in Neuroimaging (PRNI), 2011 International Workshop on (pp. 41-44) IEEE.*

*Cohen, M. S., Bookheimer, S. Y. (1994). Localization of brain function using magnetic resonance imaging. Trends in Neuroscience, 17(7), 268-277.*



Comon, P. (1994). *Independent component analysis, a new concept?* *Signal Processing*, 36(3), 287-314.

Erdős, P., Rényi, A. (1961). *On the strength of connectedness of a random graph.* *Acta Mathematica Hungarica*, 12(1-2), 261-267.

Esposito, F., Formisano, E., Seifritz, E., Goebel, R., Morrone, R., Tedeschi, G., Di Salle, F. (2002). *Spatial independent component analysis of functional MRI time-series: To what extent do results depend on the algorithm used?* *Human Brain Mapping*, 16(3), 146-157.

Esposito, F., Scarabino, T., Hyvarinen, A., Himberg, J., Formisano, E., Comani, S. & Di Salle, F. (2005). *Independent component analysis of fMRI group studies by self-organizing clustering.* *Neuroimage*, 25(1), 193-205.

Ford, J., Farid, H., Makedon, F., Flashman, L. A., McAllister, T. W., Megalooikonomou, V., Saykin, A. J. (2003). *Patient classification of fMRI activation maps.* In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003* (pp. 58-65). Springer Berlin Heidelberg.

Friston, K. J., Frith, C. D., Liddle, P. F., Frackowiak, R. S. J. (1993). *Functional connectivity: the principal-component analysis of large (PET) data sets.* *Journal of Cerebral Blood Flow & Metabolism*, 13(1), 5-14.

Friston, K. J. (1994). *Functional and effective connectivity in neuroimaging: a synthesis.* *Human Brain Mapping*, 2(1-2), 56-78.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., Frackowiak, R. S. (1994). *Statistical parametric maps in functional imaging: a general linear approach.* *Human Brain Mapping*, 2(4), 189-210.

Garrity, A. G., Pearlson, G. D., McKiernan, K., Lloyd, D., Kiehl, K. A., Calhoun, V. D. (2007). *Aberrant “default mode” functional connectivity in schizophrenia.* *American Journal of Psychiatry*.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., Gore, J. C. (1999). *Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects.* *Nature Neuroscience*, 2(6), 568-573.

Georgiev, P., Theis, F., Cichocki, A., Bakardjian, H. (2007). *Sparse component analysis: a new tool for data mining.* In *Data mining in biomedicine* (pp. 91-116). Springer US.

Grama, A., Gupta, A., Karypis, G. (1994). *Introduction to parallel computing: design and analysis of algorithms* (Vol. 400). Redwood City, CA: Benjamin/Cummings.

Heine, L., Soddu, A., Gómez, F., Vanhaudenhuyse, A., Tshibanda, L., Thonnard, M., Demertzi, A. (2012). *Resting State Networks and Consciousness: Alterations of Multiple Resting State Network Connectivity in Physiological, Pharmacological, and Pathological Consciousness States.* *Frontiers in Psychology*, 3, 295.

Humphries, M. D., Gurney, K. (2008). Network 'small-world-ness': a quantitative method for determining canonical network equivalence. *PloS one*, 3(4), e0002051.

Hyvarinen, A. (1998). New approximations of differential entropy for independent component analysis and projection pursuit. *Advances in Neural Information Processing Systems*, 10(2), 273-279.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, 13(4), 411-430.

Jafri, M. J., Pearlson, G. D., Stevens, M., Calhoun, V. D. (2008). A method for functional network connectivity among spatially independent resting-state components in schizophrenia. *Neuroimage*, 39(4), 1666-1681.

Jolliffe, I. (2002). *Principal component analysis*. John Wiley & Sons, Ltd.

Kim, S. G., Rostrup, E., Larsson, H. B., Ogawa, S., Paulson, O. B. (1999). Determination of relative CMRO<sub>2</sub> from CBF and BOLD changes: significant increase of oxygen consumption rate during visual stimulation. *Magnetic Resonance in Medicine*, 41(6), 1152-1161.

Koshino, H., Carpenter, P. A., Minshew, N. J., Cherkassky, V. L., Keller, T. A., Just, M. A. (2005). Functional connectivity in an fMRI working memory task in high-functioning autism. *Neuroimage*, 24(3), 810-821.

