

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΕΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

ΟΜΑΔΟΠΟΙΗΣΗ ΗΜΕΡΟΛΟΓΙΩΝ ΚΙΝΗΣΗΣ

ΣΠΥΡΙΔΟΥΔΑ Ν. ΜΗΤΣΗ

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος
Ειδίκευσης στην Εφαρμοσμένη Στατιστική

ΠΕΙΡΑΙΑΣ, ΣΕΠΤΕΜΒΡΙΟΣ 2014

Πανεπιστήμιο Πειραιώς

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΕΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

ΟΜΑΔΟΠΟΙΗΣΗ ΗΜΕΡΟΛΟΓΙΩΝ ΚΙΝΗΣΗΣ

ΣΠΥΡΙΔΟΥΔΑ Ν. ΜΗΤΣΗ

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος
Ειδίκευσης στην Εφαρμοσμένη Στατιστική

ΠΕΙΡΑΙΑΣ, ΣΕΠΤΕΜΒΡΙΟΣ 2014

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική

Τα μέλη της Επιτροπής ήταν:

- Πελέκης Ν. (Επιβλέπων)
- Θεοδωρίδης Ι.
- Ψαρράκος Γ.

Η έγκριση της Διπλωματικής Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα.

UNIVERSITY OF PIRAEUS



**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN
APPLIED STATISTICS**

CLUSTERING IN TRAVEL DIARIES

Spyridoula N. Mitsi

MSc Dissertation

submitted to the Department of Statistics and Insurance Science
of the University of Piraeus in partial fulfilment of the
requirements for the degree of Master of Science in Applied
Statistics

Piraeus, September 2014

Πανεπιστήμιο Πειραιώς

Στην Ιωάννα και στη Μάρι

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

Ευχαριστίες

Η παρούσα διπλωματική εργασία είναι το αποτέλεσμα μίας σειράς αλληλεπιδράσεων με διάφορα άτομα, κάθε ένα από τα οποία έπαιξε σημαντικό ρόλο στην εξέλιξη της. Πρώτα απ' όλα θα ήθελα να ευχαριστήσω την αδελφή μου Ιωάννα και την αγαπημένη μου φίλη-συμφοιτήτρια, σε όλα τα χρόνια της φοιτητικής μου ζωής, Μαρία Αγγελάκη γιατί χωρίς την πολύτιμη υποστήριξη τους η ολοκλήρωση της παρούσας διπλωματικής εργασίας θα ήταν αδύνατη.

Δεν θα μπορούσα να ξεχάσω και να μην ευχαριστήσω θερμά τους συναδέλφους στο Τμήμα Τεχνικής Βοήθειας στη ΜΟΔ και ιδιαίτερα τη Γιάννα, τη Βαρβάρα, την Ελένη και τη Βασούλα για την αμέριστη ψυχολογική- ηθική τους συμπαράσταση σε όλη τη διάρκεια της προετοιμασίας μου.

Επιπλέον θα ήθελα να ευχαριστήσω τον επιβλέπων καθηγητή μου κύριο Νίκο Πελέκη για τη συμβολή και την κατανόηση που έδειξε όλο αυτό το χρονικό διάστημα.

Τέλος ένα μεγάλο ευχαριστώ στους γονείς μου για τη στήριξη και την εμπιστοσύνη που μου έδειξαν όλα τα χρόνια των σπουδών μου.

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

Περίληψη

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η διερεύνηση πιθανών αλληλεπιδράσεων μεταξύ μεταβλητών που αναφέρονται σε τροχιές κινούμενων αντικειμένων. Τα δεδομένα που χρησιμοποιήθηκαν προήλθαν από μία πόλη του Βελγίου και συλλέχθηκαν από δύο πηγές δεδομένων, τα «δεδομένα GPS» τα οποία αποτελούν εγγραφές GPS ανά χρήστη και πιο συγκεκριμένα προέρχονται από τις ενεργοποιημένες συσκευές GPS και τα «Ημερολόγια» κίνησης στα οποία κάθε χρήστης καταγράφει τη σημασιολογική πληροφορία για το ταξίδι του.

Πιο συγκεκριμένα προκειμένου να επιτευχθεί ο σκοπός μας εφαρμόσαμε στα δεδομένα μας τη μέθοδο της συσταδοποίησης. Εφαρμόσαμε τον στοχαστικό ιεραρχικό αλγόριθμο συσταδοποίησης συσσωρευτικά, ομαδοποιώντας τα δεδομένα μας σε συμπαγείς συστάδες δηλαδή βάση της μέγιστης τετραγωνικής Ευκλείδειας απόστασης που έχουν μεταξύ τους οι συστάδες. Τρία διαφορετικά ζεύγη δοκιμάστηκαν και τα αποτελέσματα στα οποία καταλήξαμε μας έδειξαν ότι μόνο σε ένα από αυτά οι μεταβλητές συσχετίζονται ικανοποιητικά.

Πανεπιστήμιο Πειραιώς

Abstract

The subject of this thesis is to investigate possible interactions between variables which are listed in trajectories of moving objects. The data that were used came from a city in Belgium and were collected from two sources, the «data GPS» which are GPS records per user and more specifically from the GPS enabled devices and the «Diaries» in which each user records the semantic information about the trip.

More specifically, in order to achieve our purpose we applied to our data the method of clustering. We applied the stochastic hierarchical clustering algorithm cumulatively by grouping our data in compact clusters ie based on the maximum square Euclidean distance that these clusters have between them. Three different pairs of clustering were tested and the results that we have reached, indicated that only one pair of these variables correlate quite well.

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

Περιεχόμενα

Κατάλογος Πινάκων.....	xv
Κατάλογος Σχημάτων.....	xvii
ΕΙΣΑΓΩΓΗ.....	1
ΚΕΦΑΛΑΙΟ 1	5
Η ΕΝΝΟΙΑ ΤΗΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ ΣΤΙΣ ΣΗΜΑΣΙΟΛΟΓΙΚΕΣ ΤΡΟΧΙΕΣ	5
1.1 Βασικές έννοιες της τροχιάς ενός κινούμενου αντικειμένου	6
1.1.1 Η έννοια της ακατέργαστης τροχιάς.....	7
1.1.2 Από τις ακατέργαστες τροχιές στις σημασιολογικές τροχιές	8
1.2 Τμηματοποίηση Τροχιάς	9
1.3 Σχολιασμός Επεισοδίου.....	12
1.4 Σχολιασμός Τροχιάς.....	13
1.5 Σχετικές Εργασίες.....	14
ΚΕΦΑΛΑΙΟ 2.....	18
ΘΕΩΡΗΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ ΜΕΘΟΔΟΥ	18
2.1 Συσταδοποίηση.....	18
2.1.1 Αλγόριθμος Συσταδοποίησης.....	19
2.2 Απόσταση- Ομοιότητα Συστάδων.....	19
2.3 Είδη Συσταδοποίησης.....	19
2.4 Ιεραρχική Συσταδοποίηση.....	21
2.4.1 Συσσωρευτικοί Ιεραρχικοί Αλγόριθμοι.....	22
2.4.2 Κριτήρια καθορισμού της απόστασης μεταξύ των συστάδων	24
2.5 Complete linkage/Πλήρης Σύνδεση	25
2.6 Δεδομένα	27
2.7 Βήματα Μεθόδου	29
ΚΕΦΑΛΑΙΟ 3	31
ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	31
3.1 Συσταδοποίηση epis_id- epis_def_tag	31
3.2 Συσταδοποίηση epis_id- sem_traj_id.....	35
3.3 Συσταδοποίηση sem_traj_id- avg_speed.....	39
ΚΕΦΑΛΑΙΟ 4.....	44
ΣΥΜΠΕΡΑΣΜΑΤΑ.....	44
ΠΑΡΑΡΤΗΜΑ ΠΙΝΑΚΩΝ	46

Πανεπιστήμιο Πειραιώς

Κατάλογος Πινάκων

3-1	Βήματα Συγχώνευσης των Συστάδων (epis_id- epis_def_tag)	32
3-2	Δημιουργία Συστάδων (epis_id- epis_def_tag)	33
3-3	Κέντρα Συστάδων (epis_id- epis_def_tag)	33
3-4	Αποστάσεις μεταξύ των κέντρων των συστάδων (epis_id- epis_def_tag)	34
3-5	Βήματα Συγχώνευσης των Συστάδων (epis_id- sem-traj_id)	36
3-6	Δημιουργία Συστάδων (epis_id- sem-traj_id)	37
3-7	Κέντρα Συστάδων (epis_id- sem-traj_id)	37
3-8	Αποστάσεις μεταξύ των κέντρων των συστάδων (epis_id- sem-traj_id)	38
3-9	Βήματα Συγχώνευσης των Συστάδων (sem_traj_id- avg_speed)	40
3-10	Δημιουργία Συστάδων (sem_traj_id- avg_speed)	41
3-11	Κέντρα Συστάδων (sem_traj_id- avg_speed)	41
3-12	Αποστάσεις μεταξύ των κέντρων των συστάδων (sem_traj_id- avg_speed)	42

Πανεπιστήμιο Πειραιώς

Κατάλογος Σχημάτων

1-1	Οπτικοποίηση σε ένα φόντο ενός 2D χάρτη από το ίχνος που άφησε ένας τουρίστας κατά τη διάρκεια μιας μέρας στο Παρίσι.	7
1-2	Ένα χρονο-γεωγραφικό διάγραμμα [Hagerstrand 1970] το οποίο δείχνει μέρος της διαδρομής που ακολούθησε ο τουρίστας νωρίτερα.	10
1-3	Μία σημασιολογική τροχιά ενός τουρίστα εμπλουτισμένη με σχολιασμούς stops και moves	13
2-1	Γραφική απεικόνιση αλγορίθμου συσταδοποίησης	19
2-2	Δενδρόγραμμα Ιεραρχικής Συσταδοποίησης	22
2-3	Αλγόριθμος Ιεραρχικής Συσταδοποίησης	23
2-4	Δενδρόγραμμα ιεραρχικής συσσωρευτικής συσταδοποίησης	24
2-5	Complete-link: ελάχιστη ομοιότητα	25
2-6	Ένα σύνολο από έξι διδιάστα σημεία	26
2-7	Complete linkage clustering	26
2-8	Complete linkage dendrogram	26
3-1	Δενδρόγραμμα (epis_id- epis_def_tag)	35
3-2	Δενδρόγραμμα (epis_id- sem_traj_id)	39
3-3	Δενδρόγραμμα (sem_traj_id- avg_speed)	43

Πανεπιστήμιο Πειραιώς

ΕΙΣΑΓΩΓΗ

Στη σημερινή εποχή οι κινητές συσκευές είναι ευρέως διαθέσιμες και η τεράστια αύξηση της χρήσης τους παράγει ένα τεράστιο ποσό από δεδομένα κίνησης. Η παρουσία των κινητών τηλεφώνων, των ασύρματων δικτύων αισθητήρων, συσκευών που είναι εξοπλισμένες με Global Positioning Systems (GPS) και Radio- Frequency Identification (RFID), προκαλεί μία εκρηκτική ανάπτυξη των γεω- χωρικών δεδομένων που επιτρέπει την παρακολούθηση των μετακινήσεων, με χαμηλό κόστος, με συνεχή «συλλογή» των σημείων σε μία χρονική ακολουθία. Αυτός ο καταγισμός δεδομένων έχει οδηγήσει στην εμφάνιση μίας σειράς από αλγορίθμους και αναλύσεις με σκοπό την καλύτερη αντίληψη, τη μοντελοποίηση και την αξιοποίηση αυτών των δεδομένων. Τα αποτελέσματα των μελετών αυτών θα δώσουν μία εικόνα για την ανθρώπινη συμπεριφορά από την ανίχνευση της κίνησης τους, θα ανακαλύψουν τα ήθη, τις συνήθειες και τα ενδιαφέροντα τους. Οι πληροφορίες και οι γνώσεις σχετικά με τον τρόπο και τον λόγο που οι άνθρωποι κινούνται είναι απαραίτητες για τους επιστήμονες, τους μελετητές και τους υπευθύνους για τη λήψη αποφάσεων σε διαφορετικά πεδία όπως η διαχείριση της κυκλοφορίας, η μηχανική μάθηση, οι κοινωνικές επιστήμες και άλλα (Giannotti and Pedreschi, 2008).

Πολλές εφαρμογές δεν ενδιαφέρονται για τη διατήρηση και την ανάλυση όλων των εγγραφών της κίνησης ενός κινούμενου αντικειμένου. Για το λόγο αυτό επιλέγουν εκείνα τα τμήματα που παρουσιάζουν ενδιαφέρον για την ανάλυση. Τα τμήματα των διαδρομών των κινούμενων αντικειμένων που παρουσιάζουν ενδιαφέρον για τις δεδομένες εφαρμογές ονομάζονται τροχιές (trajectories), οι οποίες προσδιορίζονται από δύο ειδικές χωροχρονικές θέσεις κινούμενων αντικειμένων που καλούνται «Αρχή» και «Τέλος» της πορείας (Spaccapietra et all 2008).

Τα τελευταία χρόνια οι τεχνικές που εφαρμόζονται στη μελέτη της κίνησης έχουν επεκταθεί σε σημαντικό βαθμό. Πιο συγκεκριμένα μία καινούργια υποσχόμενη προσέγγιση έχει σχεδιαστεί για να παρέχει εφαρμογές οι οποίες εμπεριέχουν πιο πλούσια και πιο ουσιώδη γνώση σχετικά με την κίνηση των αντικειμένων. Αυτό έχει επιτευχθεί συνδυάζοντας τις ακατέργαστες διαδρομές της κίνησης (δηλαδή για παράδειγμα οι εγγραφές που συλλέγονται από τις συσκευές εντοπισμού θέσης) με συναφή συμφραζόμενα δεδομένα (contextual data).

Αυτές οι εμπλουτισμένες εγγραφές διαδρομής είναι γνωστές ως σημασιολογικές τροχιές (semantic trajectories).

Μία σημασιολογική τροχιά είναι μία τροχιά που έχει σχολιαστεί σε διαφορετικά επίπεδα όπως στο σύνολο της τροχιάς, στα επεισόδια (ουσιώδη τμήματα της τροχιάς) και στις θέσεις. Αυτό που παρουσιάζει ενδιαφέρον είναι ο σχολιασμός σε επίπεδο επεισοδίου, δηλαδή ο διαχωρισμός της διαδρομής ενός αντικειμένου σε χρονικές περιόδους που θεωρείται στάσιμο (stop episodes) και σε χρονικές περιόδους που κινείται (move episodes).

Η χωρο-χρονική συσταδοποίηση (clustering) είναι μια διαδικασία ομαδοποίησης αντικειμένων με βάση τη χωρική και χρονική τους ομοιότητα. Είναι ένα «υποπεδίο» της εξόρυξης δεδομένων (data mining) το οποίο έχει αποκτήσει μεγάλη δημοτικότητα, κυρίως στις γεωγραφικές επιστήμες πληροφόρησης, λόγω της μεγάλης διάδοσης όλων των ειδών των location – based συσκευών που καταγράφουν τη θέση, το χρόνο ή/ και τις περιβαλλοντικές ιδιότητες ενός αντικειμένου ή ενός συνόλου αντικειμένων σε πραγματικό χρόνο.

Πιο συγκεκριμένα η cluster analysis ή clustering είναι η ταξινόμηση ενός συνόλου αντικειμένων σε ομάδες που ονομάζονται clusters. Η συσταδοποίηση ομαδοποιεί τα αντικείμενα (τις παρατηρήσεις, τις μεταβλητές) με βάση τις πληροφορίες που περιέχονται στα δεδομένα που περιγράφουν τα αντικείμενα ή τις σχέσεις τους. Στόχος είναι αυτά που ανήκουν στην ίδια ομάδα να είναι παρόμοια (ή συναφή) το ένα με το άλλο και διαφορετικά (ή δεν σχετίζονται) με τα αντικείμενα άλλων ομάδων. Όσο μεγαλύτερη είναι η ομοιότητα (ή ομοιογένεια) σε μία ομάδα και όσο μεγαλύτερη είναι η διαφορά μεταξύ των ομάδων, τόσο καλύτερη είναι η ανάλυση της συσταδοποίησης. Η συσταδοποίηση είναι μία ταξινόμηση των αντικειμένων από τα δεδομένα, όπου με την έννοια ταξινόμηση εννοούμε μία επισήμανση (labeling) των αντικειμένων με ετικέτες κατηγορίας. Ως εκ τούτου η συσταδοποίηση μερικές φορές αναφέρεται και ως «ταξινόμηση χωρίς επίβλεψη» και διακρίνεται από την «εποπτευόμενη ταξινόμηση» ή αλλιώς classification, η οποία επιδιώκει να βρει κανόνες για να ταξινομήσει αντικείμενα από ένα σύνολο προ-ταξινομημένων αντικειμένων.

Οι αλγόριθμοι συσταδοποίησης διακρίνονται σε δύο βασικές κατηγορίες:

- i. Με βάση την ιεράρχηση (Hierarchical clustering)
- ii. Με βάση τον επιμερισμό ή διαχωρισμό (Partitional clustering)

Στη διαμεριστική συσταδοποίηση, οι αλγόριθμοι χωρίζουν ένα σύνολο D που αποτελείται από N αντικείμενα σε ένα σύνολο K συστάδων. Το K είναι μία παράμετρος εισόδου που επιλέγεται από τον χρήστη ανάλογα με τις γνώσεις που έχει όσον αφορά τα δεδομένα του. Οι

αλγόριθμοι αρχίζουν με μία αρχική κατάτμηση των δεδομένων και εφαρμόζουν μία επαναληπτική στρατηγική ελέγχου για τη βελτιστοποίηση των αποτελεσμάτων της αντικειμενικής συνάρτησης για κάθε συστάδα. Η υψηλή πολυπλοκότητα και η μη ικανότητα να παράγει αυθαίρετης μορφής συστάδες είναι δύο βασικά μειονεκτήματα αυτής της μεθόδου.

Στην ιεραρχική συσταδοποίηση, οι αλγόριθμοι δημιουργούν μία ιεραρχική «αποσύνθεση» των αντικειμένων ενός συνόλου δεδομένων D , η οποία αντιπροσωπεύεται από ένα δενδρόγραμμα. Το δενδρόγραμμα χωρίζει το σύνολο D σε μικρότερα υποσύνολα με μία επαναληπτική προσέγγιση, η οποία μπορεί να είναι είτε συσσωρευτική (agglomerative) είτε διαιρετική (divisive).

Η συσσωρευτική προσέγγιση ξεκινάει λαμβάνοντας κάθε αντικείμενο ως μία ξεχωριστή συστάδα και στη συνέχεια σε κάθε βήμα συγχωνεύει το πιο κοντινό ζευγάρι συστάδων με βάση ενός μέτρου απόστασης. Η συσταδοποίηση θα σταματήσει μέχρι να μείνει μόνο μία συστάδα (ή k συστάδες) ή όταν το επιθυμεί ο χρήστης.

Αντίθετα στη διαιρετική προσέγγιση ο αλγόριθμος ξεκινάει με μία συστάδα που περιέχει όλα τα αντικείμενα και σε κάθε βήμα διαχωρίζει μία συστάδα σε μικρότερες έως ότου κάθε συστάδα να περιέχει ένα μόνο σημείο.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η μελέτη, μέσω στοχαστικών διαδικασιών, των πιθανών αλληλεπιδράσεων που μπορεί να έχουν μεταξύ τους οι μεταβλητές που θέλουμε να μελετήσουμε. Συγκεκριμένα τρέξαμε τον στοχαστικό ιεραρχικό αλγόριθμο συσταδοποίησης συσσωρευτικά.

Καταλήξαμε στην επιλογή αυτής της μεθόδου καθώς οι αλγόριθμοι της ιεραρχικής προσέγγισης συγκριτικά με τους αλγόριθμους της διαμεριστικής προσέγγισης, και συγκεκριμένα με τον αλγόριθμο K means, δεν απαιτούν τον ορισμό του αριθμού των συστάδων, οι αποφάσεις σε κάθε βήμα δεν αναιρούνται, μπορούν να συνδυαστούν με τον k means και επιπλέον υπάρχουν μελέτες που υποστηρίζουν ότι παράγουν καλύτερες συστάδες.

Η δομή της παρούσας διπλωματικής εργασίας είναι η ακόλουθη:

Στο πρώτο κεφάλαιο παρουσιάζονται οι βασικές έννοιες της τροχιάς και της σημασιολογικής τροχιάς καθώς επίσης γίνεται αναφορά σε διάφορες εργασίες που αφορούν τη συσταδοποίηση δεδομένων κινούμενων αντικειμένων.

Στο δεύτερο κεφάλαιο δίνεται μία συνοπτική εικόνα των εννοιών της συσταδοποίησης, της ιεραρχικής συσταδοποίησης καθώς επίσης και περιγραφή της μεθόδου που χρησιμοποιήσαμε. Επιπλέον παρουσιάζεται το σύνολο των δεδομένων που χρησιμοποιήθηκε.

Στο τρίτο κεφάλαιο παρουσιάζονται τα αποτελέσματα από την εφαρμογή της μεθόδου μας μαζί με την ανάλυση τους και τέλος στο τέταρτο κεφάλαιο συνοψίζονται τα συμπεράσματα που έχουν προκύψει.

Πανεπιστήμιο Πειραιώς

ΚΕΦΑΛΑΙΟ 1

Η ΕΝΝΟΙΑ ΤΗΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ ΣΤΙΣ ΣΗΜΑΣΙΟΛΟΓΙΚΕΣ ΤΡΟΧΙΕΣ

Τα χωρο-χρονικά, γεωγραφικά σύνολα δεδομένων αυξάνονται με γρήγορους ρυθμούς και θα αυξηθούν ακόμα περισσότερο στο εγγύς μέλλον. Αυτό το φαινόμενο οφείλεται στην καθημερινή συλλογή και τη μεταφορά των δεδομένων μέσω των συστημάτων των βάσεων δεδομένων, των ελεγκτών του δικτύου κίνησης, των web servers, των δικτύων αισθητήρων. Στην προοπτική αυτή συμβάλλει μία σημαντική πηγή δεδομένων που είναι τα τηλεπικοινωνιακά δεδομένα που συλλέγονται από τα κινητά τηλέφωνα και από άλλες συσκευές εντοπισμού θέσης, δεδομένα που προκύπτουν από την ανάγκη εντοπισμού μέσω των ασύρματων φορητών συσκευών προκειμένου να υποστηριχθεί η αλληλεπίδραση τους με την υποδομή του δικτύου. Αυτή η μεγάλη διαθεσιμότητα αυτών των μορφών πληροφοριών αναμένεται να επιτρέψει νέες τάξεις εφαρμογών, όπου η ανακάλυψη της γνώσης θα είναι το βασικό βήμα. Χαρακτηριστικό παράδειγμα είναι η παρουσία ενός μεγάλου αριθμού ασύρματων συσκευών εντοπισμού θέσης η οποία δίνει μία αυξανόμενη δυνατότητα πρόσβασης σε αρχεία καταγραφής και ανακατασκευής των χωρο-χρονικών τροχιών αυτών των προσωπικών συσκευών και των ατόμων που τις κατέχουν: οι τροχιές είναι τα ίχνη των κινούμενων αντικειμένων και των ιδιωτών τους. Αυτές οι τροχιές περιέχουν λεπτομερείς πληροφορίες σχετικές με την συμπεριφορά των κινούμενων αντικειμένων και ως εκ τούτου προσφέρουν ενδιαφέρουσες πρακτικές για να βρεθούν πρότυπα συμπεριφοράς που μπορούν να χρησιμοποιηθούν για παράδειγμα στη βιώσιμη διαχείριση της κινητικότητας και της κυκλοφορίας.

Στο παρόν κεφάλαιο αρχικά γίνεται αναφορά στις βασικές έννοιες των τροχιών των κινούμενων αντικειμένων, στον σημασιολογικό τους εμπλουτισμό, στην έννοια της σημασιολογικής τροχιάς καθώς και στις διαφορές εργασίες- έρευνες που έχουν γίνει σχετικά με τη συσταδοποίηση τροχιών και αντικειμένων.

1.1 Βασικές έννοιες της τροχιάς ενός κινούμενου αντικειμένου

Η μείωση του κόστους των κινητών συσκευών όπως είναι τα GPS, τα κινητά τηλέφωνα και τα δίκτυα αισθητήρων, αύξησε σημαντικά την παραγωγή χωρο-χρονικών δεδομένων, τα οποία ονομάζονται τροχιές κινούμενων αντικειμένων. Αυτές οι συσκευές παράγουν ακατέργαστα δεδομένα που περιέχουν αναγνωριστικά χαρακτηριστικά των αντικειμένων καθώς επίσης και τις θέσεις στις οποίες βρίσκονται σε μια δεδομένη χρονική στιγμή. Αυτά τα δεδομένα είναι γνωστά ως πρωτογενείς τροχιές.

Επομένως η έννοια της τροχιάς προέρχεται από την καταγραφή της κίνησης ενός αντικειμένου στον γεωγραφικό χώρο κατά την πάροδο του χρόνου και έχει ως αποτέλεσμα την παραγωγή της τροχιάς κίνησης κάθε κινούμενου αντικειμένου.

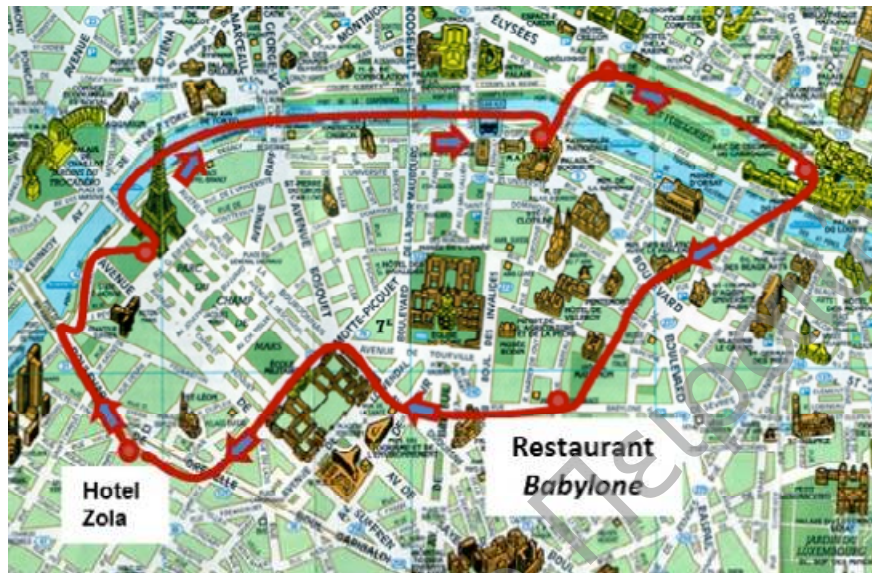
Η τροχιά της κίνησης συνιστάται κυρίως στη χρονική ακολουθία των χωροχρονικών θέσεων του αντικειμένου κατά τη διάρκεια της κίνησής του. Ωστόσο ανάλογα με τις δυνατότητες που έχει μία συσκευή πρόσθετα στοιχεία όπως είναι η στιγμιαία ταχύτητα ή η ακινησία, η επιτάχυνση, η κατεύθυνση και η περιστροφή μπορούν να συμπληρώσουν τις χωροχρονικές θέσεις του αντικειμένου. Τα δεδομένα αυτά που καταγράφονται από τη συσκευή ονομάζονται *ακατέργαστα δεδομένα*.

Οι ακατέργαστες διαδρομές κίνησης μπορούν να χρησιμοποιηθούν είτε για περαιτέρω ανάλυση είτε για να μετατραπούν σε άλλα είδη αναπαράστασης της κίνησης. Για παράδειγμα μία εφαρμογή η οποία μελετάει τον τρόπο με τον οποίο η κίνηση του αντικειμένου εξαπλώνεται σε μία δεδομένη περιοχή, επιλέγει μία ακατέργαστη διαδρομή κίνησης προκειμένου να υπάρχει μία συνεχής άποψη του χώρου κίνησης και να την καταγράψει ως ένα πεδίο διανυσμάτων. Τα διανύσματα αυτά συγκεντρώνουν στοιχεία από τις μεμονωμένες τροχιές ώστε να παρουσιάσουν, για μία δεδομένη χρονική στιγμή, χαρακτηριστικά της κίνησης για κάθε θέση στο χώρο.

Πολλές εφαρμογές, προκειμένου να αναλύσουν ένα μεγάλο όγκο εγγραφών, επιλέγουν εκείνα τα τμήματα της κίνησης του αντικειμένου που έχουν ενδιαφέρον, τα οποία ονομάζονται *τροχιές* (trajectories), και τα οποία μελετώνται για ένα συγκεκριμένο χρονικό διάστημα.

Για παράδειγμα, στη μελέτη των δραστηριοτήτων που πραγματοποιούνται από ένα τουρίστα κατά τη διάρκεια της παραμονής του στο Παρίσι, θεωρούμε ως μία ενιαία τροχιά ολόκληρο το κομμάτι της κίνησης που περιλαμβάνει αυτές τις δραστηριότητες. Προκειμένου

να μελετηθούν οι δραστηριότητες του τουρίστα κατά τη διάρκεια μίας ημέρας του στο Παρίσι ή συγκεκριμένων ημερών, κάθε καθημερινή του διαδρομή θεωρείται ως μία ξεχωριστή τροχιά, όπως φαίνεται στο Σχήμα 1.1 που ακολουθεί.



Σχήμα 1-1: Οπτικοποίηση σε ένα φόντο ενός 2D χάρτη από το ίχνος που άφησε ένας τουρίστας κατά τη διάρκεια μιας μέρας στο Παρίσι.

1.1.1 Η έννοια της ακατέργαστης τροχιάς

Στην εργασία των Spaccapietra et al. (2008) κάθε μία τροχιά προσδιορίζεται από δύο συγκεκριμένες χωροχρονικές θέσεις της κίνησης, οι οποίες ονομάζονται Αρχή (Begin) και Τέλος (End) της κάθε τροχιάς και αποτελούν την πρώτη και την τελευταία θέση του κινούμενου αντικειμένου. Πιο αναλυτικά σύμφωνα με τους συγγραφείς η τροχιά αποτελεί μία γενική χωροχρονική έννοια. Η κίνηση θεωρείται ως ένα χαρακτηριστικό κάποιων αντικειμένων το οποίο τα διαφοροποιεί από τα αντίστοιχα ακίνητα και η γενική ιδέα της υπονοεί ότι η κίνηση προορίζεται να εκπληρώσει έναν ουσιαστικό στόχο ο οποίος απαιτεί τη μεταφορά από τη μία θέση σε μία άλλη. Αυτή η μεταφορά αξιώνει ένα πεπερασμένο χρονικό διάστημα με βάση το οποίο ορίζονται οι τροχιές. Αυτό το χρονικό διάστημα οριοθετείται στιγμιαία όταν το αντικείμενο ξεκινάει την κίνησή του t_{begin} και όταν την τερματίζει t_{end} . Η

ταυτοποίηση του t_{begin} και t_{end} με το χρονικό πλαίσιο στο οποίο το αντικείμενο κινείται είναι μία απόφαση η οποία λαμβάνεται από την εφαρμογή.

Ως ακατέργαστη τροχιά ορίζεται μία τροχιά η οποία προέρχεται από τις αρχικές διαδρομές του κινούμενου αντικειμένου και περιέχει μόνο πρωτογενή δεδομένα για το διάστημα Αρχή-Τέλος.

1.1.2 Από τις ακατέργαστες τροχιές στις σημασιολογικές τροχιές

Οι αρχικές τροχιές προσαρμόζονται σε εφαρμογές που έχουν ως στόχο τον εντοπισμό ενός κινούμενου αντικειμένου ή στην εφαρμογή στατιστικής ανάλυσης πάνω στα χωροχρονικά χαρακτηριστικά των τροχιών.

Ωστόσο οι περισσότερες εφαρμογές ανάλυσης απαιτούν τον εμπλουτισμό των αρχικών δεδομένων με πρόσθετες πληροφορίες στο πλαίσιο της εφαρμογής. Για παράδειγμα, η ερμηνεία των τροχιών των ανθρώπων που κινούνται μέσα σε μία πόλη απαιτεί κάποια γνώση των χαρακτηριστικών της πόλης όπως είναι ο χάρτης της και κάποια σημεία ενδιαφέροντος. Βάση των πληροφοριών που έχουμε για αυτήν οι χωροχρονικές συντεταγμένες μπορούν να αντικατασταθούν με ονόματα δρόμων ή με τα ονόματα των πολυσύχναστων τοποθεσιών όπως είναι τα καταστήματα, τα εστιατόρια και τα μουσεία.

Ο σημασιολογικός εμπλουτισμός αποτελεί μία διαδικασία προσθήκης σχετικών δεδομένων του πλαισίου εφαρμογής στις αρχικές τροχιές. Είναι το κλειδί για την υποστήριξη της κινητικότητας και της ανάλυσης συμπεριφοράς που παρουσιάζουν ενδιαφέρον σε μία εφαρμογή. Η βιβλιογραφία πάνω στην κινητικότητα είναι πλούσια σε προτάσεις και τεχνικές σημασιολογικού εμπλουτισμού όπου η καθεμία έχει βασιστεί σε κάποιες δημοφιλείς εφαρμογές και έχει εστιάσει στην απόκτηση γνώσεων ειδικά σε αυτόν τον τομέα. Η ανθρώπινη αστική κινητικότητα είναι ένας από τους τομείς που προκαλεί μεγάλη προσοχή εξαιτίας της δυνατότητας που υπάρχει να συλλέγονται δεδομένα, από συσκευές GPS που έχουν εγκατασταθεί σε οχήματα και σε προσωπικά κινητά τηλέφωνα, με ιδιαίτερη ευκολία. Επιπλέον, οι ανησυχίες του κοινού για την πόλη και την διαχείριση της κυκλοφορίας έχουν παρακινήσει, σε μεγάλο βαθμό, την έρευνα σε αυτόν τον τομέα.

Μία τυπική διαδικασία σημασιολογικού εμπλουτισμού της τροχιάς έχει ως είσοδο ένα σύνολο από αρχικές, θορυβώδεις τροχιές (sound trajectories), μία σχετική αποθήκη

δεδομένων και παράγει ως έξοδο ένα σύνολο από σημασιολογικά σχολιασμένες τροχιές ή πιο απλά σημασιολογικές τροχιές.

Σημασιολογική Τροχιά (Semantic Trajectory) ορίζεται ως μία τροχιά η οποία έχει ενισχυθεί με σχολιασμούς και/ ή ένα ή περισσότερα συμπληρωματικά τμήματα.

1.2 Τμηματοποίηση Τροχιάς

Οι τροχιές μπορούν να σχολιαστούν σε τρία διαφορετικά επίπεδα:

- i. Τροχιές
- ii. Επεισόδια, δηλαδή ομοιογενή τμήματα που ορίζονται ως η μέγιστη υποακολουθία της τροχιάς με τέτοιο τρόπο ώστε όλες οι χωροχρονικές θέσεις να συμμορφώνονται με ένα συγκεκριμένο κατηγορημα
- iii. Θέσεις

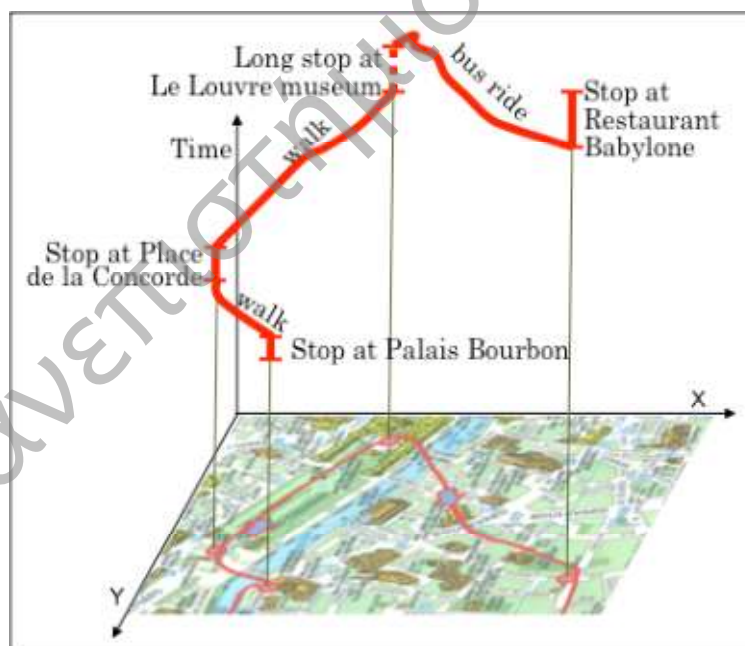
Ο σχολιασμός των θέσεων μιας τροχιάς δεν είναι αποτελεσματικός καθώς μπορεί να δημιουργήσει ένα μεγάλο αριθμό από επαναλαμβανόμενους σχολιασμούς. Ως εκ τούτου, ο σχολιασμός είναι καλύτερα να πραγματοποιείται στα επεισόδια και στα επίπεδα της τροχιάς. Ο σχολιασμός των επεισοδίων αρχικά απαιτεί τον προσδιορισμό των επεισοδίων μέσα στις τροχιές. Η διαδικασία η οποία αναγνωρίζει τα επεισόδια ονομάζεται *τμηματοποίηση τροχιάς* (trajectory segmentation).

Η τμηματοποίηση μίας τροχιάς καθορίζεται από εξαρτώμενα κριτήρια εφαρμογής. Ένας τρόπος τμηματοποίησης μιας τροχιάς που δεν περιέχει επεισόδια αλλά πηγαινεί απευθείας από την Αρχή στο Τέλος είναι ο χωρισμός της διαδρομής σε χρονικές περιόδους όπου το αντικείμενο θεωρείται ακίνητο, οι περίοδοι αυτοί ονομάζονται stops, και σε περιόδους όπου το αντικείμενο κινείται, δηλαδή τα moves. Σύμφωνα με αυτή την οπτική, μία τροχιά είναι μία ακολουθία από εναλλασσόμενα stop και moves επεισόδια, πάντα ξεκινώντας με ένα move και έχοντας κανένα ή περισσότερα stop. Στόχος των ερευνητών είναι να επεξεργαστούν τρόπους εντοπισμού των stop.

Ο χαρακτηρισμός ενός stop και move επεισοδίου, εξαρτάται από τις απαιτήσεις των εφαρμογών, όπως άλλωστε συμβαίνει και για κάθε τμηματοποίηση μίας τροχιάς.

Για παράδειγμα, σε μία εφαρμογή που μελετά τη συμπεριφορά των τουριστών όταν δεν υπάρχει καθόλου κίνηση για κάποιο ελάχιστο χρονικό διάστημα, ένα επεισόδιο χαρακτηρίζεται ως stop όταν το κινούμενο αντικείμενο κοιμάται, τρώει ή παρακολουθεί μία θεατρική παράσταση ενώ αντίθετα όταν υπάρχει κίνηση μέσα σε μία μικρή περιοχή και για ένα μικρό χρονικό διάστημα, το επεισόδιο χαρακτηρίζεται ως stop όταν το αντικείμενο επισκεφθεί μία έκθεση.

Στο Σχήμα 1-2 τα stop έχουν βρεθεί εντοπίζοντας την αργή κίνηση σε μία μικρή περιοχή η οποία αποτελεί ένα σημείο ενδιαφέροντος του τουρίστα (POI). Τα τρία κάθετα τμήματα που φαίνονται στο ίχνος της τροχιάς, αντιστοιχούν στις στάσεις που πραγματοποιεί ο τουρίστας μέσα σε ένα χρονικό διάστημα ανάλογο με το μήκος του τμήματος. Οι στάσεις οι οποίες εντοπίστηκαν είναι οι εξής: Πύργος του Άιφελ, Palais Bourbon, Place de la Concorde, Le Louvre Museum, Restaurant Babylone. Ο χαρακτηρισμός ενός επεισοδίου stop μπορεί να υποδηλώνει καθόλου κίνηση, χαμηλή ταχύτητα, κίνηση μέσα σε μία περιορισμένη περιοχή ή κοντά σε κάποια σημεία ενδιαφέροντος.



Σχήμα 1-2 : Ένα χρονο-γεωγραφικό διάγραμμα [Hagerstrand 1970] το οποίο δείχνει μέρη της διαδρομής που ακολούθησε ο τουρίστας νωρίτερα.

Αρχικά η ταυτοποίηση των stop έχει εφαρμοστεί στις ακατέργαστες τροχιές. Μία πρώτη προσέγγιση απλά συνδέει τα stop με τα πρωτογενή τμήματα μιας τροχιάς ενός αυτοκινήτου που είτε δεν υπάρχουν λόγω απώλειας ενός σήματος (GPS είναι απενεργοποιημένο ή το αυτοκίνητο μπήκε σε γκαράζ) είτε η ταχύτητα είναι μηδέν για ένα δεδομένο χρονικό διάστημα (το αυτοκίνητο είναι σταθμευμένο) (Ashbrook et al. 2003). Ο περιορισμός αυτής της προσέγγισης έγκειται στην δυνατότητα ότι μία πραγματική στάση δεν μπορεί να προσδιοριστεί εξαιτίας των σφαλμάτων κίνησης και της ταχύτητας η οποία είναι ελαφρώς πάνω από το μηδέν.

Μεταγενέστερες προσεγγίσεις χρησιμοποιούν κάπως πιο εξελιγμένες τεχνικές. Ένα παράδειγμα δίνεται από τα κριτήρια που χρησιμοποιούνται στους (Krumm et al. 2006) προκειμένου να χωριστεί ένα κομμάτι της κίνησης ενός αυτοκινήτου σε τροχιές. Τα συγκεκριμένα κριτήρια έχουν τη δυνατότητα να εντοπίσουν παρόμοιες στάσεις μέσα σε μία τροχιά:

- i. Ένα πεντάλεπτο κενό υποδεικνύει ότι το αυτοκίνητο δεν κινείται,
- ii. Τουλάχιστον πέντε λεπτά με πολύ χαμηλή ταχύτητα γεγονός που υποδηλώνει ότι το αυτοκίνητο είναι ακίνητο ενώ το GPS στέλνει σήματα δίνοντας την ίδια θέση προσεγγιστικά λόγω του θορύβου.

Η προσέγγιση των (Andrienko and Andrienko 2007a) χρησιμοποιεί ένα μεγαλύτερο χρονικά κατώτερο όριο (δύο ωρών) προκειμένου να αναγνωρίσει τα σημαντικά μέρη που έχει επισκεφθεί ένα κινούμενο άτομο. Η υπόθεση είναι ότι όσος περισσότερος χρόνος δαπανάται σε ένα μέρος, τόσο πιο σημαντικό είναι το μέρος αυτό για το κινούμενο άτομο. Μία παρόμοια υπόθεση χρησιμοποιείται στους (Zheng et al. 2011), στην οποία η στάσεις υπολογίζονται ως ακολουθίες διαδοχικών θέσεων GPS, έτσι ώστε η χωρική τους απόσταση να είναι κάτω από ένα κατώτερο όριο ενώ η χρονική διάρκεια πάνω από ένα άλλο κατώτερο όριο. Ο αλγόριθμος CB-SMoT (Palma et al. 2008) χρησιμοποιεί μία άλλη προσέγγιση, στην οποία τα stop επεισόδια υπολογίζονται βάση της μεταβολής της ταχύτητας της τροχιάς. Τα stop είναι εκείνα τα τμήματα της τροχιάς στα οποία η ταχύτητα είναι χαμηλότερη από τη μέση ταχύτητα της τροχιάς.

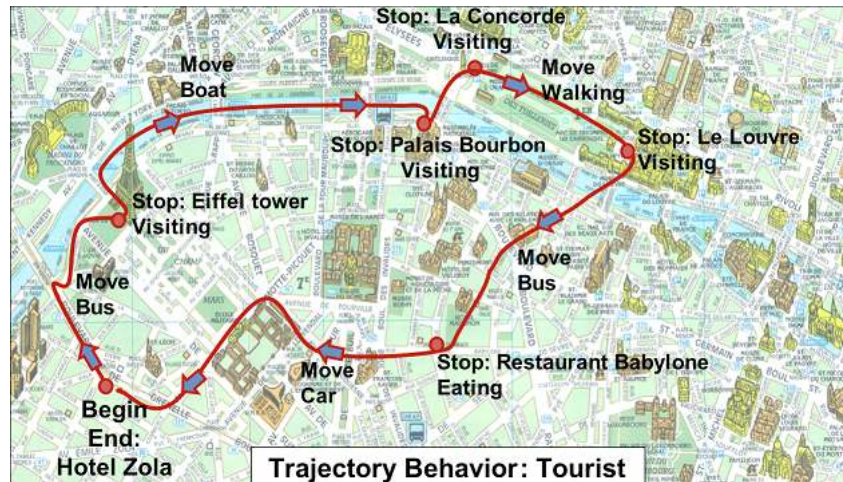
Ένα άλλο κριτήριο για την αναγνώριση ενός stop είναι η αλλαγή της κατεύθυνσης (Rocha et al. 2010), όπως συμβαίνει στη μελέτη της κίνησης ενός αλιευτικού σκάφους όπου μία

στάση του χαρακτηρίζεται από την ξαφνική αλλαγή της κατεύθυνσης προκειμένου να κινηθεί στην περιοχή αλίευσης.