Lee, S. P., Duong, T. Q., Yang, G., Iadecola, C., Kim, S. G. (2001). Relative changes of cerebral arterial and venous blood volumes during increased cerebral blood flow: implications for BOLD fMRI. *Magnetic Resonance in Medicine*, 45(5), 791-800.

Li, K., Guo, L., Nie, J., Li, G., Liu, T. (2009). Review of methods for functional brain connectivity detection using fMRI. *Computerized Medical Imaging and Graphics*, 33(2), 131-139.

Liang, M., Zhou, Y., Jiang, T., Liu, Z., Tian, L., Liu, H., Hao, Y. (2006). Widespread functional disconnectivity in schizophrenia with resting-state functional Magnetic Resonance Imaging. *Neuroreport*, 17(2), 209-213.

Liow, J. S., Rehm, K., Strother, S. C., Anderson, J. R., Morch, N., Hansen, L. K., Schaper, K. (2000). Comparison of Voxel - and - Volume - of - Interest Based Analyses in FDG PET Scans of HIV Positive and Healthy Individuals.

Liu, Y., Liang, M., Zhou, Y., He, Y., Hao, Y., Song, M., Jiang, T. (2008). Disrupted small-world networks in schizophrenia. *Brain: A Journal of Neurology*, 131(Pt 4), 945-961.

Lukic, A. S., Wernick, M. N., Hansen, L. K., Strother, S. C. (2002). An ICA algorithm for analyzing multiple data sets. In *Image Processing. 2002. Proceedings. 2002 International Conference on* (Vol. 2, pp. II-821) IEEE.

Ma, L., Wang, B., Chen, X., Xiong, J. (2007). *Detecting functional connectivity in the resting brain: a comparison between ICA and CCA. Magnetic Resonance Imaging, 25(1), 47-56.*

Mardia, K. V., & Kent, J. T. (1979). *Bibby. JM Multivariate analysis. London: Academic.*

McKeown, M., Makeig, S., Brown, S., Jung, T. P., Kindermann, S., Bell, A. J. & Sejnowski, T. (1998). *Blind separation of functional magnetic resonance imaging (fMRI) data. Human Brain Mapping, 6(5-6), 368-372.*

McKeown, M. J., Hansen, L. K., Sejnowski, T. J. (2003). *Independent component analysis of functional MRI: what is signal and what is noise? Current opinion in Neurobiology, 13(5), 620-629.*

Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., Buckner, R. L. (2000). *Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. Neuroimage, 11(6), 735-759.*

Ogawa, S., Lee, T. M., Nayak, A. S., Glynn, P. (1990). *Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magnetic Resonance in Medicine, 14(1), 68-78.*

Roebroek, Alard, Elia Formisano, and Rainer Goebel. "Mapping directed influence over the brain using Granger causality and fMRI." *Neuroimage 25.1* (2005): 230-242.

Roy, C. S., Sherrington, C. S. (1890). *On the regulation of the blood-supply of the brain. The Journal of Physiology, 11(1-2), 85.*

Rubinov, M., Sporns, O. (2010). *Complex network measures of brain connectivity: uses and interpretations. Neuroimage, 52(3), 1059-1069.*

Sato, J. R., Fujita, A., Cardoso, E. F., Thomaz, C. E., Brammer, M. J., Amaro, E. (2010). *Analyzing the connectivity between regions of interest: an approach based on cluster Granger causality for fMRI data analysis. Neuroimage, 52(4), 1444-1455.*

Schmithorst, V. J., Holland, S. K. (2004). *Comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data. Journal of Magnetic Resonance Imaging, 19(3), 365-368.*

Smith, A. M., Lewis, B. K., Ruttimann, U. E., Frank, Q. Y., Sinnwell, T. M., Yang, Y. Frank, J. A. (1999). *Investigation of low frequency drift in fMRI signal. Neuroimage, 9(5), 526-533.*

Stevens, M. C., Kiehl, K. A., Pearlson, G., Calhoun, V. D. (2007). *Functional neural circuits for mental timekeeping. Human Brain Mapping, 28(5), 394-408.*