Τέλος οι τροχιές μπορούν να τμηματοποιηθούν και με άλλα κριτήρια, εκτός από τα stop και τα moves. Για παράδειγμα, στην έρευνα (Xie et al. 2009) οι συγγραφείς χρησιμοποιούν ένα ενισχυμένο γεωγραφικό πλαίσιο προκειμένου να χωρίσουν την τροχιά με βάση το είδος του σημείου ενδιαφέροντος. Ένα επεισόδιο καθορίζεται ως το τμήμα της τροχιάς του οποίου οι θέσεις επηρεάζονται από ένα δοσμένο POI. Δηλαδή ένα POI επηρεάζει μία θέση p της τροχιάς εάν η απόσταση μεταξύ του p και του POI είναι μικρότερη από την απόσταση ανάμεσα στο p και σε ένα άλλο POI. Αξίζει να αναφερθεί ότι κάθε επεισόδιο σχολιάζεται από το αντίστοιχο POI.

1.3 Σχολιασμός Επεισοδίου

Η διαδικασία του σημασιολογικού εμπλουτισμού (semantic enrichment) εκφράζει την αντίληψη ότι τα υπάρχοντα δεδομένα συμπληρώνονται με πρόσθετα στοιχεία τα οποία ονομάζονται *σχολιασμοί* (annotations). Κάθε επεισόδιο χαρακτηρίζεται με ένα δικό του σχολιασμό. Για παράδειγμα τα επεισόδια στάσεις και κίνησης σχολιάζονται ως stop και move επεισόδια αντίστοιχα, τα επεισόδια που σχετίζονται με τον τρόπο μεταφοράς σχολιάζονται ως πόδια, λεωφορείο ή αυτοκίνητο ενώ τα επεισόδια που σχετίζονται με POI σχολιάζονται με τον κωδικό αναγνώρισης του αντίστοιχου POI. Ωστόσο, τα επεισόδια μπορούν επίσης να φέρουν και άλλους σχολιασμούς. Ένα συχνό παράδειγμα είναι στα stop επεισόδια, τα οποία μπορούν να σχολιαστούν με σχετικές πληροφορίες, όπως το αναγνωριστικό του πιθανού POI στο οποίο θα σταματήσει το αντικείμενο ή τον τύπο του (για παράδειγμα σπίτι, δουλειά εστιατόριο) ή κάποιους άλλους ειδικούς σχολιασμούς (όπως οι δραστηριότητες που γίνονται κατά τη διάρκεια της στάσης). Στο σχήμα 1.3 που ακολουθεί στα επεισόδια στάσης αντιστοιχούν οι σχολιασμοί POI και δραστηριότητα και στα επεισόδια κίνησης το μέσο μεταφοράς.



Σχήμα 1-3: Μία σημασιολογική τροχιά ενός τουρίστα εμπλουτισμένη με σχολιασμούς stops και moves

1.4 Σχολιασμός Τροχιάς

Ο σχολιασμός της τροχιάς στο σύνολο της υποθέτει ότι οι πληροφορίες της τροχιάς συνθέτουν μία ενιαία ετικέτα που χαρακτηρίζει όλη την τροχιά. Για παράδειγμα στο Σχήμα 1.1 η τροχιά έχει σχολιαστεί με ένα σχολιασμό Συμπεριφοράς του οποίου η τιμή «Τουρίστας» προέκυψε από το είδος της δραστηριότητας των στάσεων του.

Η εργασία των Spinsanti et al. (2010) προτείνει μία σειρά από κανόνες για το σχολιασμό των τροχιών με τη συνολική δραστηριότητα που εκφράζεται από τα δεδομένα τροχιάς. Λαμβάνοντας υπόψη μία τροχιά της οποίας τα επεισόδια στάσης σχολιάζονται με πιθανές δραστηριότητες και την κοινή αίσθηση των κανόνων IF_THEN, σχολιάζουν την τροχιά με την πιο πιθανή παγκόσμια δραστηριότητα. Για παράδειγμα εάν η τροχιά σχολιάζεται με τις στάσεις σπίτι και δουλειά, τότε ολόκληρη η τροχιά σχολιάζεται ως home to work. Όμοια οι τροχιές μπορούν να σχολιαστούν ως Tourist ή Shopping ανάλογα με το γεγονός αν οι περισσότερες στάσεις συμβαίνουν σε τουριστικά μέρη ή σε εμπορικά κέντρα. Ένα ενδιαφέρον χαρακτηριστικό αυτής της προσέγγισης είναι ότι χρησιμοποιεί μία σειρά από διαισθητικούς κανόνες κοινής λογικής προκειμένου να προκύψει η πιθανή συμπεριφορά σχολιασμού της τροχιάς.

Μία παρόμοια αλλά πιο εμπλουτισμένη προσέγγιση παρουσιάζεται από τους (Beghioni et al. 2012). Αυτή η εργασία σχολιάζει αυτόματα τροχιές με δραστηριότητες χρησιμοποιώντας μία οντολογία και ένα μηχανισμό εξαγωγής συμπερασμάτων. Η οντολογία ορίζει επίσημα τις

έννοιες της τροχιάς, του stop, του move, της δραστηριότητας της τροχιάς καθώς και των σχετικών πληροφοριών (για παράδειγμα Tourist Attraction, Accommodation Place). Οι διάφορες δραστηριότητες ορίζονται από οντολογικά αξιώματα δηλαδή λογικές εκφράσεις βασισμένες στα χαρακτηριστικά της τροχιάς. Για παράδειγμα μία Tourist δραστηριότητα μπορεί να οριστεί ως μία τροχιά που σταματά σε χώρους διαμονής τη νύχτα και σε τουριστικά αξιοθέατα τη μέρα. Οι τεχνικές που χρησιμοποιούνται για το σχολιασμό των τροχιών με δραστηριότητα μπορούν επίσης να χρησιμοποιηθούν για να εμπλουτίσουν τη γνώση για το ίδιο το κινούμενο αντικείμενο. Μία τροχιά, δηλαδή, που σταματά συχνά σε μουσεία, μνημεία και εστιατόρια χαρακτηρίζει πιθανότατα ένα τουρίστα. Στην περίπτωση αυτή ο σχολιασμός είναι διττός. Από τη μία η τροχιά από μόνη της δείχνει μία τουριστική δραστηριότητα, από την άλλη το κινούμενο αντικείμενο μπορεί να περιγραφεί ως τουριστικό άτομο.

1.5 Σχετικές Εργασίες

Τα βασικά ερωτήματα που προκύπτουν από τη χωρο- χρονική εξόρυξη δεδομένων είναι: τι είδους πρότυπα μπορούν να εξαχθούν από τις τροχιές; Ποιοι μέθοδοι και ποιοι αλγόριθμοι θα πρέπει να εφαρμοστούν για την εξαγωγή τους; Μία βασική μέθοδος που θα μπορούσε να «απαντήσει» στα παραπάνω ερωτήματα είναι η *συσταδοποίηση* (clustering).

Μία από τις πρώτες εργασίες που αναφέρθηκαν στην συσταδοποίηση των τροχιών ενός κινούμενου αντικειμένου είναι αυτή της Ketterlin (Ketterlin 1997). Στην έρευνα της θεωρεί ότι οι γενικές ακολουθίες (μοντελοποίηση των τροχιών ως ακολουθίες σημείων) σε συνδυασμό με μία εννοιολογική ιεραρχία στα στοιχεία της ακολουθίας μπορούν να χρησιμοποιηθούν για να υπολογιστεί τόσο η αντιπροσωπευτική ομάδα όσο και η απόσταση μεταξύ των ακολουθιών. Επιπλέον (Nanni 2002) προσαρμόζει δύο κλασσικές μεθόδους συσταδοποίησης (kmeans και hierarchical agglomerative clustering) στις τροχιές.

Μία εναλλακτική στρατηγική είναι να εφαρμοστεί στις τροχιές μία πολυδιάστατη κλιμακωτή τεχνική για μη διανυσματικά δεδομένα, όπως για παράδειγμα στην εργασία των (Faloutsos et al 1995) η οποία χαρτογραφεί ένα συγκεκριμένο χώρο δεδομένων σε έναν Ευκλείδειο χώρο διατήρησης (Κατά προσέγγιση) των αποστάσεων μεταξύ των αντικειμένων, έτσι ώστε να μπορεί να εφαρμοστεί οποιοσδήποτε αλγόριθμος συσταδοποίησης για διανυσματικά δεδομένα.

Οι Gaffney and Smyth (1999) στην εργασία τους προτείνουν ένα μοντέλο που βασίζεται στη συσταδοποίηση συνεχών τροχιών, το οποίο ομαδοποιεί αντικείμενα που είναι πιθανόν να παράγονται από ένα κοινό πυρήνα τροχιών.

Ο καθορισμός μια χωρο-χρονικής απόστασης που θα μπορούσε να ανταποκριθεί καλύτερα στη χωρο- χρονική φύση των τροχιών αποτέλεσε πρόκληση για τους Pelekis et al. (2011a), οι οποίοι στην έρευνά τους καθόρισαν μία μέθοδο η οποία ομαδοποιεί τις τροχιές χρησιμοποιώντας διάφορες συναρτήσεις απόστασης που βασίζονται στις ιδιότητες της κίνησης όπως είναι χωρική τοποθεσία, η ταχύτητα, η επιτάχυνση και η κατεύθυνση. Η μέθοδος αυτή αξιολογεί τις αποστάσεις ως προς το σύνολο των τροχιών. Σε αντίθεση, ο αλγόριθμος συσταδοποίησης που χρησιμοποιήθηκε από τους Lee et al αναζητά την ομοιότητα μεταξύ των τμημάτων των τροχιών χωρίς να απαιτεί να εφαρμόζεται στο σύνολο της τροχιάς (Lee et al 2009). Το πλεονέκτημα αυτής της μεθόδου, δηλαδή να χρησιμοποιείται η ομοιότητα ανάμεσα στα τμήματα, είναι η πιθανότητα να τελειοποιηθεί η ομαδοποίηση των τροχιών. Για παράδειγμα, μία ομάδα τουριστών στο Παρίσι περνούν το πρωί μαζί : συναντιούνται στην πλατεία Concorde, περπατούν μέχρι το Λούβρο, σταματούν εκεί για να το επισκεφθούν και έπειτα πηγαίνουν για μεσημεριανό. Μετά από αυτό χωρίζουν και ο καθένας συνεχίζει μόνος του. Οι τροχιές τους θεωρούνται παρόμοιες με βάση τα πρωινά τους τμήματα.

Όταν οι τροχιές δεν είναι ελεύθερες αλλά περιορίζονται από ένα δίκτυο, όπως είναι το οδικό δίκτυο για τις τροχιές των αυτοκινήτων, νέα μέτρα πυκνότητας και απόστασης πρέπει να ληφθούν υπόψη ώστε να υπολογίζουν την ιδιαιτερότητα του δικτύου. Συνήθως τα μέτρα απόστασης δεν βασίζονται πλέον στην Ευκλείδεια απόσταση αλλά σε αλγόριθμους που εντοπίζουν την πιο κοντινή διαδρομή σε ένα γράφημα (Tiakas et al 2009). Στους Hun et al (2012) οι συγγραφείς βρίσκουν συστάδες τμημάτων των τροχιών αυτοκινήτων ψάχνοντας για ακολουθίες συνεχόμενων οδικών τμημάτων που αποτελούνται από συνεχή ροή της κυκλοφορίας.

Η συσταδοποίηση των τροχιών έχει χρησιμοποιηθεί σε διαφορετικά πλαίσια εργασιών όπως η αναζήτηση και η εξέταση τροχιών (Panagiotakis et al 2012), η οπτικοποίηση των τροχιών (Rinzivillo et al 2008) και η επεξεργασία της αβεβαιότητας των τροχιών (Pelekis 2011b). Πιο συγκεκριμένα εισάγουν ένα μέτρο απόστασης για τις τροχιές, χρησιμοποιώντας ασαφή σύνολα (fuzzy sets) για να μοντελοποιήσουν την αβεβαιότητα των τροχιών και να υπολογίσουν αντιπροσωπευτικά πλέγματα τροχιών για τις συστάδες.

Στην εργασία των (Micro Nanni et al 2006) αντιμετωπίζεται το πρόβλημα της συσταδοποίησης των τροχιών. Βασική υπόθεση των συγγραφέων είναι ότι μία συλλογή από μεμονωμένες τροχιές των κινούμενων αντικειμένων μπορεί να ανακατασκευαστεί με προσεγγιστικό τρόπο βάσει των δεδομένων καταγραφής παρακολούθησης των αντικειμένων καθώς κινούνται μέσα σε ένα δίκτυο. Για παράδειγμα, ένα κινητό τηλέφωνο που «κινείται» μέσα σε ένα ασύρματο δίκτυο, κατά τη διάρκεια των αλληλεπιδράσεων του με αυτό, αποτελεί ένα σύνολο από τριάδες (id, loc, t) όπου το καθένα καθορίζει την τοποθεσία στο χώρο loc και στον χρόνο t της ταυτότητας του κινητού τηλεφώνου. Οι συγγραφείς γενίκευσαν την χωρική έννοια της απόστασης μεταξύ των αντικειμένων σε μία χωρο-χρονική έννοια της απόστασης μεταξύ τροχιών και πέτυχαν μία φυσική επέκταση της τεχνικής της συσταδοποίησης των τροχιών που βασίζεται στην πυκνότητα. Θεώρησαν ένα συγκεκριμένο αλγόριθμο συσταδοποίησης, OPTICS (Ankerst et al. 1999) και πρότειναν μία εμπειρική σύγκριση με αρκετούς παραδοσιακούς k-means και ιεραρχικούς αλγορίθμους.

Οι επιστήμονες διερεύνησαν επίσης τη χρήση των προσεγγίσεων συσταδοποίησης που βασίζονται στην πυκνότητα. Στην εργασία (Kalnis et al. 2005, Laube et al. 2005) χρησιμοποιούνται αλγόριθμοι που βασίζονται στην πυκνότητα για την εξαγωγή προτύπων κίνησης και την ανακάλυψη των συστάδων των κινούμενων αντικειμένων. Κινούμενοι στο ίδιο πλαίσιο οι Ester et al. (pp 226-231) προτείνουν αρκετούς αλγόριθμους που βασίζονται στη μέθοδο DBSCAN (Density Based Spatial Clustering of Application with Noise). Βασιζόμενοι σε αυτή τη μέθοδο οι Palma et al (2008 pp 863-868) ανακάλυψαν ενδιαφέροντα μέρη εντός ενός συνόλου δεδομένων των τροχιών.

Στην εργασία των (Lee et al 2007) προτείνεται ένα νέο πλαίσιο (partition and group framework) για τη συσταδοποίηση των τροχιών. Με βάση αυτό το πλαίσιο οι συγγραφείς ανέπτυξαν έναν αλγόριθμο συσταδοποίησης, TRACCLUS, κατά την εφαρμογή του οποίου αρχικά μία τροχιά διαμερίζεται σε ένα σύνολο γραμμικών τμημάτων. Στη συνέχεια τα παρόμοια γραμμικά τμήματα μίας πυκνής περιοχής, που έχουν προκύψει, ομαδοποιούνται σε μία συστάδα.

Οι διάφορες τεχνικές της συσταδοποίησης χρησιμοποιήθηκαν από τους ερευνητές και για την ανακάλυψη ενδιαφέροντων ομάδων ατόμων με κοινά χαρακτηριστικά. Πιο συγκεκριμένα οι Kumar et al (2013), λαμβάνοντας δεδομένα από τον Οργανισμό Ασφαλείων Υγείας της Ινδίας, μέσω της συσταδοποίησης και με χρήση του αλγορίθμου k means ανακάλυψαν ομάδες ατόμων που ήθελαν να ασφαλιστούν με βάση τα ίδια κριτήρια που ικανοποιούσαν τις

ανάγκες τους. Αρχικά ξεχώρισαν τρία άτομα με βάση τα οποία θα ανακάλυπταν τις ομάδες που θα προέκυπταν με χρήση του αλγορίθμου. Υπολόγισαν την απόσταση, χρησιμοποιώντας τα χαρακτηριστικά των ατόμων, και βασιζόμενοι σε αυτή την απόσταση τοποθέτησαν το κάθε άτομο στην πλησιέστερη ομάδα. Στη συνέχεια σύγκριναν τα μέσα των ομάδων που έχουν προκύψει με τις αρχικές ομάδες και υπολόγισαν εκ νέου την απόσταση. Όταν έφτασαν στο σημείο που δεν προέκυπταν νέες ομάδες, θεώρησαν αυτές τις ομάδες που είχαν υπολογίσει ως τελικές.

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

ΚΕΦΑΛΑΙΟ 2

ΘΕΩΡΗΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ ΜΕΘΟΔΟΥ

Στο παρόν κεφάλαιο γίνεται αναφορά στις βασικές έννοιες της μεθόδου συσταδοποίησης. Περιγράφεται η μέθοδος που χρησιμοποιήσαμε, το σύνολο των δεδομένων με το οποίο εργαστήκαμε και παρουσιάζεται θεωρητικά ο τρόπος λειτουργίας του στατιστικού πακέτου που χρησιμοποιήσαμε.

2.1 Συσταδοποίηση

Η ανάλυση κατά συστάδες (συσταδοποίηση) είναι μία μέθοδος που σκοπό έχει να κατατάξει σε ομάδες τις υπάρχουσες παρατηρήσεις ενός συνόλου δεδομένων χρησιμοποιώντας την πληροφορία που υπάρχει σε κάποιες μεταβλητές. Θα μπορούσαμε να πούμε πως εξετάζοντας πόσο όμοιες είναι κάποιες παρατηρήσεις ως προς κάποιον αριθμό μεταβλητών η μέθοδος τείνει να δημιουργεί ομάδες από παρατηρήσεις που μοιάζουν μεταξύ τους.

Η συσταδοποίηση χρησιμοποιείται σε πολλές επιστήμες για να ομαδοποιήσει δεδομένα. Για παράδειγμα διαφορετικά είδη ζώων μπορούν να ομαδοποιηθούν με βάση κάποια χαρακτηριστικά τους όπως επίσης και οι πελάτες σε μία έρευνα αγοράς.

Συστάδα (Cluster)

Μία συστάδα είναι ένα σύνολο από παρατηρήσεις (ή αντικείμενα) τέτοιο ώστε κάθε παρατήρηση της συστάδας είναι πιο κοντά στο «κέντρο» της από ότι στο κέντρο οποιασδήποτε άλλης συστάδας.

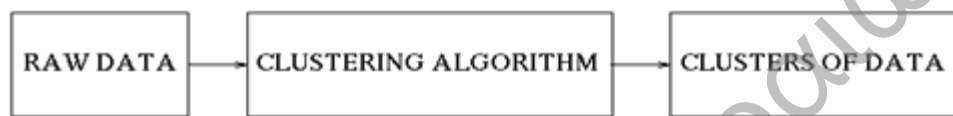
Cluster Centroid / Κέντρο Συστάδας

Το κέντρο μίας συστάδας είναι ο μέσος όρος των τιμών των παραμέτρων των παρατηρήσεων της συστάδας.

2.1.1 Αλγόριθμος Συσταδοποίησης

Ένας αλγόριθμος συσταδοποίησης προσπαθεί να βρει φυσικά ομάδες των δεδομένων με βάση κάποια ομοιότητα. Ο αλγόριθμος βρίσκει το κέντρο βάρους (centroid) μίας συστάδας ενός συνόλου δεδομένων και προκειμένου να καθορίσει τη σχέση ανάμεσα στις συστάδες, αξιολογεί την απόσταση ενός σημείου και του κέντρου βάρους της συστάδας.

Τα αποτελέσματα ενός αλγόριθμου συσταδοποίησης είναι βασικά μία στατιστική περιγραφή των κέντρων των συστάδων με τον αριθμό των παρατηρήσεων σε κάθε συστάδα.



Σχήμα 2-1: Γραφική απεικόνιση αλγορίθμου συσταδοποίησης

2.2 Απόσταση- Ομοιότητα Συστάδων

Βασικές έννοιες για την ανάλυση κατά συστάδες είναι οι έννοιες της απόστασης και της ομοιότητας καθώς είναι αυτές που μας επιτρέπουν να μετρήσουμε πόσο μοιάζουν οι παρατηρήσεις μεταξύ τους και κατά συνέπεια μας επιτρέπουν να τις τοποθετήσουμε στην ίδια ομάδα. Μπορούμε εύκολα να διαπιστώσουμε πως αυτές οι δύο έννοιες είναι αντίθετες καθώς παρατηρήσεις που είναι όμοιες μεταξύ τους θα έχουν μεγάλη ομοιότητα και μικρή απόσταση.

Ομοιότητα

Η ομοιότητα μεταξύ δύο παρατηρήσεων μετριέται συνήθως με βάση μίας συνάρτησης απόστασης, $d(i,j)$, η οποία εξαρτάται από το είδος των δεδομένων, δηλαδή από το είδος των γνωρισμάτων τους. Για παράδειγμα στην περίπτωση που έχουμε συνεχείς και τυποποιημένες μεταβλητές, όπως είναι οι μεταβλητές του συνόλου δεδομένου όπου εργαστήκαμε, η Ευκλείδεια απόσταση είναι αυτή που χρησιμοποιείται συνήθως.

Συχνές Ιδιότητες Συναρτήσεων Απόστασης

1. $d(i, j) \geq 0$
2. $d(i, i) = 0$ (ανακλαστική ιδιότητα)
3. $d(i, j) = d(j, i)$ (συμμετρική ιδιότητα)
4. $d(i, j) \leq d(i, h) + d(h, j)$ (τριγωνική ιδιότητα)

Όταν ισχύουν και οι τέσσερις ιδιότητες η συνάρτηση απόστασης ονομάζεται και μετρική απόστασης (distance metric).

Αποστάσεις

Τα είδη των αποστάσεων διακρίνονται σε δύο μεγάλες κατηγορίες:

- Ευκλείδειες αποστάσεις (βασίζονται στη θέση των σημείων)

Έστω δύο μεταβλητές i και j με n γνωρίσματα x_{ik} και x_{jk} $i = (x_{i1}, x_{i2}, \dots, x_{in})$ $j = (x_{j1}, x_{j2}, \dots, x_{jn})$, η Ευκλείδεια απόσταση δίνεται από την ακόλουθη σχέση:

$$d(i, j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{in} - x_{jn}|^2)}$$

- Μη ευκλείδειες αποστάσεις (βασίζονται σε άλλες ιδιότητες των σημείων πλην της θέσης τους). Παραδείγματα μη ευκλείδειων αποστάσεων είναι οι Jaccard distance, cosine distance, edit distance, hamming distance.

2.3 Είδη Συσταδοποίησης

Η κύρια διάκριση στις προσεγγίσεις συσταδοποίησης είναι μεταξύ των ιεραρχικών (*hierarchical*) και των διαμεριστικών (*partitional*) προσεγγίσεων. Η διαφορά τους έγκειται στο γεγονός ότι στην πρώτη περίπτωση η ομαδοποίηση των δεδομένων μας παράγει μία συγκεκριμένη δομή- σχέση μεταξύ τους, ενώ η δεύτερη ομαδοποιεί τα δεδομένα μας χωρίς να έχουν οι συστάδες αυτές σχέση μεταξύ τους, δηλαδή παράγει μια αρχιτεκτονική δομή συστάδων ασυσχέτιστων μεταξύ τους.