Sun, F. T., Miller, L. M., D'Esposito, M. (2004). *Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. Neuroimage, 21(2), 647-658.*

Tenenbaum, J. B., de Silva, V., Langford, J. C. (2000). *A Global Geometric Framework for Nonlinear Dimensionality Reduction. Ann. NY Acad. Sci, 911, 418.*

Toppi, J., de Vico Fallani, F., Vecchiato, G., Maglione, A. G., Cincotti, F., Mattia, D. & Astolfi, L. (2012). *How the statistical validation of functional connectivity patterns can prevent erroneous definition of small-world properties of a brain connectivity network. Computational and Mathematical Methods in Medicine, 2012.*

Van de Ven, V. G., Formisano, E., Prvulovic, D., Roeder, C. H., Linden, D. E. (2004). *Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. Human Brain Mapping, 22(3), 165-178.*

Warrington, E. K. (1986). *Visual deficits associated with occipital lobe lesions in man. Experimental Brain Research Supplementum, 11, 247-261.*

Whitfield-Gabrieli, S., Thermenos, H. W., Milanovic, S., Tsuang, M. T., Faraone, S. V., McCarley, R. W., Wojcik, J. (2009). *Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. Proceedings of the National Academy of Sciences, 106(4), 1279-1284.*

Worsley, K. J., Chen, J. I., Lerch, J., Evans, A. C. (2005). *Comparing functional connectivity via thresholding correlations and singular value decomposition. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360(1457), 913-920.*

Yu, Q., Sui, J., Rachakonda, S., He, H., Gruner, W., Pearlson, G. Calhoun, V. D. (2011). *Altered topological properties of functional network connectivity in schizophrenia during resting state: a small-world brain network study. PloS one, 6(9), e25423.*

Zalesky, A., Fornito, A., Bullmore, E. (2012). *On the use of correlation as a measure of network connectivity. Neuroimage, 60(4), 2096-2106.*

Zhang, L., Samaras, D., Tomasi, D., Volkow, N., Goldstein, R. (2005, June). *Machine learning for clinical diagnosis from Functional Magnetic Resonance Imaging. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 1211-1217) IEEE.*

Zhao, X., Glahn, D., Tan, L. H., Li, N., Xiong, J., Gao, J. H. (2004). *Comparison of TCA and ICA techniques in fMRI data processing. Journal of Magnetic Resonance Imaging, 19(4), 397-402.*

## Παράρτημα

---

Στο κεφάλαιο αυτό θα παραθέσουμε τον κώδικα που χρησιμοποιήθηκε για την ανάλυση λειτουργικής διασυνδεσιμότητας των FMRI δεδομένων που παρουσιάστηκαν εκτενώς στο κεφάλαιο 3.

Για την ανάλυση μας χρησιμοποιήθηκε το στατιστικό πακέτο FSL που διανέμεται δωρεάν στη διεύθυνση <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki> για την προ-επεξεργασία των δεδομένων και ανάλυση σε ανεξάρτητες συνιστώσες και η γλώσσα R η οποία επίσης διανέμεται δωρεάν στη διεύθυνση <https://www.r-project.org> και κάποιες από τις βιβλιοθήκες της για τη μετέπειτα ανάλυση. Να σημειωθεί επίσης ότι η διαδικασία της μελέτης των δεδομένων μας έλαβε χώρα σε περιβάλλον Linux και συγκεκριμένα Ubuntu.

Όλες οι εντολές δόθηκαν απευθείας στο terminal του λειτουργικού οπότε για διευκόλυνση του αναγνώστη και για τη δυνατότητά του να παράγει και αυτός τα αποτελέσματα που παράχθηκαν στην ανάλυση οι εντολές θα ακολουθήσουν κατά σειρά προτεραιότητας όπως ακριβώς εκτελέστηκαν στην ανάλυση μας.