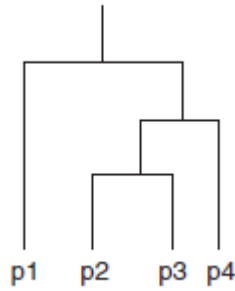
Πιο συγκεκριμένα οι *διαμεριστικοί αλγόριθμοι* συσταδοποίησης λαμβάνουν μία και μόνο διαμέριση- κατάτμηση των δεδομένων αντί για μία ενιαία δομή, όπως η δομή δέντρου που παράγεται από τις ιεραρχικές τεχνικές. Με άλλα λόγια ξεκινούν με όλο το σύνολο των δεδομένων και δημιουργούν μία διαμέριση σε αυτό, η οποία αντιπροσωπεύει μία συστάδα.

Οι *ιεραρχικές τεχνικές* παράγουν μία εμφωλευμένη αλληλουχία των διαμερίσεων, με μία ενιαία, που περιλαμβάνει όλα τα στοιχεία της συστάδας (cluster) στην κορυφή και μονήρεις συστάδες των μεμονωμένων σημείων στο κάτω μέρος. Κάθε ενδιάμεσο επίπεδο μπορεί να θεωρηθεί ως ο συνδυασμός (διάσπαση) δύο ομάδων από το αμέσως προηγούμενο (επόμενο υψηλότερο) επίπεδο.

Μία σημαντική διαφορά ανάμεσα στις δύο προσεγγίσεις είναι ότι στη διαμεριστική μέθοδο συσταδοποίησης ο αριθμός των συστάδων που θα προκύψουν πρέπει να είναι γνωστός εκ των προτέρων, κάτι που δεν ισχύει για την ιεραρχική προσέγγιση. Επιπλέον στην ιεραρχική συσταδοποίηση απαιτείται μόνο ένα μέτρο ομοιότητας ανάμεσα στις ομάδες των αντικειμένων των δεδομένων μας σε αντίθεση με τους αλγόριθμους που εφαρμόζονται στην διαχωριστική συσταδοποίηση (όπως για παράδειγμα στον k means) όπου για να τους εφαρμόσουμε εκτός από τον αριθμό των συστάδων που θα πρέπει να γνωρίζουμε, όπως αναφέραμε, απαιτείται ένα μέτρο απόστασης μεταξύ των δεδομένων καθώς και μία αρχική ανάθεση των δεδομένων σε συστάδες.

2.4 Ιεραρχική Συσταδοποίηση

Η Ιεραρχική Συσταδοποίηση (Hierarchical Clustering) είναι μία μέθοδος συσταδοποίησης κατά την οποία παράγονται εμφωλιασμένες ή εμφωλευμένες (nested) συστάδες, δηλαδή η μικρότερη συστάδα περικλείεται στην αμέσως μεγαλύτερη, που δομούν μία ιεραρχία ή αλλιώς ένα ιεραρχικό δέντρο. Κάθε κόμβος (συστάδα) στο δέντρο (εκτός από τους κόμβους φύλλα όπου κάθε στοιχείο αποτελεί μία ξεχωριστή συστάδα) είναι η ένωση των συστάδων- παιδιών (υποομάδες) και η ρίζα του δέντρου είναι η ομάδα- συστάδα που περιέχει όλα τα στοιχεία.



Σχήμα 2-2: Δενδρόγραμμα Ιεραρχικής Συσταδοποίησης

Στην ιεραρχική συσταδοποίηση υπάρχουν δύο βασικές προσεγγίσεις:

- Η συσσωρευτική προσέγγιση (*Agglomerative/ Amalgamative/Bottom- Up*). Ξεκινά με τα σημεία ως ξεχωριστές συστάδες και σε κάθε βήμα συγχωνεύει το πιο κοντινό ζευγάρι συστάδων μέχρι να μείνει μόνο μία ή k συστάδες.
- Η διαιρετική (*Divisive/ Top- Down*) η οποία είναι ακριβώς αντίστροφη της συσσωρευτικής, δηλαδή ξεκινά με μία συστάδα που περιέχει όλα τα σημεία- δεδομένα και έπειτα σε κάθε βήμα διαχωρίζει μία συστάδα μέχρι κάθε συστάδα να περιέχει ένα μόνο σημείο.

Οι τεχνικές ιεραρχικής συσσωρευτικής συσταδοποίησης απαιτούν λιγότερο χρόνο υπολογισμού και είναι πιο δημοφιλείς.

2.4.1 Συσσωρευτικοί Ιεραρχικοί Αλγόριθμοι

Οι συσσωρευτικοί αλγόριθμοι που χρησιμοποιούνται έχουν την ίδια δομή και διαφέρουν μόνο στον τρόπο που υπολογίζονται οι αποστάσεις μεταξύ των συστάδων. Τα βήματα έχουν ως εξής:

1. Ξεκίνησε με τόσες συστάδες όσες και οι παρατηρήσεις. Οι αποστάσεις μεταξύ των συστάδων είναι οι αποστάσεις μεταξύ των αρχικών παρατηρήσεων.
2. Επέλεξε τις δύο περισσότερο όμοιες παρατηρήσεις στον πίνακα αποστάσεων και με αυτές σχημάτισε μία νέα συστάδα.

3. Αντικατέστησε τις δύο παρατηρήσεις που χρησιμοποιήθηκαν στο βήμα 2 με τη νέα συστάδα. Οι αποστάσεις μεταξύ της νέας συστάδας και των υπολοίπων υπολογίζεται με βάση μίας απόστασης σύνδεσης (linkage method).
4. Πήγαινε πίσω στο βήμα 2 και επανέλαβε τα βήματα 2 και 3 έως ότου όλες οι συστάδες να έχουν συγχωνευθεί σε μία ενιαία συστάδα.

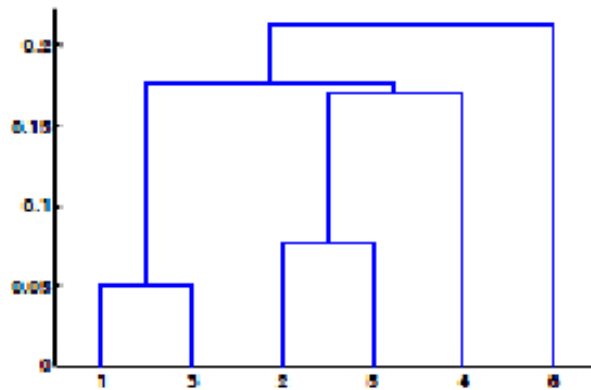
Διαφορετικά:

-
- 1: Υπολογισμός του Πίνακα Γειτνίασης
 - 2: Έστω κάθε σημείο αποτελεί και μια συστάδα
 - 3: **Repeat**
 - 4: Συγχώνευση των δύο κοντινότερων συστάδων
 - 5: Ενημέρωση του Πίνακα Γειτνίασης
 - 6: **Until** να μείνει μία μόνο συστάδα
-

Σχήμα 2-3: Αλγόριθμος Ιεραρχικής Συσταδοποίησης

Δενδρόγραμμα

Μία ιεραρχική συσσωρευτική συσταδοποίηση μπορεί να απεικονιστεί με ένα δενδρόγραμμα. Το δενδρόγραμμα εμφανίζει τις πληροφορίες του πίνακα συγχώνευσης με τη μορφή ενός διαγράμματος δέντρου. Οι αρχικές παρατηρήσεις, η καθεμία από τις οποίες αποτελούν μία συστάδα, απεικονίζονται στο κάτω μέρος του γραφήματος. Οι συγχωνεύσεις μεταξύ των συστάδων αντιπροσωπεύονται από τρεις ευθείες γραμμές, οι οποίες ενώνονται στους κόμβους των οποίων η θέση κατά μήκος του άξονα x δείχνει το επίπεδο στο οποίο συμβαίνει η συγχώνευση.



Σχήμα 2-4: Δενδρόγραμμα ιεραρχικής συσσωρευτικής συσταδοποίησης

2.4.2 Κριτήρια καθορισμού της απόστασης μεταξύ των συστάδων

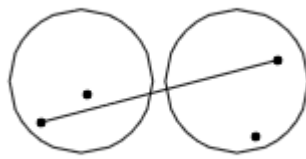
Το κλειδί της λειτουργίας ενός συσσωρευτικού αλγορίθμου είναι ο υπολογισμός της απόστασης (ή διαφορετικά της απόστασης σύνδεσης) μεταξύ δύο συστάδων και είναι ο καθορισμός αυτής της απόστασης που διαφοροποιεί τις διάφορες τεχνικές της συσσωρευτικής ιεραρχικής συσταδοποίησης (linkage methods).

Η *single linkage* or nearest neighbor or min/ απλή σύνδεση ή πλησιέστερος γείτονας τεχνική καθορίζει την απόσταση των συστάδων ως την ελάχιστη απόσταση μεταξύ των παρατηρήσεων στις συστάδες. Η *complete linkage* or furthest neighbor or max/ πλήρης σύνδεση ή απότερος γείτονας λαμβάνει ως απόσταση των συστάδων την απόσταση μεταξύ των πιο απομακρυσμένων παρατηρήσεων στις διαφορετικές συστάδες. Μία άλλη προσέγγιση, η *group average*/ μέσος όρος της συστάδας τεχνική, καθορίζει την απόσταση των συστάδων ως τον μέσο όρο των αποστάσεων όλων των ζευγών των παρατηρήσεων από τις διαφορετικές συστάδες. Τέλος στην περίπτωση που κάθε συστάδα αναπαριστάται από ένα κέντρο η απόσταση μεταξύ των συστάδων καθορίζεται ως η απόσταση μεταξύ των κέντρων των συστάδων (*centroid linkage*).

Στην παρούσα διπλωματική εργασία προκειμένου να υπολογίσουμε τις αποστάσεις μεταξύ των συστάδων βασιστήκαμε στην complete linkage τεχνική περαιτέρω ανάλυση της οποίας ακολουθεί παρακάτω.

2.5 Complete linkage/Πλήρης Σύνδεση

Η complete-linkage τεχνική είναι μία από τις πολλές μεθόδους συσσωρευτικής ιεραρχικής ομαδοποίησης. Στην αρχή της διαδικασίας, κάθε παρατήρηση είναι μία συστάδα από μόνη της. Στη συνέχεια οι συστάδες διαδοχικά συγχωνεύονται σε μεγαλύτερες συστάδες, μέχρις ότου όλες οι παρατηρήσεις να ανήκουν στην ίδια συστάδα. Σε κάθε βήμα, οι δύο συστάδες οι οποίες έχουν διαχωριστεί βάση της μεγαλύτερης απόστασης, συγχωνεύονται.



Σχήμα 2-5: : Complete-link: ελάχιστη ομοιότητα

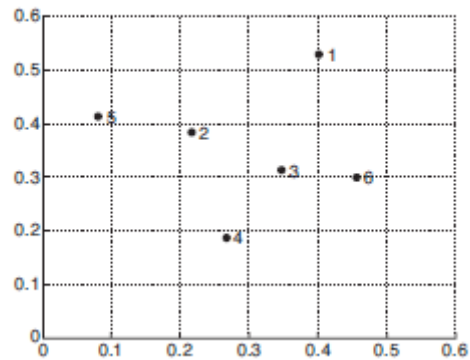
Ο ορισμός της «μικρότερης απόστασης» είναι αυτό που διαφοροποιεί τις διαφορετικές συσσωρευτικές μεθόδους συσταδοποίησης. Στην complete linkage συσταδοποίηση, η σύνδεση μεταξύ δύο συστάδων περιέχει όλα τα ζεύγη των στοιχείων, και η απόσταση μεταξύ των συστάδων ισούται με την απόσταση μεταξύ των δύο αυτών στοιχείων (ένα σε κάθε συστάδα), που είναι η μεγαλύτερη απόσταση. Η μικρότερη από αυτές τις συνδέσεις που παραμένει σε οποιοδήποτε στάδιο προκαλεί τη συγχώνευση των δύο συστάδων, των οποίων τα στοιχεία είναι που συγχωνεύονται. Η μέθοδος είναι επίσης γνωστή ως συσταδοποίηση του «απομακρυσμένου γείτονα».

Η complete linkage συνάρτηση, η απόσταση $D(X,Y)$ μεταξύ των συστάδων X και Y , περιγράφεται από την ακόλουθη σχέση:

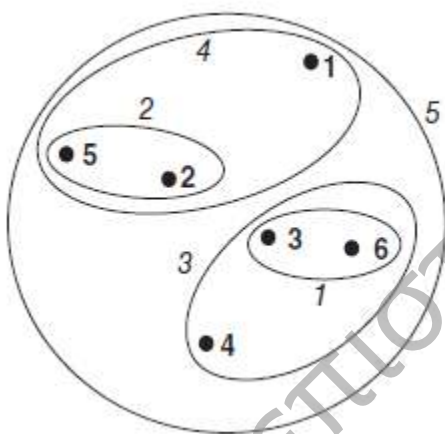
$$D(X, Y) = \max_{x \in X, y \in Y} d(x, y)$$

όπου: $d(x,y)$ είναι η απόσταση μεταξύ των στοιχείων $x \in X$ και $y \in Y$ ενώ X και Y είναι τα δύο σύνολα των στοιχείων (συστάδες).

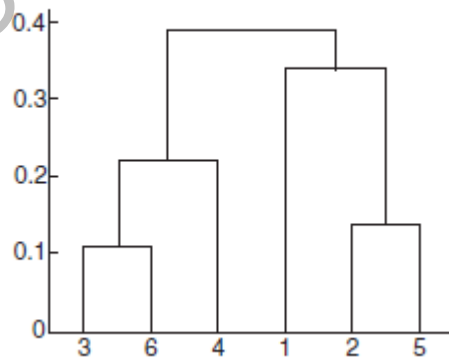
Το αποτέλεσμα της συσταδοποίησης μπορεί να απεικονιστεί ως ένα δενδρόγραμμα που δείχνει την αλληλουχία της συγχώνευσης των συστάδων και την απόσταση στην οποία η κάθε συγχώνευση πραγματοποιήθηκε.



Σχήμα 2-6: Ένα σύνολο από έξι διδιάστα σημεία



Σχήμα 2-7: Complete linkage clustering



Σχήμα 2-8: Complete linkage dendrogram

Στο Σχήμα 2.6 απεικονίζεται ένα σύνολο έξι διδιάστατων σημείων. Στα Σχήματα 2.7 και 2.8 παρουσιάζονται η συσταδοποίηση και το αντίστοιχο δενδρόγραμμα αυτών των σημείων με βάση την complete linkage τεχνική.

Σύγκριση με τις υπόλοιπες τεχνικές

Η complete-linkage τεχνική συσσωρευτικής συσταδοποίησης αποφεύγει ένα μειονέκτημα της εναλλακτικής single-linkage μεθόδου - το λεγόμενο φαινόμενο αλυσιδωτή σύνδεση, όπου οι συστάδες που σχηματίζονται, μέσω της μεθόδου αυτής, μπορεί να αναγκαστούν να είναι μαζί λόγω μεμονωμένων στοιχείων που είναι κοντά το ένα στο άλλο, παρόλο που πολλά από τα στοιχεία σε κάθε συστάδα μπορεί να είναι πολύ απομακρυσμένα μεταξύ τους. Επιπλέον σε αντίθεση με τη single-linkage τεχνική συσσωρευτικής ιεραρχικής συσταδοποίησης και τον αλγόριθμο k means της διαμεριστικής προσέγγισης, η μέθοδος αυτή έχει μικρότερη ευαισθησία σε θόρυβο και outliers. Η centroid linkage (κέντρου βάρους) τεχνική εμφανίζει ένα χαρακτηριστικό μειονέκτημα: την πιθανότητα αναστροφής. Συγκεκριμένα, δύο συστάδες που συγχωνεύονται μπορεί να είναι περισσότερο όμοιες (να έχουν δηλαδή μικρότερη απόσταση) με συστάδες που έχουν συγχωνευτεί σε προηγούμενα βήματα.

2.6 Δεδομένα

Τα διαθέσιμα δεδομένα που χρησιμοποιήθηκαν σε αυτή την εργασία συλλέχθηκαν τη περίοδο από 13 Απριλίου 2006 έως τις 14 Απριλίου 2008 σε μία πόλη του Βελγίου και συγκεκριμένα στη Φλάνδρα. Το σύνολο των δεδομένων προέρχεται από δύο διαφορετικές πηγές δεδομένων. Τα «δεδομένα GPS» τα οποία αποτελούν εγγραφές GPS ανά χρήστη και προέρχονται από τις ενεργοποιημένες συσκευές GPS και τα «Ημερολόγια» κίνησης (Diaries) στα οποία κάθε χρήστης καταγράφει τη σημασιολογική πληροφορία για το ταξίδι του (trip). Ένα ταξίδι για ένα χρήστη είναι μία διαδρομή από το σημείο A στο σημείο B. Το συγκεκριμένο αρχείο επικεντρώνεται σε δεδομένα τα οποία προέρχονται από τις δύο παραπάνω πηγές και αφορούν 186 χρήστες που κατέγραψαν τα ταξίδια τους (τις διαδρομές τους).

Τα ημερολόγια κίνησης περιέχουν την μορφή των ταξιδιών εμπλουτισμένα με σημασιολογικές πληροφορίες για κάθε ταξίδι. . Επιπλέον κάθε ταξίδι έχει σχέση με τον αριθμό των ταξιδιών που βρέθηκε στην πηγή των «δεδομένων GPS» για κάθε χρήστη, και έτσι μπορούμε να αντιστοιχίσουμε τις εγγραφές GPS με τα ταξίδια στα ημερολόγια κίνησης. Επιπλέον για κάθε ταξίδι στα ημερολόγια υπάρχουν σημασιολογικές πληροφορίες σχετικές

με την προέλευση και τον προορισμό, τη δραστηριότητα του χρήστη σε κάποιο προορισμό και το μεταφορικό μέσο που αυτός χρησιμοποίησε. Αυτά τα ημερολόγια κίνησης υποβλήθηκαν σε επεξεργασία για την παραγωγή σημασιολογικών τροχιών ως εξής: αρχικά συνδυάστηκαν πληροφορίες από τις παραπάνω πηγές και επαναλαμβάνεται για κάθε χρήστη μέσω των ημερολόγιων ταξιδιών τους. Ένα ημερολόγιο κίνησης θεωρείται ένα επεισόδιο κίνησης (move episode) οπότε επεισόδια στάσης (stop episode) πρέπει να προστεθούν πριν και μετά από αυτό το ταξίδι. Σημασιολογικές τροχιές δημιουργούνται κάθε φορά που ένας καινούριος χρήστης ή μία καινούρια ημέρα εισέρχονται.

Από το σύνολο δεδομένων που είχαμε στη διάθεση μας επιλέξαμε τις 600 πρώτες εγγραφές και εργαστήκαμε με τις εξής ακόλουθες μεταβλητές:

- `epis_id` η οποία δηλώνει την ταυτότητα του επεισοδίου για κάθε μία τροχιά
- `sem_traj_id` η οποία δηλώνει την ταυτότητα της τροχιάς για κάθε χρήστη και ανάλογα με τον χρήστη διακρίνεται σε δύο έως τέσσερις κατηγορίες (κάθε κατηγορία αντιστοιχεί σε ένα επεισόδιο είτε στάσης είτε κίνησης)
- `epis_def_tag` η οποία δηλώνει το σχολιασμό του επεισοδίου ως επεισόδιο στάσης (stop) ή επεισόδιο κίνησης (move)
- `avg_speed` η οποία δηλώνει τη μέση ταχύτητα του χρήστη κατά τη διάρκεια του κάθε επεισοδίου

Σκοπός της παρούσας διπλωματικής εργασίας ήταν να μελετήσουμε, μέσω στοχαστικών διαδικασιών, τις πιθανές αλληλεπιδράσεις που μπορούν να έχουν μεταξύ τους οι μεταβλητές που επιλέξαμε να μελετήσουμε. Εφαρμόσαμε λοιπόν τον στοχαστικό ιεραρχικό αλγόριθμο συσταδοποίησης συσσωρευτικά, ομαδοποιώντας δηλαδή τα δεδομένα μας σε συμπαγείς συστάδες δηλαδή βάση της μέγιστης τετραγωνικής Ευκλείδειας απόστασης (complete linkage) που έχουν μεταξύ τους οι συστάδες, στα ακόλουθα ζεύγη των παραπάνω μεταβλητών: `epis_id` και `sem_traj_id`, `epis_id` και `epis_def_tag` και `sem_traj_id` και `avg_speed`.

Να σημειώσουμε ότι προκειμένου να χρησιμοποιήσουμε τη μεταβλητή `epis_def_tag` ορίσαμε δύο ψευδομεταβλητές σύμφωνα με τις οποίες όπου stop βάλουμε τον αριθμό 1 και όπου move τον αριθμό 2.

2.7 Βήματα Μεθόδου

Στην παρούσα διπλωματική εργασία θα χρησιμοποιήσουμε, όπως αναφέρθηκε, το συσσωρευτικό αλγόριθμο Ιεραρχικής Συσταδοποίησης εφαρμόζοντας τον στο στατιστικό λογισμικό πακέτο Minitab 16. Πριν προχωρήσουμε στην παρουσίαση των αποτελεσμάτων θα δώσουμε μία θεωρητική περιγραφή του τρόπου λειτουργίας του πακέτου καθώς και της ερμηνείας των πινάκων που προκύπτουν από αυτό.

Στην αρχή της μεθόδου κάθε παρατήρηση αποτελεί μία συστάδα από μόνη της. Στο πρώτο στάδιο οι δύο πιο κοντινές παρατηρήσεις ενώνονται με αποτέλεσμα να έχουμε μία καινούργια συστάδα η οποία αποτελείται από αυτές τις δύο παρατηρήσεις και έτσι η νέα συστάδα «περικλείει» δύο υποσυστάδες. Στο δεύτερο στάδιο είτε μία τρίτη παρατήρηση ενώνεται με τις δύο πρώτες είτε δύο άλλες παρατηρήσεις ενώνονται σε μία διαφορετική συστάδα. Αυτή η διαδικασία θα συνεχιστεί έως ότου όλες οι συστάδες να ενωθούν σε μία.

Το λογισμικό πακέτο δίνει τέσσερις πίνακες. Ο πρώτος πίνακας δείχνει ποιες συστάδες ενώθηκαν, την μεταξύ τους απόσταση, το αντίστοιχο επίπεδο ομοιότητας, τον αριθμό των καινούργιων συστάδων (ο οποίος είναι πάντα μικρότερος από τους δύο αριθμούς των συστάδων που ενώνονται), τον αριθμό των παρατηρήσεων σε κάθε συστάδα και τον αριθμό των συστάδων. Η συγχώνευση των συστάδων συνεχίζεται μέχρι να υπάρχει μόνο μία συστάδα που να περιέχει όλες τις παρατηρήσεις.

Όταν καθοριστεί και η τελευταία συγχώνευση εμφανίζονται τρεις επιπλέον πίνακες. Ο πρώτος πίνακας συνοψίζει κάθε συστάδα με τον αριθμό των παρατηρήσεων, το «within cluster sum of squares», το «average distance from centroid» και «maximum distance from centroid». Πιο συγκεκριμένα:

- i. Το «within sum of squares» είναι η τετραγωνική Ευκλείδεια απόσταση μεταξύ των μέσων των δύο συστάδων (centroid) και μετράει ουσιαστικά τη τετραγωνική διαφορά. Συγκεκριμένα το «centroid» ή πρότυπη-μέση της συστάδας είναι ένα διάνυσμα που περιέχει ένα αριθμό από κάθε μία μεταβλητή της συστάδας, ο οποίος αριθμός αυτός είναι η μέση τιμή της κάθε μεταβλητής της εν λόγω συστάδας.
- ii. Το «Average Distance from centroid» είναι η μέση τιμή των αποστάσεων των παρατηρήσεων των μεταβλητών της συστάδας από τη μέση της συστάδας. Το «maximum distance from centroid» είναι η μέγιστη απόσταση των παρατηρήσεων των μεταβλητών της συστάδας από τη μέση της συστάδας.

Σε γενικές γραμμές, μία συστάδα με μικρό άθροισμα τετραγώνων (sum of squares) είναι πιο συμπαγής από μία συστάδα με μεγάλο άθροισμα.

Ο δεύτερος πίνακας εμφανίζει τα κέντρα των μεμονωμένων συστάδων ενώ ο τρίτος δίνει τις αποστάσεις μεταξύ των κέντρων των συστάδων.

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

ΚΕΦΑΛΑΙΟ 3

ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Σε αυτό το κεφάλαιο παρουσιάζονται τα αποτελέσματα που πήραμε εφαρμόζοντας στο λογισμικό πακέτο Minitab 16 το στοχαστικό αλγόριθμο συσταδοποίησης συσσωρευτικά, ομαδοποιώντας τα δεδομένα μας σε συμπαγείς συστάδες δηλαδή βάση της μέγιστης τετραγωνικής Ευκλείδειας απόστασης (complete-linkage) που έχουν μεταξύ τους (οι συστάδες). Επιπλέον επειδή όλες οι μεταβλητές του συνόλου δεδομένων μας έχουν διαφορετικές μονάδες μέτρησης, εφαρμόσαμε τις τυποποιημένες τιμές τους.

Εφαρμόσαμε τον αλγόριθμο συσταδοποίησης στα ακόλουθα ζεύγη μεταβλητών: epis_id-sem_traj_id, epis_id-epis_deg_tag και sem_traj_id-avg_speed, οι οποίες έχουν αναλυθεί στο κεφάλαιο 2. Τα αποτελέσματα που πήραμε και η ανάλυση τους ακολουθεί παρακάτω.

3.1 Συσταδοποίηση epis_id- epis_def_tag

Από τον Πίνακα 3.1 του Minitab, ο οποίος είναι ο πρώτος σε σειρά που εμφανίζεται και αφορά τη συγχώνευση (Amalgamation/ Agglomeration) των συστάδων, παρατηρούμε ότι σε κάθε βήμα (step) τα δεδομένα μας ομαδοποιούνται σε δύο συστάδες που στη συνέχεια αυτές συγχωνεύονται και προκύπτει μία νέα συστάδα. Για παράδειγμα στην πρώτη δοκιμή, βάσει 100% ομοιότητας και παράλληλα μηδενικής απόστασης μεταξύ των συστάδων, δημιουργήθηκαν δύο συστάδες. Η πρώτη συστάδα αποτελείται από τις πρώτες 595 παρατηρήσεις και είναι μέρος της δεύτερης συστάδας η οποία αποτελείται από 600 παρατηρήσεις. Οι δύο αυτές συστάδες συγχωνεύονται και η νέα πλέον συστάδα (η οποία έχει αριθμό αναγνώρισης 595) απαρτίζεται μόνο από 2 παρατηρήσεις (ουσιαστικά απαρτίζεται από δύο υποσυστάδες). Η διαδικασία αυτή συνεχίζεται, ώσπου η τελευταία 599^η παρατήρηση να συμπληρώσει άθροισμα παρατηρήσεων τελευταίας συστάδας ίσο με 600, όσο είναι δηλαδή και το μέγεθος του δείγματός μας.

Cluster Analysis of Observations: epis_id; epis_def_tag

Standardized Variables, Euclidean Distance, Complete Linkage, Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	599	100,000	0,00000	595 600	595	2
2	598	100,000	0,00000	592 599	592	2
3	597	100,000	0,00000	591 598	591	2
4	596	100,000	0,00000	590 597	590	2
5	595	100,000	0,00000	589 596	589	2
6	594	100,000	0,00000	588 595	588	3
7	593	100,000	0,00000	570 594	570	2
8	592	100,000	0,00000	569 593	569	2
9	591	100,000	0,00000	587 592	587	3
10	590	100,000	0,00000	586 591	586	3
11	589	100,000	0,00000	585 590	585	3
12	588	100,000	0,00000	584 589	584	3
13	587	100,000	0,00000	583 588	583	4
14	586	100,000	0,00000	582 587	582	4
15	585	100,000	0,00000	581 586	581	4
16	584	100,000	0,00000	580 585	580	4
17	583	100,000	0,00000	579 584	579	4
18	582	100,000	0,00000	578 583	578	5
19	581	100,000	0,00000	577 582	577	5
20	580	100,000	0,00000	576 581	576	5
21	579	100,000	0,00000	575 580	575	5
22	578	100,000	0,00000	574 579	574	5
23	577	100,000	0,00000	573 578	573	6
24	576	100,000	0,00000	568 577	568	6
25	575	100,000	0,00000	567 576	567	6
26	574	100,000	0,00000	566 575	566	6
27	573	100,000	0,00000	565 574	565	6
28	572	100,000	0,00000	564 573	564	7
29	571	100,000	0,00000	541 572	541	2
30	570	100,000	0,00000	540 571	540	2
31	569	100,000	0,00000	539 570	539	3
32	568	100,000	0,00000	538 569	538	3
33	567	100,000	0,00000	563 568	563	7
580	20	90,152	0,56361	340 342	340	6
581	19	90,152	0,56361	90 341	90	11
582	18	90,152	0,56361	1 3	1	179
583	17	90,152	0,56361	87 89	87	20
584	16	90,152	0,56361	29 88	29	34
585	15	90,152	0,56361	26 28	26	57
586	14	90,152	0,56361	5 27	5	115
587	13	90,152	0,56361	2 4	2	169
588	12	90,152	0,56361	416 418	416	3
589	11	90,152	0,56361	415 417	415	3
590	10	80,304	1,12722	90 343	90	14
591	9	70,456	1,69083	1 5	1	294
592	8	70,456	1,69083	87 340	87	26
593	7	70,456	1,69083	2 26	2	226
594	6	61,719	2,19089	415 416	415	6
595	5	60,608	2,25444	29 90	29	48
596	4	50,651	2,82428	29 87	29	74
597	3	50,651	2,82428	1 2	1	520
598	2	35,339	3,70059	29 415	29	80
599	1	0,000	5,72310	1 29	1	600

Πίνακας 3-1: Βήματα Συγχώνευσης Συστάδων

Επιπλέον από τον Πίνακα 3.1 παρατηρούμε ότι από την 580^η επανάληψη και μετά η ομοιότητα αρχίζει να μειώνεται και να αυξάνεται η απόσταση. Επομένως μπορούμε μία πρώτη ματιά ότι η συσταδοποίηση δεν είναι καλώς ορισμένη μόνο κατά 3,3% ($\{600-580/600\} \cdot 100$) στο σύνολο των παρατηρήσεων. Η πρώτη εντύπωση που μας δίνεται είναι ότι οι συστάδες θα είναι κατά κύριο λόγο καλά ορισμένες.

Final Partition
Number of clusters: 2

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	520	686,359	1,13449	1,68065
Cluster2	80	134,451	1,22993	2,67043

Πίνακας 3-2: Δημιουργία Συστάδων

Από τον Πίνακα 3-2 που αφορά τον επιμερισμό των συσσωρευμένων συστάδων, παρατηρούμε ότι τα δεδομένα χωρίστηκαν σε δύο συστάδες. Παρατηρούμε λοιπόν ότι η πρώτη συστάδα η οποία έχει τις περισσότερες παρατηρήσεις (520 παρατηρήσεις), έχει και την μεγαλύτερη τετραγωνική αθροιστική απόσταση των παρατηρήσεων της από τη μέση της συστάδας, αλλά τελικά την μικρότερη μέγιστη απόσταση (τετραγωνική ρίζα της απόστασης) από τη μέση της συστάδας. Το αντίστροφο συμβαίνει στη δεύτερη συστάδα που έχει τις λιγότερες παρατηρήσεις (80 παρατηρήσεις). Μεταξύ των δύο συστάδων θα επιλέξουμε αυτή που έχει χαμηλότερη μέγιστη απόσταση ως πιο συμπαγής, δηλαδή παρουσιάζει χαμηλότερη μεταβλητότητα, δηλαδή την πρώτη συστάδα.

Cluster Centroids

Variable	Cluster1	Cluster2	Grand centroid
epis_id	-0,310383	2,01749	-0,0000000
epis_def_tag	0,019435	-0,12633	-0,0000000

Πίνακας 3-3: Κέντρα των Συστάδων

Από τον Πίνακα 3-3 που αναφέρεται στα κέντρα βάρη των συστάδων με βάση τις δύο μεταβλητές που πήραμε, παρατηρούμε ότι η μεταβλητή «epis_id» εμφανίζει μεγαλύτερη ομαδοποίηση ως προς το κέντρο βάρους της πρώτης συστάδας, γιατί έχει την χαμηλότερη μέση τιμή κατά απόλυτη τιμή. Αυτό σημαίνει ότι η πρώτη συστάδα η οποία απαρτίζεται συνολικά από 520 παρατηρήσεις, περιέχει περισσότερο τιμές παρατηρήσεων από τη μεταβλητή «epis_id». Όμοια και η μεταβλητή «epis_def_tag» εμφανίζει μεγαλύτερη ομαδοποίηση ως προς το κέντρο βάρους της πρώτης.

Distances Between Cluster Centroids

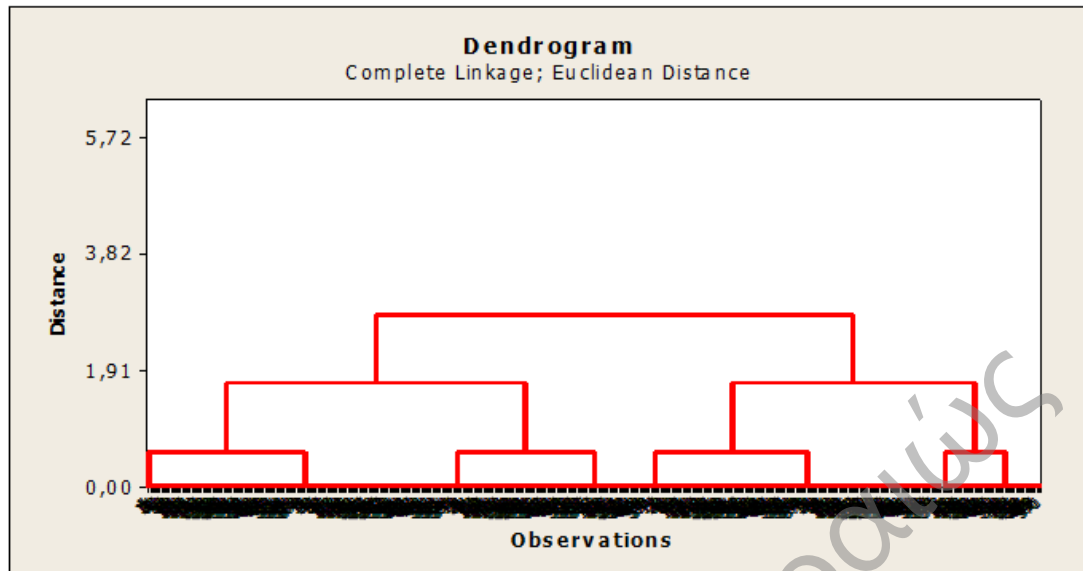
	Cluster1	Cluster2
Cluster1	0,00000	2,33243
Cluster2	2,33243	0,00000

Πίνακας 3-4: Αποστάσεις μεταξύ των κέντρων των συστάδων

Από τον Πίνακα 3-4 που αφορά τις αποστάσεις μεταξύ των κέντρων βαρών των συστάδων με βάση τις δύο μεταβλητές που πήραμε, παρατηρούμε, ότι οι συστάδες είναι μόνο δύο. Δεδομένου ότι η πρώτη συστάδα αποτελείται από 520 παρατηρήσεις, είναι συμπαγής και στη πρώτη συστάδα και οι δύο μεταβλητές συσχετίζονται (ομαδοποιούνται), συμπεραίνουμε λοιπόν με βάση την ιεραρχική συσταδοποίηση ότι οι μεταβλητές «epis_id» και «epis_def_tag» εμφανίζουν μεγάλη ομοιότητα μεταξύ τους, άρα η μία επηρεάζει την άλλη σε ποσοστό 233% (η μεταξύ τους απόσταση*100) περίπου, στο σύνολο των παρατηρήσεων (δηλαδή στο $520/600*100=87\%$ των ζευγιώντων παρατηρήσεων παρατηρούμε αλληλεπίδραση 233%). Η ταυτότητα του επεισοδίου καθορίζεται από το σχολιασμό του επεισοδίου, ως στάση ή κίνηση, και αντίστροφα. Να σημειώσουμε σε αυτό το σημείο ότι μπορούμε να «ταιριάξουμε» τη μία μεταβλητή με την άλλη αλλά δεν μπορούμε να συμπεράνουμε για το ποια κατηγορία της μιας «ταιριάζει» με την κατηγορία της άλλης.

Το αποτέλεσμα της συσταδοποίησης απεικονίζεται στο ακόλουθο δενδρόγραμμα, που δείχνει την αλληλουχία της συγχώνευσης των συστάδων και την απόσταση στην οποία η κάθε συγχώνευση πραγματοποιήθηκε.

ΔΕΝΔΡΟΓΡΑΜΜΑ



Σχήμα 3-1: Δενδρόγραμμα (epis_id- epis_def_tag)

3.2 Συσταδοποίηση epis_id- sem_traj_id

Στον ίδιο σχολιασμό καταλήγουμε μελετώντας τον Πίνακα 3.5 του Minitab, ο οποίος είναι ο πρώτος σε σειρά που εμφανίζεται και αφορά τη συγχώνευση (Amalgamation/ Agglomeration) των συστάδων. Πιο συγκεκριμένα παρατηρούμε ότι σε κάθε βήμα (step) τα δεδομένα μας ομαδοποιούνται σε δύο συστάδες που στη συνέχεια αυτές συγχωνεύονται και προκύπτει μία νέα συστάδα. Για παράδειγμα στην πρώτη δοκιμή, βάσει 100% ομοιότητας και παράλληλα μηδενικής απόστασης μεταξύ των συστάδων, δημιουργήθηκαν δύο συστάδες. Η πρώτη συστάδα αποτελείται από τις πρώτες 595 παρατηρήσεις και είναι μέρος της δεύτερης συστάδας η οποία αποτελείται από 600 παρατηρήσεις. Οι δύο αυτές συστάδες συγχωνεύονται και η νέα πλέον συστάδα (η οποία έχει αριθμό αναγνώρισης 595) απαρτίζεται μόνο από 2 παρατηρήσεις (ουσιαστικά απαρτίζεται από δύο υποσυστάδες). Η διαδικασία αυτή συνεχίζεται, ώσπου η τελευταία 599^η παρατήρηση να συμπληρώσει άθροισμα παρατηρήσεων τελευταίας συστάδας ίσο με 600, όσο είναι δηλαδή και το μέγεθος του δείγματός μας.

Cluster Analysis of Observations: epis_id; sem_traj_id

Standardized Variables, Euclidean Distance, Complete Linkage, Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	599	100,000	0,00000	595 600	595	2
2	598	100,000	0,00000	592 599	592	2
3	597	100,000	0,00000	591 598	591	2
4	596	100,000	0,00000	590 597	590	2
5	595	100,000	0,00000	589 596	589	2
6	594	100,000	0,00000	588 595	588	3
7	593	100,000	0,00000	539 594	539	2
8	592	100,000	0,00000	538 593	538	2
9	591	100,000	0,00000	587 592	587	3
10	590	100,000	0,00000	586 591	586	3
11	589	100,000	0,00000	585 590	585	3
12	588	100,000	0,00000	584 589	584	3
13	587	100,000	0,00000	583 588	583	4
14	586	100,000	0,00000	582 587	582	4
15	585	100,000	0,00000	581 586	581	4
16	584	100,000	0,00000	580 585	580	4
17	583	100,000	0,00000	579 584	579	4
18	582	100,000	0,00000	578 583	578	5
19	581	100,000	0,00000	563 582	563	5
20	580	100,000	0,00000	562 581	562	5
21	579	100,000	0,00000	561 580	561	5
22	578	100,000	0,00000	560 579	560	5
23	577	100,000	0,00000	559 578	559	6
24	576	100,000	0,00000	525 577	525	2
25	575	100,000	0,00000	524 576	524	2
26	574	100,000	0,00000	523 575	523	2
27	573	100,000	0,00000	522 574	522	2
28	572	100,000	0,00000	521 573	521	2
29	571	100,000	0,00000	427 572	427	2
30	570	100,000	0,00000	426 571	426	2
31	569	100,000	0,00000	462 570	462	2
549	51	96,050	0,28181	340 341	340	6
550	50	96,050	0,28181	470 471	470	2
551	49	96,050	0,28181	137 138	137	14
552	48	96,050	0,28181	120 121	120	6
595	5	62,368	2,68499	137 141	137	50
596	4	58,995	2,92561	1 5	1	481
597	3	52,603	3,38167	29 415	29	69
598	2	31,020	4,92161	1 137	1	531
599	1	0,000	7,13481	1 29	1	600

Πίνακας 3-5: Βήματα Συγχώνευσης των Συστάδων

Επιπλέον από τον Πίνακα 3-5 παρατηρούμε ότι από την 549^η επανάληψη και μετά η ομοιότητα και η απόσταση αρχίζουν σταδιακά να μειώνονται και να αυξάνονται αντίστοιχα. Οι μεταβολές αυτές αρχίζουν να γίνονται μεγαλύτερες από την 597^η επανάληψη και μετά. Θα μπορούσαμε λοιπόν να πούμε ότι οι τέσσερις συστάδες είναι λογικά επαρκείς για την οριστική

συγχώνευση (θα μπορούσαμε δηλαδή να σταματήσουμε τη διαδικασία στις τέσσερις συστάδες).

Final Partition
Number of clusters: 4

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	481	377,275	0,814981	2,05857
Cluster2	63	41,901	0,713915	1,52321
Cluster3	50	34,756	0,718150	1,87831
Cluster4	6	0,543	0,250494	0,51664

Πίνακας 3-6: Δημιουργία Συστάδων

Από τον Πίνακα 3.6 , που αφορά τον επιμερισμό των συσσωρευμένων συστάδων, παρατηρούμε αρχικά ότι τα δεδομένα χωρίστηκαν σε τέσσερις συστάδες. Παρατηρούμε λοιπόν ότι η πρώτη συστάδα η οποία έχει τις περισσότερες παρατηρήσεις, 481 παρατηρήσεις, έχει και την μεγαλύτερη απόσταση των παρατηρήσεων της από τη μέση ή διαφορετικά από το κέντρο βάρους (centroid) της συστάδας (Average Distance from centroid), γεγονός που μας οδηγεί στο να συμπεράνουμε ότι η συγκεκριμένη συστάδα δεν είναι συμπαγής, δηλαδή ομαδοποιούνται χαλαρά τα δεδομένα μας σε αυτή. Αντίθετα η τέταρτη συστάδα είναι η πιο συμπαγής, αφού έχει τη μικρότερη απόσταση των παρατηρήσεων της από τη μέση της συστάδας, και απαρτίζεται μόνο από 6 παρατηρήσεις.

Cluster Centroids					
Variable	Cluster1	Cluster2	Cluster3	Cluster4	Grand centroid
epis_id	-0,280240	1,94631	-0,24470	4,06880	-0,0000000
sem_traj_id	-0,204996	-0,42001	2,57390	-0,60519	-0,0000000

Πίνακας 3-7: Κέντρα των Συστάδων

Από τον Πίνακα 3-7 που αφορά τα κέντρα βάρη των συστάδων με βάση τις δύο μεταβλητές που πήραμε, παρατηρούμε ότι η μεταβλητή «epis_id» (η ταυτότητα του επεισοδίου για κάθε τροχιά) εμφανίζει μεγαλύτερη ομαδοποίηση ως προς το κέντρο βάρους της

στην τρίτη συστάδα, γιατί έχει την χαμηλότερη μέση τιμή κατά απόλυτη τιμή. Αυτό σημαίνει ότι η τρίτη συστάδα η οποία απαρτίζεται συνολικά από 50 παρατηρήσεις, περιέχει περισσότερες τιμές παρατηρήσεων από τη μεταβλητή «epis_id». Επιπλέον η μεταβλητή «sem_traj_id» εμφανίζει μεγαλύτερη ομαδοποίηση ως προς το κέντρο βάρους της, στην πρώτη συστάδα η οποία όμως παρουσιάζει και την μεγαλύτερη μεταβλητότητα αφού σε σχέση με τις υπόλοιπες τρεις παρουσιάζει τη μεγαλύτερη απόσταση των παρατηρήσεων της από το μέσο της συστάδας.

Να σημειώσουμε ότι τα αρνητικά πρόσημα δηλώνουν κάτω από τη μέση της συστάδας και τα θετικά πρόσημα το αντίστροφο, δηλαδή για παράδειγμα οι τιμές της epis_id απέχουν από το κέντρο βάρους του cluster 3 απόσταση ίση με -0,24470 επομένως οι τιμές της βρίσκονται κάτω από το μέσο του κέντρο βάρους της συστάδας.