Για την προ-επεξεργασία των δεδομένων και χωρική (spatial) ανάλυση ανεξάρτητων συνιστωσών με τη χρήση της συνάρτησης `melodic` έχουμε το παρακάτω κομμάτι του κώδικα στον οποίο με κόκκινο χρώμα και με # μπροστά από κάθε γραμμή εκτέλεσης βρίσκονται σχόλια για τη διευκόλυνση του αναγνώστη :

```
#!/bin/sh
```

```
#Τοποθεσία φακέλου δεδομένων COBRE, ο χρήστης θα πρέπει να αλλάξει την  
#τοποθεσία ανάλογα με την τοποθεσία των φακέλων
```

```
STUDY_DIR="/home/User/Desktop/COBRE";
```

```
#####
```

```
# CONSTANTS #
```

```
#####
```

```
subjects="0040000 0040005 0040010 0040045 0040050 0040055 0040060 0040015  
0040065 0040020 0040070 0040025 0040075 0040030 0040080 0040035 0040085  
0040040 0040090 0040095 0040100 0040105 0040110 0040115 0040120 0040125  
0040130 0040135 0040140 0040145 0040001 0040006 0040011 0040016 0040021  
0040026 0040031 0040036 0040041 0040046 0040051 0040056 0040061 0040066  
0040071 0040076 0040081 0040086 0040091 0040096 0040101 0040106 0040111  
0040116 0040121 0040126 0040131 0040136 0040141 0040146 0040002 0040007  
0040012 0040017 0040022 0040027 0040032 0040037 0040042 0040047 0040052  
0040057 0040062 0040067 0040072 0040077 0040082 0040087 0040092 0040097  
0040102 0040107 0040112 0040117 0040122 0040127 0040132 0040137 0040142  
0040147 0040003 0040008 0040013 0040018 0040023 0040028 0040033 0040038  
0040043 0040048 0040053 0040058 0040063 0040068 0040073 0040078 0040083  
0040088 0040093 0040098 0040103 0040108 0040113 0040118 0040123 0040128  
0040133 0040138 0040143 0040004 0040009 0040014 0040019 0040024 0040029  
0040034 0040039 0040044 0040049 0040054 0040059 0040064 0040069 0040074
```

```

0040079 0040084 0040089 0040094 0040099 0040104 0040109 0040114 0040119
0040124 0040129 0040134 0040139 0040144";
umask 0002;
#####
# PROCESSING COMMANDS #
#####
# Μετάβαση στο φάκελο STUDY_DIR
cd $STUDY_DIR;
mkdir COBRE_MELODIC
# Loop through subjects
for i in $subjects; do
if [ ! -f "$STUDY_DIR/COBRE_MELODIC/${i}_melodic_mix_new" ]; then
cd $STUDY_DIR/${i}/session_1/rest_1
rm -r *.ica*
rm rest_mcf*
rm prefiltered*
rm filtered*
fsl5.0-mcflirt -in rest.nii.gz -out prefiltered_func_data_mcf -mats -rmsrel -rmsabs
fsl5.0-fslmaths prefiltered_func_data_mcf -Tmean mean_func
fsl5.0-bet2 mean_func mask -f 0.3 -n -m; fsl5.0-immv mask_mask mask
fsl5.0-fslmaths prefiltered_func_data_mcf -mas mask prefiltered_func_data_bet
fsl5.0-fslstats prefiltered_func_data_bet -p 2 -p 98
fsl5.0-fslmaths prefiltered_func_data_bet -thr 100.8095459 -Tmin -bin mask -odt char
fsl5.0-fslstats prefiltered_func_data_mcf -k mask -p 50
fsl5.0-fslmaths mask -dilF mask
fsl5.0-fslmaths prefiltered_func_data_mcf -mas mask prefiltered_func_data_thresh
fsl5.0-fslmaths prefiltered_func_data_thresh -Tmean mean_func
fsl5.0-susan prefiltered_func_data_thresh 614.340225 2.12314225053 3 1 1
mean_func 614.340225 prefiltered_func_data_smooth
fsl5.0-fslmaths prefiltered_func_data_smooth -mas mask
prefiltered_func_data_smooth
fsl5.0-fslmaths prefiltered_func_data_smooth -mul 12.2082189881
prefiltered_func_data_intnorm
fsl5.0-fslmaths prefiltered_func_data_intnorm -bptf 25.0 -1
prefiltered_func_data_tempfilt
fsl5.0-fslmaths prefiltered_func_data_tempfilt filtered_func_data
fsl5.0-fslmaths filtered_func_data -Tmean mean_func
fsl5.0-melodic -i filtered_func_data --nobet --bgthreshold=3 --tr=2.0000000000 -d 0 --
mmthresh=0.5
mkdir $STUDY_DIR/COBRE_MELODIC/${i}_melodic_mix_new
cp filtered_func_data.ica/melodic_mix
$STUDY_DIR/COBRE_MELODIC/${i}_melodic_mix_new
fi
done
#END "Loop through subjects..."
echo "Processing complete.";