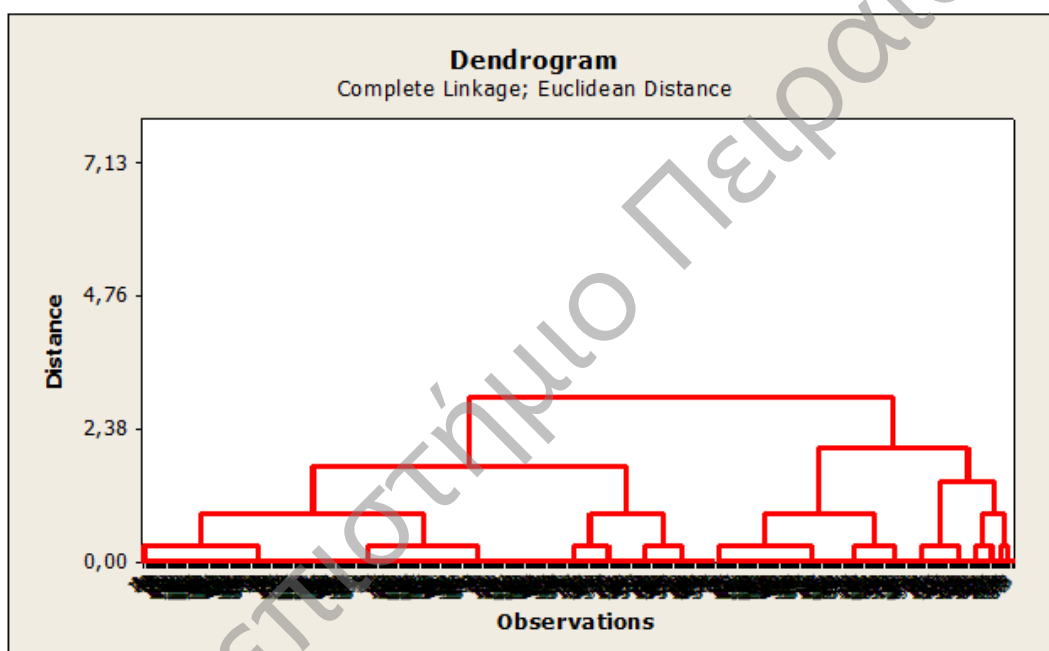
Distances Between Cluster Centroids				
	Cluster1	Cluster2	Cluster3	Cluster4
Cluster1	0,00000	2,23691	2,77912	4,36742
Cluster2	2,23691	0,00000	3,71000	2,13055
Cluster3	2,77912	3,71000	0,00000	5,35845
Cluster4	4,36742	2,13055	5,35845	0,00000

Πίνακας 3-8: Αποστάσεις μεταξύ των κέντρων των συστάδων

Από τον Πίνακα 3-8 που αφορά τις αποστάσεις μεταξύ των κέντρων βαρών των συστάδων με βάση τις δύο μεταβλητές που πήραμε, παρατηρούμε ότι η μικρότερη απόσταση μεταξύ δύο συστάδων βρίσκεται μεταξύ της πρώτης και δεύτερης συστάδας. Αυτό μπορεί να φάνει περίεργο σε κάποιον που μπορεί να μελετήσει τον Πίνακα 3.8, καθώς υπάρχουν και αποστάσεις μικρότερες από τη συγκεκριμένη. Επιλέξαμε να μελετήσουμε τις αποστάσεις της πρώτης συστάδας σε σχέση με τις υπόλοιπες τρεις καθώς είναι αυτή που παρουσιάζει τη μεγαλύτερη μεταβλητότητα. Επιπλέον η πρώτη συστάδα επηρεάζει καθοριστικά τη συσταδοποίηση καθώς είναι αυτή που περιέχει το μεγαλύτερο ποσοστό των παρατηρήσεων (περισσότερο από το 50%). Εάν η πρώτη συστάδα δεν περιείχε τόσο μεγάλο ποσοστό παρατηρήσεων θα κάναμε περαιτέρω ανάλυση και για τις υπόλοιπες. Από τη στιγμή που η μεγαλύτερη συστάδα αποδεικνύεται «ελαττωματική» οδηγούμαστε στο συμπέρασμα ότι δεν έχουμε μία καλώς ορισμένη συσταδοποίηση της «sem_traj_id». Με άλλα λόγια δεν παράγεται ουσιαστικό ερευνητικό αποτέλεσμα.

Σύμφωνα λοιπόν με τα παραπάνω καταλήγουμε στο συμπέρασμα ότι αφού οι δύο πιο κοντινές συστάδες είναι η συστάδα 1 και η συστάδα 2, όπου και στις δύο υπερτερούν οι τιμές της μεταβλητής «sem_traj_id» (ταυτότητα της τροχιάς για κάθε χρήστη), με βάση την ιεραρχική συσταδοποίηση οι μεταβλητές «epis_id» (ταυτότητα του επεισοδίου) και «sem_traj_id» δεν εμφανίζουν κανένα σημείο τομής-ομοιότητας μεταξύ τους, άρα η μία δεν αλληλεπιδρά στην άλλη. Με άλλα λόγια η ταυτότητα της τροχιάς του χρήστη δεν αλληλεπιδρά με την ταυτότητα του επεισοδίου.

Στο δενδρόγραμμα που ακολουθεί εμφανίζονται οι πληροφορίες του πίνακα συγχώνευσης με τη μορφή ενός διαγράμματος δέντρου.



Σχήμα 3-2: Δενδρόγραμμα (epis_id- sem_traj_id)

3.3 Συσταδοποίηση sem_traj_id- avg_speed

Από τον Πίνακα 3.9 του Minitab, ο οποίος είναι ο πρώτος σε σειρά που εμφανίζεται και αφορά τη συγχώνευση (Amalgamation/ Agglomeration) των συστάδων, παρατηρούμε ότι σε κάθε βήμα (step) τα δεδομένα μας ομαδοποιούνται σε δύο συστάδες που στη συνέχεια αυτές συγχωνεύονται και προκύπτει μία νέα συστάδα. Για παράδειγμα στην πρώτη δοκιμή, βάσει

100% ομοιότητας και παράλληλα μηδενικής απόστασης μεταξύ των συστάδων, δημιουργήθηκαν δύο συστάδες . Η πρώτη συστάδα αποτελείται από τις πρώτες 160 παρατηρήσεις και είναι μέρος της δεύτερης συστάδας η οποία αποτελείται από 594 παρατηρήσεις. Οι δύο αυτές συστάδες συγχωνεύονται και η νέα πλέον συστάδα (η οποία έχει αριθμό αναγνώρισης 160) απαρτίζεται μόνο από 2 παρατηρήσεις (ουσιαστικά απαρτίζεται από δύο υποσυστάδες). Η διαδικασία αυτή συνεχίζεται, ώσπου η τελευταία 599^η παρατήρηση να συμπληρώσει άθροισμα παρατηρήσεων τελευταίας συστάδας ίσο με 600 , όσο είναι δηλαδή και το μέγεθος του δείγματός μας.

Να σημειωθεί ότι από την 155^η επανάληψη και μετά η ομοιότητα και η απόσταση μεταξύ των συστάδων μειώνεται και αυξάνεται αντίστοιχα. Παρατηρείται επομένως μεγάλη ανομοιογένεια στο δείγμα κατά ποσοστό 74,33% ($\{[600-154]600\} * 100$) και από πολύ νωρίς θα μπορούσαμε να πούμε ότι εμφανίζονται «προβληματικές» συστάδες, δηλαδή όχι συμπαγής. Παρατηρείται μεγάλη δυσαναλογία τιμών στο σύνολο του δείγματος.

Cluster Analysis of Observations: sem_traj_id; avg_speed

Standardized Variables, Euclidean Distance, Complete Linkage, Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	599	100,000	0,0000	160 594	160	2
2	598	100,000	0,0000	430 434	430	2
3	597	100,000	0,0000	533 543	533	2
4	596	100,000	0,0000	258 572	258	2
5	595	100,000	0,0000	583 587	583	2
6	594	100,000	0,0000	78 90	78	2
7	593	100,000	0,0000	44 202	44	2
8	592	100,000	0,0000	29 72	29	2
9	591	100,000	0,0000	150 234	150	2
10	590	100,000	0,0000	96 104	96	2
11	589	100,000	0,0000	559 563	559	2
12	588	100,000	0,0000	178 186	178	2
13	587	100,000	0,0000	105 479	105	2
14	586	100,000	0,0000	56 430	56	3
15	585	100,000	0,0000	349 353	349	2
16	584	100,000	0,0000	11 284	11	2
17	583	100,000	0,0000	150 558	150	3
18	582	100,000	0,0000	95 553	95	2
19	581	100,000	0,0000	78 548	78	3
155	445	99,999	0,0001	8 132	8	2
156	444	99,999	0,0001	84 588	84	2
157	443	99,999	0,0001	393 427	393	2
158	442	99,999	0,0001	75 597	75	3
159	441	99,999	0,0001	141 463	141	8

Πίνακας 3-9: Βήματα Συγχώνευσης των Συστάδων

Final Partition
 Number of clusters: 4

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	523	484,405	0,892537	1,90343
Cluster2	1	0,000	0,000000	0,00000
Cluster3	26	15,111	0,641390	1,43607
Cluster4	50	49,110	0,922209	1,61683

Πίνακας 3-10: Δημιουργία Συστάδων

Από τον Πίνακα 3-10 που αφορά τον επιμερισμό των συσσωρευμένων συστάδων, παρατηρούμε ότι τα δεδομένα χωρίστηκαν σε τέσσερις συστάδες όσο είναι και ο αριθμός κατηγοριών της μεταβλητής *sem_traj_id*. Επιπλέον παρατηρούμε ότι η τρίτη συστάδα η οποία έχει τις λιγότερες παρατηρήσεις (26 παρατηρήσεις) έχει και την μικρότερη απόσταση των παρατηρήσεων της από τη μέση της συστάδας, άρα είναι συμπαγής, δηλαδή ομαδοποιούνται με χαμηλή μεταβλητότητα τα δεδομένα μας σε αυτή τη συστάδα. Αντίθετα η πρώτη συστάδα παρουσιάζει τη μεγαλύτερη μεταβλητότητα παρόλο που απαρτίζεται από τις περισσότερες παρατηρήσεις (523 παρατηρήσεις).

Cluster Centroids					
Variable	Cluster1	Cluster2	Cluster3	Cluster4	Grand centroid
<i>sem_traj_id</i>	-0,228769	-0,6052	-0,32475	2,57390	-0,0000000
<i>avg_speed</i>	-0,133672	11,9944	2,27365	-0,02398	-0,0000000

Πίνακας 3-11: Κέντρα των Συστάδων

Από τον Πίνακα 3-11, που αφορά τα κέντρα βάρη των συστάδων με βάση τις δύο μεταβλητές που πήραμε, παρατηρούμε ότι η μεταβλητή «*avg_speed*» εμφανίζει μεγαλύτερη ομαδοποίηση ως προς το κέντρο βάρους της, στην τέταρτη συστάδα, γιατί έχει την χαμηλότερη μέση τιμή κατά απόλυτη τιμή. Αυτό σημαίνει ότι η τέταρτη συστάδα, η οποία απαρτίζεται συνολικά από 50 παρατηρήσεις, περιέχει περισσότερο τιμές παρατηρήσεων από τη μεταβλητή «*avg_speed*» και παρουσιάζει σχετικά χαμηλή μεταβλητότητα ως είναι δηλαδή συμπαγής. Αντίθετα η μεταβλητή «*sem_traj_id*» εμφανίζει μεγαλύτερη ομαδοποίηση ως προς

το κέντρο βάρους της, στην πρώτη συστάδα η οποία όμως έχει την μεγαλύτερη μεταβλητότητα.

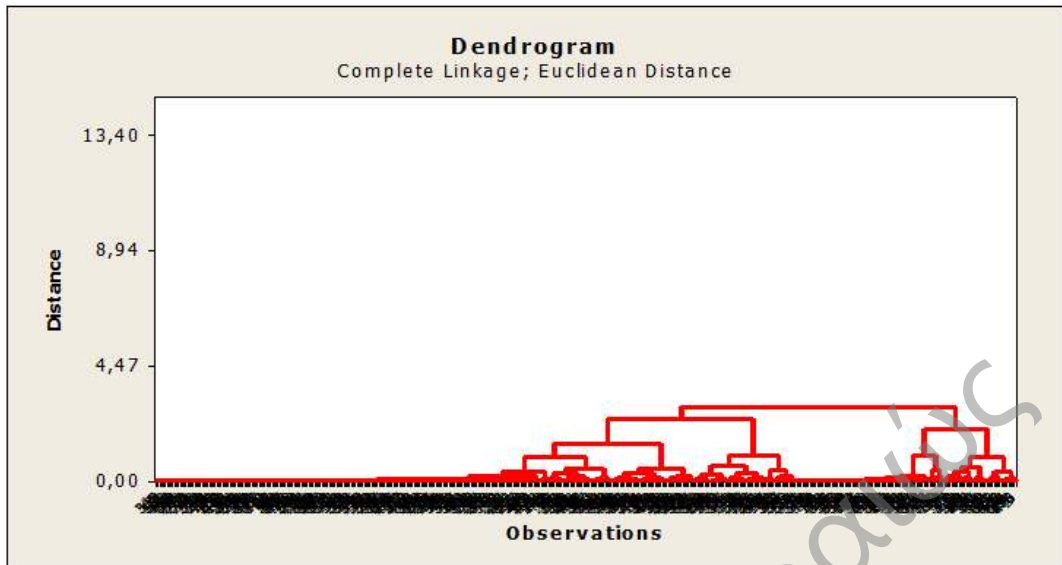
Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4
Cluster1	0,0000	12,1339	2,40923	2,8048
Cluster2	12,1339	0,0000	9,72481	12,4317
Cluster3	2,4092	9,7248	0,00000	3,6988
Cluster4	2,8048	12,4317	3,69882	0,0000

Πίνακας 3-12: Αποστάσεις μεταξύ των κέντρων των συστάδων

Από τον Πίνακα 3-12 κατά σειρά, που αφορά τις αποστάσεις μεταξύ των κέντρων βαρών των συστάδων με βάση τις δύο μεταβλητές που πήραμε, παρατηρούμε ότι η μικρότερη απόσταση μεταξύ δύο συστάδων βρίσκεται μεταξύ της πρώτης και της τρίτης συστάδας. Δεδομένου ότι η τρίτη συστάδα αποτελείται από μόνο 26 παρατηρήσεις και η πρώτη συστάδα που αποτελείται από 523 παρατηρήσεις, παρουσιάζει τη μεγαλύτερη μεταβλητότητα και σε αυτήν υπερτερούν επίσης οι τιμές της μεταβλητής «sem_traj_id», συμπεραίνουμε ότι με βάση την ιεραρχική συσταδοποίηση οι μεταβλητές «avg_speed» και «sem_traj_id» δεν εμφανίζουν κανένα σημείο τομής-ομοιότητας μεταξύ τους, άρα η μία δεν επηρεάζει την άλλη.

Το αποτέλεσμα της συσταδοποίησης απεικονίζεται στο παρακάτω δενδρόγραμμα, που δείχνει την αλληλουχία της συγχώνευσης των συστάδων και την απόσταση στην οποία πραγματοποιήθηκε η κάθε συγχώνευση.



Σχήμα 3-3: Δενδρόγραμμα (sem_traj_id- avg_speed)

Πανεπιστήμιο Πειραιώς

ΚΕΦΑΛΑΙΟ 4

ΣΥΜΠΕΡΑΣΜΑΤΑ

Τα διαθέσιμα δεδομένα που χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία συλλέχθηκαν σε μία πόλη του Βελγίου και προέρχονται από δύο πηγές δεδομένων, τα «δεδομένα GPS» τα οποία αποτελούν εγγραφές GPS ανά χρήστη και πιο συγκεκριμένα προέρχονται από τις ενεργοποιημένες συσκευές GPS και τα «Ημερολόγια» κίνησης (Diaries) στα οποία κάθε χρήστης καταγράφει τη σημασιολογική πληροφορία για το ταξίδι του (trip).

Το σύνολο των κινούμενων αντικειμένων χρησιμοποιήθηκε προκειμένου να μελετήσουμε τις πιθανές αλληλεπιδράσεις που μπορεί να έχουν οι μεταβλητές που θέλουμε να μελετήσουμε. Για το λόγο αυτό πραγματοποιήθηκαν τρεις διαφορετικές συσταδοποιήσεις στα ακόλουθα ζεύγη μεταβλητών: `epis_id- sem-traj-id`, `epis_id- epis_def_tag` και `sem_traj_id- avg_speed`.

Τα αποτελέσματα που πήραμε εφαρμόζοντας τον ιεραρχικό αλγόριθμο συσταδοποίησης στα ζεύγη `epis_id- sem-traj-id` και `sem_traj_id- avg_speed` δεν μας οδήγησαν σε ουσιώδη στατιστικά συμπεράσματα αφού οι μεταβλητές αυτές δεν εμφάνισαν σημεία ομοιότητας, δηλαδή δεν αλληλεπιδρούν μεταξύ τους. Αντίθετα όσον αναφορά τη συσταδοποίηση που πραγματοποιήθηκε ανάμεσα στο ζεύγος των μεταβλητών `epis_id- epis_def_tag` καταλήξαμε στο συμπέρασμα ότι εμφανίζουν μεγάλο ποσοστό αλληλεπίδρασης δηλαδή η ταυτότητα του επεισοδίου (`epis_id`) καθορίζεται από το σχολιασμό του (`epis_def_tag`) ως στάση ή κίνηση και αντίστροφα. Επομένως θα μπορούσαμε να πούμε ότι σε μία ενδεχόμενη μελλοντική χρήση του συγκεκριμένου συνόλου δεδομένων για περαιτέρω ανάλυση, όπως για παράδειγμα είναι η πρόβλεψη, οι παραπάνω αυτές μεταβλητές θα μπορούσαν να αξιοποιηθούν.

Για να βελτιώσουμε την υπολογιστική πολυπλοκότητα και την ακρίβεια των συμπερασμάτων μας και δεδομένου ότι υπάρχει περίπτωση να θέλουμε μελλοντικά να αναλύσουμε μεγαλύτερα μεγέθη δείγματος, μπορούμε εναλλακτικά να τρέξουμε αλγόριθμους τεχνητής νοημοσύνης και πιο συγκεκριμένα χασοτικά μοντέλα νευρωνικών δικτύων ή ασαφούς λογικής.

Πανεπιστήμιο Πειραιώς

ΠΑΡΑΡΤΗΜΑ ΠΙΝΑΚΩΝ

Πίνακας συγχώνευσης συστάδων στη συσταδοποίηση των μεταβλητών epis_id- epis_def_tag

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	599	100,000	0,00000	595 600	595	2
2	598	100,000	0,00000	592 599	592	2
3	597	100,000	0,00000	591 598	591	2
4	596	100,000	0,00000	590 597	590	2
5	595	100,000	0,00000	589 596	589	2
6	594	100,000	0,00000	588 595	588	3
7	593	100,000	0,00000	570 594	570	2
8	592	100,000	0,00000	569 593	569	2
9	591	100,000	0,00000	587 592	587	3
10	590	100,000	0,00000	586 591	586	3
11	589	100,000	0,00000	585 590	585	3
12	588	100,000	0,00000	584 589	584	3
13	587	100,000	0,00000	583 588	583	4
14	586	100,000	0,00000	582 587	582	4
15	585	100,000	0,00000	581 586	581	4
16	584	100,000	0,00000	580 585	580	4
17	583	100,000	0,00000	579 584	579	4
18	582	100,000	0,00000	578 583	578	5
19	581	100,000	0,00000	577 582	577	5
20	580	100,000	0,00000	576 581	576	5
21	579	100,000	0,00000	575 580	575	5
22	578	100,000	0,00000	574 579	574	5
23	577	100,000	0,00000	573 578	573	6
24	576	100,000	0,00000	568 577	568	6
25	575	100,000	0,00000	567 576	567	6
26	574	100,000	0,00000	566 575	566	6
27	573	100,000	0,00000	565 574	565	6
28	572	100,000	0,00000	564 573	564	7
29	571	100,000	0,00000	541 572	541	2
30	570	100,000	0,00000	540 571	540	2
31	569	100,000	0,00000	539 570	539	3
32	568	100,000	0,00000	538 569	538	3
33	567	100,000	0,00000	563 568	563	7
34	566	100,000	0,00000	562 567	562	7
35	565	100,000	0,00000	561 566	561	7
36	564	100,000	0,00000	560 565	560	7
37	563	100,000	0,00000	559 564	559	8
38	562	100,000	0,00000	558 563	558	8
39	561	100,000	0,00000	557 562	557	8
40	560	100,000	0,00000	556 561	556	8
41	559	100,000	0,00000	555 560	555	8
42	558	100,000	0,00000	554 559	554	9
43	557	100,000	0,00000	553 558	553	9
44	556	100,000	0,00000	552 557	552	9
45	555	100,000	0,00000	551 556	551	9
46	554	100,000	0,00000	550 555	550	9
47	553	100,000	0,00000	549 554	549	10
48	552	100,000	0,00000	548 553	548	10
49	551	100,000	0,00000	547 552	547	10
50	550	100,000	0,00000	546 551	546	10
51	549	100,000	0,00000	545 550	545	10
52	548	100,000	0,00000	544 549	544	11

53	547	100,000	0,00000	537	548	537	11
54	546	100,000	0,00000	536	547	536	11
55	545	100,000	0,00000	535	546	535	11
56	544	100,000	0,00000	534	545	534	11
57	543	100,000	0,00000	533	544	533	12
58	542	100,000	0,00000	499	543	499	2
59	541	100,000	0,00000	498	542	498	2
60	540	100,000	0,00000	497	541	497	3
61	539	100,000	0,00000	496	540	496	3
62	538	100,000	0,00000	532	539	532	4
63	537	100,000	0,00000	531	538	531	4
64	536	100,000	0,00000	530	537	530	12
65	535	100,000	0,00000	529	536	529	12
66	534	100,000	0,00000	528	535	528	12
67	533	100,000	0,00000	527	534	527	12
68	532	100,000	0,00000	526	533	526	13
69	531	100,000	0,00000	495	532	495	5
70	530	100,000	0,00000	494	531	494	5
71	529	100,000	0,00000	525	530	525	13
72	528	100,000	0,00000	524	529	524	13
73	527	100,000	0,00000	523	528	523	13
74	526	100,000	0,00000	522	527	522	13
75	525	100,000	0,00000	521	526	521	14
76	524	100,000	0,00000	520	525	520	14
77	523	100,000	0,00000	519	524	519	14
78	522	100,000	0,00000	518	523	518	14
79	521	100,000	0,00000	517	522	517	14
80	520	100,000	0,00000	516	521	516	15
81	519	100,000	0,00000	515	520	515	15
82	518	100,000	0,00000	514	519	514	15
83	517	100,000	0,00000	513	518	513	15
84	516	100,000	0,00000	512	517	512	15
85	515	100,000	0,00000	511	516	511	16
86	514	100,000	0,00000	493	515	493	16
87	513	100,000	0,00000	492	514	492	16
88	512	100,000	0,00000	510	513	510	16
89	511	100,000	0,00000	509	512	509	16
90	510	100,000	0,00000	508	511	508	17
91	509	100,000	0,00000	491	510	491	17
92	508	100,000	0,00000	490	509	490	17
93	507	100,000	0,00000	489	508	489	18
94	506	100,000	0,00000	416	507	416	2
95	505	100,000	0,00000	415	506	415	2
96	504	100,000	0,00000	414	505	414	2
97	503	100,000	0,00000	413	504	413	2
98	502	100,000	0,00000	412	503	412	2
99	501	100,000	0,00000	411	502	411	2
100	500	100,000	0,00000	410	501	410	2
101	499	100,000	0,00000	409	500	409	2
102	498	100,000	0,00000	429	499	429	3
103	497	100,000	0,00000	428	498	428	3
104	496	100,000	0,00000	478	497	478	4
105	495	100,000	0,00000	477	496	477	4
106	494	100,000	0,00000	476	495	476	6
107	493	100,000	0,00000	475	494	475	6
108	492	100,000	0,00000	488	493	488	17
109	491	100,000	0,00000	487	492	487	17
110	490	100,000	0,00000	486	491	486	18
111	489	100,000	0,00000	485	490	485	18
112	488	100,000	0,00000	484	489	484	19
113	487	100,000	0,00000	483	488	483	18
114	486	100,000	0,00000	482	487	482	18
115	485	100,000	0,00000	481	486	481	19
116	484	100,000	0,00000	480	485	480	19
117	483	100,000	0,00000	479	484	479	20