```

Μόλις η διαδικασία προ-επεξεργασίας και ανάλυσης χωρικά ανεξάρτητων συνιστωσών τελειώσει, μια διαδικασία που μπορεί να πάρει από 10 ώρες έως και μέρες

ανάλογα με το σύστημα που θα χρησιμοποιήσει κάποιος, τότε γράφουμε ξανά στο terminal τις παρακάτω εντολές που περιλαμβάνουν τα υπόλοιπα βήματα που αναλύσαμε εκτενώς στο κεφάλαιο 3 (με κόκκινο χρώμα και με # μπροστά από κάθε γραμμή εκτέλεσης βρίσκονται και πάλι σχόλια για τη διευκόλυνση του αναγνώστη) :

```
#Ανοίγω την R
```

```
R
```

```
#Καλώ τις βιβλιοθήκες που θα χρειαστώ στην ανάλυση μου
```

```
library(igraph)
```

```
library(vegan)
```

```
library(e1071)
```

```
#Ορίζω το χώρο στον οποίο θα δουλέψω (working space)
```

```
setwd("/home/User/Desktop/COBRE/COBRE_MELODIC")
```

```
#Ως filenames ορίζω μόνο φακέλους με το παρακάτω πρότυπο
```

```
filenames <- dir (pattern="melodic_mix_new")
```

```
num_subjects <- length(filenames)
```

```
s <- c(rep(0,num_subjects))
```

```
#Διαβάζω κάθε melodic mix για να δω πόσες συνιστώσες έχει (dimension)
```

```
for (i in 1:num_subjects)
```

```
{ s[i] <- dim((read.table(as.character(filenames[i]))))[2] }
```

```
#Φτιάχνω ένα πίνακα με διαστάσεις των αριθμό των ασθενών επί το μέγιστο αριθμό
```

```
#ανεξάρτητων συνιστωσών
```

```
#Υπολογίζω την μετρική απόσταση ανάμεσα στις συνιστώσες των υποκειμένων με
```

```
#lag max(καθυστερήσεις)=3
```

```
#Πίνακας συμμετρικός με διαγώνιο 0
```

```
data_array_distance <- array(NA, c(num_subjects, max(s), max(s)))
```

```
for (i in 1:num_subjects)
```

```
{
```

```
temp <- as.matrix(read.table(as.character(filenames[i])))
```

```
for(j in 1:s[i])
```

```
{
```

```
for(k in j:s[i])
```

```
{
```

```
data_array_distance[i,k,j] <- 1-max(abs(ccf(temp[,j], temp[,k],
```

```
plot = FALSE, lag.max = 3)$acf))
```

```
data_array_distance[i,j,k] <- data_array_distance[i,k,j] }}
```

```
diag(data_array_distance[i,,]) <- 0
}
```

*#Παράγουμε μία συνάρτηση που θα χρησιμοποιήσουμε αργότερα. Η συνάρτηση αυτή  
#θα παίρνει ως είσοδο το αποτέλεσμα από την κλίση της μεθόδου ISOMAP και  
#φτιάχνει ένα γράφο για το κάθε υποκείμενο με κόμβους τις ανεξάρτητες συνιστώσες.*

```
makegraph <- function(my_iso)
{
my_dist <- as.matrix(dist(my_iso$points[]))
k <- dim(my_dist)[1]
#initialize gia to diktuo mas
my_net <- matrix(0, nrow = k, ncol = k)
#topothethsh timwn mas apo to isomap object
which.rows <- my_iso$net[,1]
which.cols <- my_iso$net[,2]
for(j in 1:length(which.rows))
{
my_net[which.rows[j],which.cols[j]]<- 1/my_dist[which.rows[j],which.cols[j]]
my_net[which.cols[j],which.rows[j]]<- 1/my_dist[which.cols[j],which.rows[j]]
}
my_net
}
```