118	482	100,000	0,00000	474	483	474	19
119	481	100,000	0,00000	473	482	473	19
120	480	100,000	0,00000	472	481	472	20
121	479	100,000	0,00000	471	480	471	20
122	478	100,000	0,00000	470	479	470	21
123	477	100,000	0,00000	450	478	450	5
124	476	100,000	0,00000	449	477	449	5
125	475	100,000	0,00000	469	476	469	7
126	474	100,000	0,00000	468	475	468	7
127	473	100,000	0,00000	467	474	467	20
128	472	100,000	0,00000	466	473	466	20
129	471	100,000	0,00000	465	472	465	21
130	470	100,000	0,00000	464	471	464	21
131	469	100,000	0,00000	463	470	463	22
132	468	100,000	0,00000	462	469	462	8
133	467	100,000	0,00000	461	468	461	8
134	466	100,000	0,00000	460	467	460	21
135	465	100,000	0,00000	459	466	459	21
136	464	100,000	0,00000	458	465	458	22
137	463	100,000	0,00000	457	464	457	22
138	462	100,000	0,00000	456	463	456	23
139	461	100,000	0,00000	448	462	448	9
140	460	100,000	0,00000	447	461	447	9
141	459	100,000	0,00000	455	460	455	22
142	458	100,000	0,00000	454	459	454	22
143	457	100,000	0,00000	453	458	453	23
144	456	100,000	0,00000	452	457	452	23
145	455	100,000	0,00000	451	456	451	24
146	454	100,000	0,00000	446	455	446	23
147	453	100,000	0,00000	445	454	445	23
148	452	100,000	0,00000	444	453	444	24
149	451	100,000	0,00000	443	452	443	24
150	450	100,000	0,00000	442	451	442	25
151	449	100,000	0,00000	427	450	427	6
152	448	100,000	0,00000	426	449	426	6
153	447	100,000	0,00000	441	448	441	10
154	446	100,000	0,00000	440	447	440	10
155	445	100,000	0,00000	439	446	439	24
156	444	100,000	0,00000	438	445	438	24
157	443	100,000	0,00000	437	444	437	25
158	442	100,000	0,00000	436	443	436	25
159	441	100,000	0,00000	435	442	435	26
160	440	100,000	0,00000	425	441	425	11
161	439	100,000	0,00000	424	440	424	11
162	438	100,000	0,00000	434	439	434	25
163	437	100,000	0,00000	433	438	433	25
164	436	100,000	0,00000	432	437	432	26
165	435	100,000	0,00000	431	436	431	26
166	434	100,000	0,00000	430	435	430	27
167	433	100,000	0,00000	423	434	423	26
168	432	100,000	0,00000	422	433	422	26
169	431	100,000	0,00000	421	432	421	27
170	430	100,000	0,00000	420	431	420	27
171	429	100,000	0,00000	419	430	419	28
172	428	100,000	0,00000	408	429	408	4
173	427	100,000	0,00000	407	428	407	4
174	426	100,000	0,00000	406	427	406	7
175	425	100,000	0,00000	405	426	405	7
176	424	100,000	0,00000	404	425	404	12
177	423	100,000	0,00000	403	424	403	12
178	422	100,000	0,00000	402	423	402	27
179	421	100,000	0,00000	401	422	401	27
180	420	100,000	0,00000	400	421	400	28
181	419	100,000	0,00000	399	420	399	28
182	418	100,000	0,00000	398	419	398	29

183	417	100,000	0,00000	343	414	343	3
184	416	100,000	0,00000	342	413	342	3
185	415	100,000	0,00000	341	412	341	3
186	414	100,000	0,00000	340	411	340	3
187	413	100,000	0,00000	380	410	380	3
188	412	100,000	0,00000	379	409	379	3
189	411	100,000	0,00000	378	408	378	5
190	410	100,000	0,00000	377	407	377	5
191	409	100,000	0,00000	376	406	376	8
192	408	100,000	0,00000	375	405	375	8
193	407	100,000	0,00000	392	404	392	13
194	406	100,000	0,00000	391	403	391	13
195	405	100,000	0,00000	397	402	397	28
196	404	100,000	0,00000	396	401	396	28
197	403	100,000	0,00000	395	400	395	29
198	402	100,000	0,00000	394	399	394	29
199	401	100,000	0,00000	393	398	393	30
200	400	100,000	0,00000	390	397	390	29
201	399	100,000	0,00000	389	396	389	29
202	398	100,000	0,00000	388	395	388	30
203	397	100,000	0,00000	387	394	387	30
204	396	100,000	0,00000	386	393	386	31
205	395	100,000	0,00000	374	392	374	14
206	394	100,000	0,00000	373	391	373	14
207	393	100,000	0,00000	385	390	385	30
208	392	100,000	0,00000	384	389	384	30
209	391	100,000	0,00000	383	388	383	31
210	390	100,000	0,00000	382	387	382	31
211	389	100,000	0,00000	381	386	381	32
212	388	100,000	0,00000	372	385	372	31
213	387	100,000	0,00000	371	384	371	31
214	386	100,000	0,00000	370	383	370	32
215	385	100,000	0,00000	369	382	369	32
216	384	100,000	0,00000	368	381	368	33
217	383	100,000	0,00000	339	380	339	4
218	382	100,000	0,00000	338	379	338	4
219	381	100,000	0,00000	337	378	337	6
220	380	100,000	0,00000	336	377	336	6
221	379	100,000	0,00000	367	376	367	9
222	378	100,000	0,00000	366	375	366	9
223	377	100,000	0,00000	365	374	365	15
224	376	100,000	0,00000	364	373	364	15
225	375	100,000	0,00000	363	372	363	32
226	374	100,000	0,00000	362	371	362	32
227	373	100,000	0,00000	361	370	361	33
228	372	100,000	0,00000	360	369	360	33
229	371	100,000	0,00000	359	368	359	34
230	370	100,000	0,00000	335	367	335	10
231	369	100,000	0,00000	334	366	334	10
232	368	100,000	0,00000	333	365	333	16
233	367	100,000	0,00000	332	364	332	16
234	366	100,000	0,00000	358	363	358	33
235	365	100,000	0,00000	357	362	357	33
236	364	100,000	0,00000	356	361	356	34
237	363	100,000	0,00000	355	360	355	34
238	362	100,000	0,00000	354	359	354	35
239	361	100,000	0,00000	353	358	353	34
240	360	100,000	0,00000	352	357	352	34
241	359	100,000	0,00000	351	356	351	35
242	358	100,000	0,00000	350	355	350	35
243	357	100,000	0,00000	349	354	349	36
244	356	100,000	0,00000	348	353	348	35
245	355	100,000	0,00000	347	352	347	35
246	354	100,000	0,00000	346	351	346	36
247	353	100,000	0,00000	345	350	345	36

248	352	100,000	0,00000	344	349	344	37
249	351	100,000	0,00000	331	348	331	36
250	350	100,000	0,00000	330	347	330	36
251	349	100,000	0,00000	329	346	329	37
252	348	100,000	0,00000	328	345	328	37
253	347	100,000	0,00000	327	344	327	38
254	346	100,000	0,00000	279	339	279	5
255	345	100,000	0,00000	278	338	278	5
256	344	100,000	0,00000	326	337	326	7
257	343	100,000	0,00000	325	336	325	7
258	342	100,000	0,00000	324	335	324	11
259	341	100,000	0,00000	323	334	323	11
260	340	100,000	0,00000	322	333	322	17
261	339	100,000	0,00000	321	332	321	17
262	338	100,000	0,00000	320	331	320	37
263	337	100,000	0,00000	319	330	319	37
264	336	100,000	0,00000	318	329	318	38
265	335	100,000	0,00000	317	328	317	38
266	334	100,000	0,00000	316	327	316	39
267	333	100,000	0,00000	277	326	277	8
268	332	100,000	0,00000	276	325	276	8
269	331	100,000	0,00000	275	324	275	12
270	330	100,000	0,00000	274	323	274	12
271	329	100,000	0,00000	273	322	273	18
272	328	100,000	0,00000	272	321	272	18
273	327	100,000	0,00000	315	320	315	38
274	326	100,000	0,00000	314	319	314	38
275	325	100,000	0,00000	313	318	313	39
276	324	100,000	0,00000	312	317	312	39
277	323	100,000	0,00000	311	316	311	40
278	322	100,000	0,00000	310	315	310	39
279	321	100,000	0,00000	309	314	309	39
280	320	100,000	0,00000	308	313	308	40
281	319	100,000	0,00000	307	312	307	40
282	318	100,000	0,00000	306	311	306	41
283	317	100,000	0,00000	299	310	299	40
284	316	100,000	0,00000	298	309	298	40
285	315	100,000	0,00000	305	308	305	41
286	314	100,000	0,00000	304	307	304	41
287	313	100,000	0,00000	303	306	303	42
288	312	100,000	0,00000	302	305	302	42
289	311	100,000	0,00000	301	304	301	42
290	310	100,000	0,00000	300	303	300	43
291	309	100,000	0,00000	297	302	297	43
292	308	100,000	0,00000	296	301	296	43
293	307	100,000	0,00000	295	300	295	44
294	306	100,000	0,00000	294	299	294	41
295	305	100,000	0,00000	293	298	293	41
296	304	100,000	0,00000	292	297	292	44
297	303	100,000	0,00000	291	296	291	44
298	302	100,000	0,00000	290	295	290	45
299	301	100,000	0,00000	289	294	289	42
300	300	100,000	0,00000	288	293	288	42
301	299	100,000	0,00000	287	292	287	45
302	298	100,000	0,00000	286	291	286	45
303	297	100,000	0,00000	285	290	285	46
304	296	100,000	0,00000	284	289	284	43
305	295	100,000	0,00000	283	288	283	43
306	294	100,000	0,00000	282	287	282	46
307	293	100,000	0,00000	281	286	281	46
308	292	100,000	0,00000	280	285	280	47
309	291	100,000	0,00000	271	284	271	44
310	290	100,000	0,00000	270	283	270	44
311	289	100,000	0,00000	269	282	269	47
312	288	100,000	0,00000	268	281	268	47

313	287	100,000	0,00000	267	280	267	48
314	286	100,000	0,00000	218	279	218	6
315	285	100,000	0,00000	217	278	217	6
316	284	100,000	0,00000	229	277	229	9
317	283	100,000	0,00000	228	276	228	9
318	282	100,000	0,00000	253	275	253	13
319	281	100,000	0,00000	252	274	252	13
320	280	100,000	0,00000	251	273	251	19
321	279	100,000	0,00000	250	272	250	19
322	278	100,000	0,00000	263	271	263	45
323	277	100,000	0,00000	262	270	262	45
324	276	100,000	0,00000	266	269	266	48
325	275	100,000	0,00000	265	268	265	48
326	274	100,000	0,00000	264	267	264	49
327	273	100,000	0,00000	261	266	261	49
328	272	100,000	0,00000	260	265	260	49
329	271	100,000	0,00000	259	264	259	50
330	270	100,000	0,00000	258	263	258	46
331	269	100,000	0,00000	257	262	257	46
332	268	100,000	0,00000	256	261	256	50
333	267	100,000	0,00000	255	260	255	50
334	266	100,000	0,00000	254	259	254	51
335	265	100,000	0,00000	249	258	249	47
336	264	100,000	0,00000	248	257	248	47
337	263	100,000	0,00000	247	256	247	51
338	262	100,000	0,00000	246	255	246	51
339	261	100,000	0,00000	245	254	245	52
340	260	100,000	0,00000	227	253	227	14
341	259	100,000	0,00000	226	252	226	14
342	258	100,000	0,00000	241	251	241	20
343	257	100,000	0,00000	240	250	240	20
344	256	100,000	0,00000	239	249	239	48
345	255	100,000	0,00000	238	248	238	48
346	254	100,000	0,00000	244	247	244	52
347	253	100,000	0,00000	243	246	243	52
348	252	100,000	0,00000	242	245	242	53
349	251	100,000	0,00000	237	244	237	53
350	250	100,000	0,00000	236	243	236	53
351	249	100,000	0,00000	235	242	235	54
352	248	100,000	0,00000	225	241	225	21
353	247	100,000	0,00000	224	240	224	21
354	246	100,000	0,00000	234	239	234	49
355	245	100,000	0,00000	233	238	233	49
356	244	100,000	0,00000	232	237	232	54
357	243	100,000	0,00000	231	236	231	54
358	242	100,000	0,00000	230	235	230	55
359	241	100,000	0,00000	223	234	223	50
360	240	100,000	0,00000	222	233	222	50
361	239	100,000	0,00000	221	232	221	55
362	238	100,000	0,00000	220	231	220	55
363	237	100,000	0,00000	219	230	219	56
364	236	100,000	0,00000	216	229	216	10
365	235	100,000	0,00000	215	228	215	10
366	234	100,000	0,00000	214	227	214	15
367	233	100,000	0,00000	213	226	213	15
368	232	100,000	0,00000	212	225	212	22
369	231	100,000	0,00000	211	224	211	22
370	230	100,000	0,00000	210	223	210	51
371	229	100,000	0,00000	209	222	209	51
372	228	100,000	0,00000	208	221	208	56
373	227	100,000	0,00000	207	220	207	56
374	226	100,000	0,00000	206	219	206	57
375	225	100,000	0,00000	122	218	122	7
376	224	100,000	0,00000	121	217	121	7
377	223	100,000	0,00000	120	216	120	11

378	222	100,000	0,00000	119	215	119	11
379	221	100,000	0,00000	202	214	202	16
380	220	100,000	0,00000	201	213	201	16
381	219	100,000	0,00000	200	212	200	23
382	218	100,000	0,00000	199	211	199	23
383	217	100,000	0,00000	198	210	198	52
384	216	100,000	0,00000	197	209	197	52
385	215	100,000	0,00000	205	208	205	57
386	214	100,000	0,00000	204	207	204	57
387	213	100,000	0,00000	203	206	203	58
388	212	100,000	0,00000	196	205	196	58
389	211	100,000	0,00000	195	204	195	58
390	210	100,000	0,00000	194	203	194	59
391	209	100,000	0,00000	186	202	186	17
392	208	100,000	0,00000	185	201	185	17
393	207	100,000	0,00000	193	200	193	24
394	206	100,000	0,00000	192	199	192	24
395	205	100,000	0,00000	191	198	191	53
396	204	100,000	0,00000	190	197	190	53
397	203	100,000	0,00000	189	196	189	59
398	202	100,000	0,00000	188	195	188	59
399	201	100,000	0,00000	187	194	187	60
400	200	100,000	0,00000	184	193	184	25
401	199	100,000	0,00000	183	192	183	25
402	198	100,000	0,00000	182	191	182	54
403	197	100,000	0,00000	181	190	181	54
404	196	100,000	0,00000	180	189	180	60
405	195	100,000	0,00000	179	188	179	60
406	194	100,000	0,00000	178	187	178	61
407	193	100,000	0,00000	150	186	150	18
408	192	100,000	0,00000	149	185	149	18
409	191	100,000	0,00000	167	184	167	26
410	190	100,000	0,00000	166	183	166	26
411	189	100,000	0,00000	177	182	177	55
412	188	100,000	0,00000	176	181	176	55
413	187	100,000	0,00000	175	180	175	61
414	186	100,000	0,00000	174	179	174	61
415	185	100,000	0,00000	173	178	173	62
416	184	100,000	0,00000	172	177	172	56
417	183	100,000	0,00000	171	176	171	56
418	182	100,000	0,00000	170	175	170	62
419	181	100,000	0,00000	169	174	169	62
420	180	100,000	0,00000	168	173	168	63
421	179	100,000	0,00000	165	172	165	57
422	178	100,000	0,00000	164	171	164	57
423	177	100,000	0,00000	163	170	163	63
424	176	100,000	0,00000	162	169	162	63
425	175	100,000	0,00000	161	168	161	64
426	174	100,000	0,00000	148	167	148	27
427	173	100,000	0,00000	147	166	147	27
428	172	100,000	0,00000	160	165	160	58
429	171	100,000	0,00000	159	164	159	58
430	170	100,000	0,00000	158	163	158	64
431	169	100,000	0,00000	157	162	157	64
432	168	100,000	0,00000	156	161	156	65
433	167	100,000	0,00000	155	160	155	59
434	166	100,000	0,00000	154	159	154	59
435	165	100,000	0,00000	153	158	153	65
436	164	100,000	0,00000	152	157	152	65
437	163	100,000	0,00000	151	156	151	66
438	162	100,000	0,00000	146	155	146	60
439	161	100,000	0,00000	145	154	145	60
440	160	100,000	0,00000	144	153	144	66
441	159	100,000	0,00000	143	152	143	66
442	158	100,000	0,00000	142	151	142	67

443	157	100,000	0,00000	118	150	118	19
444	156	100,000	0,00000	117	149	117	19
445	155	100,000	0,00000	136	148	136	28
446	154	100,000	0,00000	135	147	135	28
447	153	100,000	0,00000	141	146	141	61
448	152	100,000	0,00000	140	145	140	61
449	151	100,000	0,00000	139	144	139	67
450	150	100,000	0,00000	138	143	138	67
451	149	100,000	0,00000	137	142	137	68
452	148	100,000	0,00000	134	141	134	62
453	147	100,000	0,00000	133	140	133	62
454	146	100,000	0,00000	132	139	132	68
455	145	100,000	0,00000	131	138	131	68
456	144	100,000	0,00000	130	137	130	69
457	143	100,000	0,00000	129	136	129	29
458	142	100,000	0,00000	128	135	128	29
459	141	100,000	0,00000	127	134	127	63
460	140	100,000	0,00000	126	133	126	63
461	139	100,000	0,00000	125	132	125	69
462	138	100,000	0,00000	124	131	124	69
463	137	100,000	0,00000	123	130	123	70
464	136	100,000	0,00000	116	129	116	30
465	135	100,000	0,00000	115	128	115	30
466	134	100,000	0,00000	114	127	114	64
467	133	100,000	0,00000	113	126	113	64
468	132	100,000	0,00000	112	125	112	70
469	131	100,000	0,00000	111	124	111	70
470	130	100,000	0,00000	110	123	110	71
471	129	100,000	0,00000	90	122	90	8
472	128	100,000	0,00000	89	121	89	8
473	127	100,000	0,00000	88	120	88	12
474	126	100,000	0,00000	87	119	87	12
475	125	100,000	0,00000	104	118	104	20
476	124	100,000	0,00000	103	117	103	20
477	123	100,000	0,00000	102	116	102	31
478	122	100,000	0,00000	101	115	101	31
479	121	100,000	0,00000	109	114	109	65
480	120	100,000	0,00000	108	113	108	65
481	119	100,000	0,00000	107	112	107	71
482	118	100,000	0,00000	106	111	106	71
483	117	100,000	0,00000	105	110	105	72
484	116	100,000	0,00000	100	109	100	66
485	115	100,000	0,00000	99	108	99	66
486	114	100,000	0,00000	98	107	98	72
487	113	100,000	0,00000	97	106	97	72
488	112	100,000	0,00000	96	105	96	73
489	111	100,000	0,00000	86	104	86	21
490	110	100,000	0,00000	85	103	85	21
491	109	100,000	0,00000	84	102	84	32
492	108	100,000	0,00000	83	101	83	32
493	107	100,000	0,00000	95	100	95	67
494	106	100,000	0,00000	94	99	94	67
495	105	100,000	0,00000	93	98	93	73
496	104	100,000	0,00000	92	97	92	73
497	103	100,000	0,00000	91	96	91	74
498	102	100,000	0,00000	82	95	82	68
499	101	100,000	0,00000	81	94	81	68
500	100	100,000	0,00000	80	93	80	74
501	99	100,000	0,00000	79	92	79	74
502	98	100,000	0,00000	78	91	78	75
503	97	100,000	0,00000	29	86	29	22
504	96	100,000	0,00000	28	85	28	22
505	95	100,000	0,00000	56	84	56	33
506	94	100,000	0,00000	55	83	55	33
507	93	100,000	0,00000	77	82	77	69