*#Ανάλυση μικρού-κόσμου (small-worldness) για το γράφο με  
#κατασκευή συνάρτησης για τον υπολογισμό των μέτρων small-world*

```
smallworld <- function(n,m)
{ smallworld <- matrix(nrow=5000,ncol=2)
for(k in 1:5000)
{
g <- erdos.renyi.game(n,m, type="gnm")
smallworld[k,1] <- transitivity(g)
smallworld[k,2] <- average.path.length(g) }
colMeans(smallworld)
}
```

*#Εδώ έχουμε την υλοποίηση των μεθόδων κατά-σειρά χρησιμοποιώντας τις  
#συναρτήσεις που έχουμε κατασκευάσει παραπάνω.*

```
my_small_worldness <- matrix(NA, nrow = num_subjects, ncol = 1)
my_feature_matrix <- matrix(NA, nrow = num_subjects, ncol = 12)
randomsmallstore <- matrix(NA, nrow = num_subjects, ncol = 2)

for(i in 1:num_subjects)
{
d <- matrix(data_array_distance[i,1:s[i],1:s[i]], nrow = s[i])
my_iso <- isomap(d[1:s[i],1:s[i]],axes=3, k=max(floor(s[i]/10),2), ndim =
15,fragmentedOK=TRUE)
```



```

#Αποθήκευση των διαγραμμάτων που δίνει το Isomap
#Απαιτείται αλλαγή μονοπατιού στο σημείο που θέλει την αποθήκευση ο χρήστης

mypath <- file.path("/", "home", "User", "Desktop", "plots
isomap", paste("Isomap_plot", filenames[i], ".jpg", sep = ""))
jpeg(filename=mypath)
plot(my_iso)
dev.off()

#Υπολογισμός γραφοθεωρητικών στοιχείων και χαρακτηριστικών small-world του
#γράφου που πήραμε από το ISOMAP

my_net <- makegraph(my_iso)
d2 <- graph.adjacency(my_net )
transitivity(d2)
n=vcount(d2)
m = ecount(d2)
randomsmall <- smallworld(n,m)
sigma <-(transitivity(d2)/randomsmall[1])/(average.path.length(d2)/randomsmall[2])
randomsmallstore[i,] <- c(randomsmall)
my_small_worldness[i,] <- sigma
my_net <- makegraph(my_iso)
d2 <- graph.adjacency(my_net, weighted = TRUE )
my_feature_matrix[i,] <-
c(average.path.length(d2),clique.number(d2),graph.density(d2),
edge.connectivity(d2),median(closeness(d2)),median(graph.coreness(d2)),
max(degree(d2)),median(degree(d2)),min(degree(d2)),vcount(d2),ecount(d2),
transitivity(d2))
}
my_feature_matrix <- cbind(my_feature_matrix,my_small_worldness)

colnames(my_feature_matrix) <- c("Average Path Length", "Clique Number","Graph
Density","Edge Connectivity", "Median Closeness", "Median Graph Coreness","Max
Degree","Median Degree","Min Degree", "Vertex Count", "Edge
Count", "Transitivity","Small Worldness")

#Διαβάζω τα δεδομένα των φαινότυπων του δείγματος και ξεσκαρτάρω αυτούς με
#την επιγραφή disenrolled

cobre <- read.csv("COBRE_phenotypic_data.csv")
cobre <- cobre[order(cobre[,1]),]
my_cat <- cobre[,5]
my_data_matrix <- my_feature_matrix[my_cat!="Disenrolled",]
my_cat <- my_cat[my_cat!="Disenrolled"]
my_data_matrix <- cbind(my_cat,my_data_matrix)

#Καλώ τη συνάρτηση για τη χρήση του SVM #ταξινομητή με την επιλογή cross=10
#για 10-fold cross validation

my_svm <- svm(as.factor(my_cat)~.,data = as.data.frame(my_data_matrix),cross=10)

```

```
pred <- fitted(my_svm)
```

```
#Ελεγχοι υποθέσεων για τα 13 μεγέθη που υπολογίστηκαν για τις δύο ομάδες
```

```
for(i in 2:14)
```

```
{
```

```
print(t.test(na.omit(my_data_matrix[my_cat=="Patient",i]),na.omit(my_data_matrix[  
my_cat=="Control",i]))) }
```