508	92	100,000	0,00000	76	81	76	69
509	91	100,000	0,00000	75	80	75	75
510	90	100,000	0,00000	74	79	74	75
511	89	100,000	0,00000	73	78	73	76
512	88	100,000	0,00000	72	77	72	70
513	87	100,000	0,00000	71	76	71	70
514	86	100,000	0,00000	70	75	70	76
515	85	100,000	0,00000	69	74	69	76
516	84	100,000	0,00000	68	73	68	77
517	83	100,000	0,00000	61	72	61	71
518	82	100,000	0,00000	60	71	60	71
519	81	100,000	0,00000	67	70	67	77
520	80	100,000	0,00000	66	69	66	77
521	79	100,000	0,00000	65	68	65	78
522	78	100,000	0,00000	64	67	64	78
523	77	100,000	0,00000	63	66	63	78
524	76	100,000	0,00000	62	65	62	79
525	75	100,000	0,00000	59	64	59	79
526	74	100,000	0,00000	58	63	58	79
527	73	100,000	0,00000	57	62	57	80
528	72	100,000	0,00000	54	61	54	72
529	71	100,000	0,00000	53	60	53	72
530	70	100,000	0,00000	52	59	52	80
531	69	100,000	0,00000	51	58	51	80
532	68	100,000	0,00000	50	57	50	81
533	67	100,000	0,00000	36	56	36	34
534	66	100,000	0,00000	35	55	35	34
535	65	100,000	0,00000	49	54	49	73
536	64	100,000	0,00000	48	53	48	73
537	63	100,000	0,00000	47	52	47	81
538	62	100,000	0,00000	46	51	46	81
539	61	100,000	0,00000	45	50	45	82
540	60	100,000	0,00000	41	49	41	74
541	59	100,000	0,00000	40	48	40	74
542	58	100,000	0,00000	44	47	44	82
543	57	100,000	0,00000	43	46	43	82
544	56	100,000	0,00000	42	45	42	83
545	55	100,000	0,00000	39	44	39	83
546	54	100,000	0,00000	38	43	38	83
547	53	100,000	0,00000	37	42	37	84
548	52	100,000	0,00000	34	41	34	75
549	51	100,000	0,00000	33	40	33	75
550	50	100,000	0,00000	32	39	32	84
551	49	100,000	0,00000	31	38	31	84
552	48	100,000	0,00000	30	37	30	85
553	47	100,000	0,00000	27	36	27	35
554	46	100,000	0,00000	26	35	26	35
555	45	100,000	0,00000	25	34	25	76
556	44	100,000	0,00000	24	33	24	76
557	43	100,000	0,00000	23	32	23	85
558	42	100,000	0,00000	22	31	22	85
559	41	100,000	0,00000	21	30	21	86
560	40	100,000	0,00000	20	25	20	77
561	39	100,000	0,00000	19	24	19	77
562	38	100,000	0,00000	18	23	18	86
563	37	100,000	0,00000	17	22	17	86
564	36	100,000	0,00000	16	21	16	87
565	35	100,000	0,00000	15	20	15	78
566	34	100,000	0,00000	14	19	14	78
567	33	100,000	0,00000	13	18	13	87
568	32	100,000	0,00000	12	17	12	87
569	31	100,000	0,00000	11	16	11	88
570	30	100,000	0,00000	10	15	10	79
571	29	100,000	0,00000	9	14	9	79
572	28	100,000	0,00000	8	13	8	88

573	27	100,000	0,00000	7	12	7	88
574	26	100,000	0,00000	6	11	6	89
575	25	100,000	0,00000	5	10	5	80
576	24	100,000	0,00000	4	9	4	80
577	23	100,000	0,00000	3	8	3	89
578	22	100,000	0,00000	2	7	2	89
579	21	100,000	0,00000	1	6	1	90
580	20	90,152	0,56361	340	342	340	6
581	19	90,152	0,56361	90	341	90	11
582	18	90,152	0,56361	1	3	1	179
583	17	90,152	0,56361	87	89	87	20
584	16	90,152	0,56361	29	88	29	34
585	15	90,152	0,56361	26	28	26	57
586	14	90,152	0,56361	5	27	5	115
587	13	90,152	0,56361	2	4	2	169
588	12	90,152	0,56361	416	418	416	3
589	11	90,152	0,56361	415	417	415	3
590	10	80,304	1,12722	90	343	90	14
591	9	70,456	1,69083	1	5	1	294
592	8	70,456	1,69083	87	340	87	26
593	7	70,456	1,69083	2	26	2	226
594	6	61,719	2,19089	415	416	415	6
595	5	60,608	2,25444	29	90	29	48
596	4	50,651	2,82428	29	87	29	74
597	3	50,651	2,82428	1	2	1	520
598	2	35,339	3,70059	29	415	29	80
599	1	0,000	5,72310	1	29	1	600

Πίνακας Συγχώνευσης Συστάδων στη συσταδοποίηση των μεταβλητών epis_id_sem_traj_id

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	599	100,000	0,00000	595 600	595	2
2	598	100,000	0,00000	592 599	592	2
3	597	100,000	0,00000	591 598	591	2
4	596	100,000	0,00000	590 597	590	2
5	595	100,000	0,00000	589 596	589	2
6	594	100,000	0,00000	588 595	588	3
7	593	100,000	0,00000	539 594	539	2
8	592	100,000	0,00000	538 593	538	2
9	591	100,000	0,00000	587 592	587	3
10	590	100,000	0,00000	586 591	586	3
11	589	100,000	0,00000	585 590	585	3
12	588	100,000	0,00000	584 589	584	3
13	587	100,000	0,00000	583 588	583	4
14	586	100,000	0,00000	582 587	582	4
15	585	100,000	0,00000	581 586	581	4
16	584	100,000	0,00000	580 585	580	4
17	583	100,000	0,00000	579 584	579	4
18	582	100,000	0,00000	578 583	578	5
19	581	100,000	0,00000	563 582	563	5
20	580	100,000	0,00000	562 581	562	5
21	579	100,000	0,00000	561 580	561	5
22	578	100,000	0,00000	560 579	560	5
23	577	100,000	0,00000	559 578	559	6
24	576	100,000	0,00000	525 577	525	2
25	575	100,000	0,00000	524 576	524	2
26	574	100,000	0,00000	523 575	523	2
27	573	100,000	0,00000	522 574	522	2

28	572	100,000	0,00000	521	573	521	2
29	571	100,000	0,00000	427	572	427	2
30	570	100,000	0,00000	426	571	426	2
31	569	100,000	0,00000	462	570	462	2
32	568	100,000	0,00000	461	569	461	2
33	567	100,000	0,00000	553	568	553	2
34	566	100,000	0,00000	552	567	552	2
35	565	100,000	0,00000	551	566	551	2
36	564	100,000	0,00000	550	565	550	2
37	563	100,000	0,00000	549	564	549	2
38	562	100,000	0,00000	558	563	558	6
39	561	100,000	0,00000	557	562	557	6
40	560	100,000	0,00000	556	561	556	6
41	559	100,000	0,00000	555	560	555	6
42	558	100,000	0,00000	554	559	554	7
43	557	100,000	0,00000	548	558	548	7
44	556	100,000	0,00000	547	557	547	7
45	555	100,000	0,00000	546	556	546	7
46	554	100,000	0,00000	545	555	545	7
47	553	100,000	0,00000	544	554	544	8
48	552	100,000	0,00000	520	553	520	3
49	551	100,000	0,00000	519	552	519	3
50	550	100,000	0,00000	518	551	518	3
51	549	100,000	0,00000	517	550	517	3
52	548	100,000	0,00000	516	549	516	3
53	547	100,000	0,00000	537	548	537	8
54	546	100,000	0,00000	536	547	536	8
55	545	100,000	0,00000	535	546	535	8
56	544	100,000	0,00000	534	545	534	8
57	543	100,000	0,00000	533	544	533	9
58	542	100,000	0,00000	499	543	499	2
59	541	100,000	0,00000	498	542	498	2
60	540	100,000	0,00000	497	541	497	2
61	539	100,000	0,00000	496	540	496	2
62	538	100,000	0,00000	532	539	532	3
63	537	100,000	0,00000	531	538	531	3
64	536	100,000	0,00000	530	537	530	9
65	535	100,000	0,00000	529	536	529	9
66	534	100,000	0,00000	528	535	528	9
67	533	100,000	0,00000	527	534	527	9
68	532	100,000	0,00000	526	533	526	10
69	531	100,000	0,00000	495	532	495	4
70	530	100,000	0,00000	494	531	494	4
71	529	100,000	0,00000	515	530	515	10
72	528	100,000	0,00000	514	529	514	10
73	527	100,000	0,00000	513	528	513	10
74	526	100,000	0,00000	512	527	512	10
75	525	100,000	0,00000	511	526	511	11
76	524	100,000	0,00000	467	525	467	3
77	523	100,000	0,00000	466	524	466	3
78	522	100,000	0,00000	465	523	465	3
79	521	100,000	0,00000	464	522	464	3
80	520	100,000	0,00000	463	521	463	3
81	519	100,000	0,00000	460	520	460	4
82	518	100,000	0,00000	459	519	459	4
83	517	100,000	0,00000	458	518	458	4
84	516	100,000	0,00000	457	517	457	4
85	515	100,000	0,00000	456	516	456	4
86	514	100,000	0,00000	493	515	493	11
87	513	100,000	0,00000	492	514	492	11
88	512	100,000	0,00000	510	513	510	11
89	511	100,000	0,00000	509	512	509	11
90	510	100,000	0,00000	508	511	508	12
91	509	100,000	0,00000	491	510	491	12
92	508	100,000	0,00000	490	509	490	12

93	507	100,000	0,00000	489	508	489	13
94	506	100,000	0,00000	416	507	416	2
95	505	100,000	0,00000	415	506	415	2
96	504	100,000	0,00000	414	505	414	2
97	503	100,000	0,00000	413	504	413	2
98	502	100,000	0,00000	412	503	412	2
99	501	100,000	0,00000	411	502	411	2
100	500	100,000	0,00000	410	501	410	2
101	499	100,000	0,00000	409	500	409	2
102	498	100,000	0,00000	408	499	408	3
103	497	100,000	0,00000	407	498	407	3
104	496	100,000	0,00000	406	497	406	3
105	495	100,000	0,00000	405	496	405	3
106	494	100,000	0,00000	404	495	404	5
107	493	100,000	0,00000	403	494	403	5
108	492	100,000	0,00000	488	493	488	12
109	491	100,000	0,00000	487	492	487	12
110	490	100,000	0,00000	486	491	486	13
111	489	100,000	0,00000	485	490	485	13
112	488	100,000	0,00000	484	489	484	14
113	487	100,000	0,00000	483	488	483	13
114	486	100,000	0,00000	482	487	482	13
115	485	100,000	0,00000	481	486	481	14
116	484	100,000	0,00000	480	485	480	14
117	483	100,000	0,00000	479	484	479	15
118	482	100,000	0,00000	455	483	455	14
119	481	100,000	0,00000	454	482	454	14
120	480	100,000	0,00000	453	481	453	15
121	479	100,000	0,00000	452	480	452	15
122	478	100,000	0,00000	451	479	451	16
123	477	100,000	0,00000	448	469	448	2
124	476	100,000	0,00000	447	468	447	2
125	475	100,000	0,00000	446	467	446	4
126	474	100,000	0,00000	445	466	445	4
127	473	100,000	0,00000	444	465	444	4
128	472	100,000	0,00000	443	464	443	4
129	471	100,000	0,00000	442	463	442	4
130	470	100,000	0,00000	441	462	441	3
131	469	100,000	0,00000	440	461	440	3
132	468	100,000	0,00000	439	460	439	5
133	467	100,000	0,00000	438	459	438	5
134	466	100,000	0,00000	437	458	437	5
135	465	100,000	0,00000	436	457	436	5
136	464	100,000	0,00000	435	456	435	5
137	463	100,000	0,00000	434	455	434	15
138	462	100,000	0,00000	433	454	433	15
139	461	100,000	0,00000	432	453	432	16
140	460	100,000	0,00000	431	452	431	16
141	459	100,000	0,00000	430	451	430	17
142	458	100,000	0,00000	385	446	385	5
143	457	100,000	0,00000	384	445	384	5
144	456	100,000	0,00000	383	444	383	5
145	455	100,000	0,00000	382	443	382	5
146	454	100,000	0,00000	381	442	381	5
147	453	100,000	0,00000	425	441	425	4
148	452	100,000	0,00000	424	440	424	4
149	451	100,000	0,00000	423	439	423	6
150	450	100,000	0,00000	422	438	422	6
151	449	100,000	0,00000	421	437	421	6
152	448	100,000	0,00000	420	436	420	6
153	447	100,000	0,00000	419	435	419	6
154	446	100,000	0,00000	402	434	402	16
155	445	100,000	0,00000	401	433	401	16
156	444	100,000	0,00000	400	432	400	17
157	443	100,000	0,00000	399	431	399	17

158	442	100,000	0,00000	398	430	398	18
159	441	100,000	0,00000	378	429	378	2
160	440	100,000	0,00000	377	428	377	2
161	439	100,000	0,00000	376	427	376	3
162	438	100,000	0,00000	375	426	375	3
163	437	100,000	0,00000	374	425	374	5
164	436	100,000	0,00000	373	424	373	5
165	435	100,000	0,00000	397	423	397	7
166	434	100,000	0,00000	396	422	396	7
167	433	100,000	0,00000	395	421	395	7
168	432	100,000	0,00000	394	420	394	7
169	431	100,000	0,00000	393	419	393	7
170	430	100,000	0,00000	343	414	343	3
171	429	100,000	0,00000	342	413	342	3
172	428	100,000	0,00000	341	412	341	3
173	427	100,000	0,00000	340	411	340	3
174	426	100,000	0,00000	339	410	339	3
175	425	100,000	0,00000	338	409	338	3
176	424	100,000	0,00000	337	408	337	4
177	423	100,000	0,00000	336	407	336	4
178	422	100,000	0,00000	367	406	367	4
179	421	100,000	0,00000	366	405	366	4
180	420	100,000	0,00000	392	404	392	6
181	419	100,000	0,00000	391	403	391	6
182	418	100,000	0,00000	390	402	390	17
183	417	100,000	0,00000	389	401	389	17
184	416	100,000	0,00000	388	400	388	18
185	415	100,000	0,00000	387	399	387	18
186	414	100,000	0,00000	386	398	386	19
187	413	100,000	0,00000	372	397	372	8
188	412	100,000	0,00000	371	396	371	8
189	411	100,000	0,00000	370	395	370	8
190	410	100,000	0,00000	369	394	369	8
191	409	100,000	0,00000	368	393	368	8
192	408	100,000	0,00000	365	392	365	7
193	407	100,000	0,00000	364	391	364	7
194	406	100,000	0,00000	363	390	363	18
195	405	100,000	0,00000	362	389	362	18
196	404	100,000	0,00000	361	388	361	19
197	403	100,000	0,00000	360	387	360	19
198	402	100,000	0,00000	359	386	359	20
199	401	100,000	0,00000	263	385	263	6
200	400	100,000	0,00000	262	384	262	6
201	399	100,000	0,00000	261	383	261	6
202	398	100,000	0,00000	260	382	260	6
203	397	100,000	0,00000	259	381	259	6
204	396	100,000	0,00000	122	380	122	2
205	395	100,000	0,00000	121	379	121	2
206	394	100,000	0,00000	326	378	326	3
207	393	100,000	0,00000	325	377	325	3
208	392	100,000	0,00000	324	376	324	4
209	391	100,000	0,00000	323	375	323	4
210	390	100,000	0,00000	322	374	322	6
211	389	100,000	0,00000	321	373	321	6
212	388	100,000	0,00000	320	372	320	9
213	387	100,000	0,00000	319	371	319	9
214	386	100,000	0,00000	318	370	318	9
215	385	100,000	0,00000	317	369	317	9
216	384	100,000	0,00000	316	368	316	9
217	383	100,000	0,00000	335	367	335	5
218	382	100,000	0,00000	334	366	334	5
219	381	100,000	0,00000	333	365	333	8
220	380	100,000	0,00000	332	364	332	8
221	379	100,000	0,00000	358	363	358	19
222	378	100,000	0,00000	357	362	357	19

223	377	100,000	0,00000	356	361	356	20
224	376	100,000	0,00000	355	360	355	20
225	375	100,000	0,00000	354	359	354	21
226	374	100,000	0,00000	353	358	353	20
227	373	100,000	0,00000	352	357	352	20
228	372	100,000	0,00000	351	356	351	21
229	371	100,000	0,00000	350	355	350	21
230	370	100,000	0,00000	349	354	349	22
231	369	100,000	0,00000	348	353	348	21
232	368	100,000	0,00000	347	352	347	21
233	367	100,000	0,00000	346	351	346	22
234	366	100,000	0,00000	345	350	345	22
235	365	100,000	0,00000	344	349	344	23
236	364	100,000	0,00000	331	348	331	22
237	363	100,000	0,00000	330	347	330	22
238	362	100,000	0,00000	329	346	329	23
239	361	100,000	0,00000	328	345	328	23
240	360	100,000	0,00000	327	344	327	24
241	359	100,000	0,00000	279	339	279	4
242	358	100,000	0,00000	278	338	278	4
243	357	100,000	0,00000	277	337	277	5
244	356	100,000	0,00000	276	336	276	5
245	355	100,000	0,00000	275	335	275	6
246	354	100,000	0,00000	274	334	274	6
247	353	100,000	0,00000	273	333	273	9
248	352	100,000	0,00000	272	332	272	9
249	351	100,000	0,00000	315	331	315	23
250	350	100,000	0,00000	314	330	314	23
251	349	100,000	0,00000	313	329	313	24
252	348	100,000	0,00000	312	328	312	24
253	347	100,000	0,00000	311	327	311	25
254	346	100,000	0,00000	120	326	120	4
255	345	100,000	0,00000	119	325	119	4
256	344	100,000	0,00000	118	324	118	5
257	343	100,000	0,00000	117	323	117	5
258	342	100,000	0,00000	193	322	193	7
259	341	100,000	0,00000	192	321	192	7
260	340	100,000	0,00000	258	320	258	10
261	339	100,000	0,00000	257	319	257	10
262	338	100,000	0,00000	256	318	256	10
263	337	100,000	0,00000	255	317	255	10
264	336	100,000	0,00000	254	316	254	10
265	335	100,000	0,00000	310	315	310	24
266	334	100,000	0,00000	309	314	309	24
267	333	100,000	0,00000	308	313	308	25
268	332	100,000	0,00000	307	312	307	25
269	331	100,000	0,00000	306	311	306	26
270	330	100,000	0,00000	299	310	299	25
271	329	100,000	0,00000	298	309	298	25
272	328	100,000	0,00000	305	308	305	26
273	327	100,000	0,00000	304	307	304	26
274	326	100,000	0,00000	303	306	303	27
275	325	100,000	0,00000	302	305	302	27
276	324	100,000	0,00000	301	304	301	27
277	323	100,000	0,00000	300	303	300	28
278	322	100,000	0,00000	297	302	297	28
279	321	100,000	0,00000	296	301	296	28
280	320	100,000	0,00000	295	300	295	29
281	319	100,000	0,00000	294	299	294	26
282	318	100,000	0,00000	293	298	293	26
283	317	100,000	0,00000	292	297	292	29
284	316	100,000	0,00000	291	296	291	29
285	315	100,000	0,00000	290	295	290	30
286	314	100,000	0,00000	289	294	289	27
287	313	100,000	0,00000	288	293	288	27

288	312	100,000	0,00000	287	292	287	30
289	311	100,000	0,00000	286	291	286	30
290	310	100,000	0,00000	285	290	285	31
291	309	100,000	0,00000	284	289	284	28
292	308	100,000	0,00000	283	288	283	28
293	307	100,000	0,00000	282	287	282	31
294	306	100,000	0,00000	281	286	281	31
295	305	100,000	0,00000	280	285	280	32
296	304	100,000	0,00000	271	284	271	29
297	303	100,000	0,00000	270	283	270	29
298	302	100,000	0,00000	269	282	269	32
299	301	100,000	0,00000	268	281	268	32
300	300	100,000	0,00000	267	280	267	33
301	299	100,000	0,00000	218	279	218	5
302	298	100,000	0,00000	217	278	217	5
303	297	100,000	0,00000	229	277	229	6
304	296	100,000	0,00000	228	276	228	6
305	295	100,000	0,00000	253	275	253	7
306	294	100,000	0,00000	252	274	252	7
307	293	100,000	0,00000	251	273	251	10
308	292	100,000	0,00000	250	272	250	10
309	291	100,000	0,00000	249	271	249	30
310	290	100,000	0,00000	248	270	248	30
311	289	100,000	0,00000	266	269	266	33
312	288	100,000	0,00000	265	268	265	33
313	287	100,000	0,00000	264	267	264	34
314	286	100,000	0,00000	247	266	247	34
315	285	100,000	0,00000	246	265	246	34
316	284	100,000	0,00000	245	264	245	35
317	283	100,000	0,00000	141	263	141	7
318	282	100,000	0,00000	140	262	140	7
319	281	100,000	0,00000	139	261	139	7
320	280	100,000	0,00000	138	260	138	7
321	279	100,000	0,00000	137	259	137	7
322	278	100,000	0,00000	191	258	191	11
323	277	100,000	0,00000	190	257	190	11
324	276	100,000	0,00000	189	256	189	11
325	275	100,000	0,00000	188	255	188	11
326	274	100,000	0,00000	187	254	187	11
327	273	100,000	0,00000	227	253	227	8
328	272	100,000	0,00000	226	252	226	8
329	271	100,000	0,00000	241	251	241	11
330	270	100,000	0,00000	240	250	240	11
331	269	100,000	0,00000	239	249	239	31
332	268	100,000	0,00000	238	248	238	31
333	267	100,000	0,00000	244	247	244	35
334	266	100,000	0,00000	243	246	243	35
335	265	100,000	0,00000	242	245	242	36
336	264	100,000	0,00000	237	244	237	36
337	263	100,000	0,00000	236	243	236	36
338	262	100,000	0,00000	235	242	235	37
339	261	100,000	0,00000	225	241	225	12
340	260	100,000	0,00000	224	240	224	12
341	259	100,000	0,00000	234	239	234	32
342	258	100,000	0,00000	233	238	233	32
343	257	100,000	0,00000	232	237	232	37
344	256	100,000	0,00000	231	236	231	37
345	255	100,000	0,00000	230	235	230	38
346	254	100,000	0,00000	223	234	223	33
347	253	100,000	0,00000	222	233	222	33
348	252	100,000	0,00000	221	232	221	38
349	251	100,000	0,00000	220	231	220	38
350	250	100,000	0,00000	219	230	219	39
351	249	100,000	0,00000	216	229	216	7
352	248	100,000	0,00000	215	228	215	7

353	247	100,000	0,00000	214	227	214	9
354	246	100,000	0,00000	213	226	213	9
355	245	100,000	0,00000	212	225	212	13
356	244	100,000	0,00000	211	224	211	13
357	243	100,000	0,00000	210	223	210	34
358	242	100,000	0,00000	209	222	209	34
359	241	100,000	0,00000	208	221	208	39
360	240	100,000	0,00000	207	220	207	39
361	239	100,000	0,00000	206	219	206	40
362	238	100,000	0,00000	90	218	90	6
363	237	100,000	0,00000	89	217	89	6
364	236	100,000	0,00000	88	216	88	8
365	235	100,000	0,00000	87	215	87	8
366	234	100,000	0,00000	202	214	202	10
367	233	100,000	0,00000	201	213	201	10
368	232	100,000	0,00000	200	212	200	14
369	231	100,000	0,00000	199	211	199	14
370	230	100,000	0,00000	198	210	198	35
371	229	100,000	0,00000	197	209	197	35
372	228	100,000	0,00000	205	208	205	40
373	227	100,000	0,00000	204	207	204	40
374	226	100,000	0,00000	203	206	203	41
375	225	100,000	0,00000	196	205	196	41
376	224	100,000	0,00000	195	204	195	41
377	223	100,000	0,00000	194	203	194	42
378	222	100,000	0,00000	186	202	186	11
379	221	100,000	0,00000	185	201	185	11
380	220	100,000	0,00000	184	200	184	15
381	219	100,000	0,00000	183	199	183	15
382	218	100,000	0,00000	182	198	182	36
383	217	100,000	0,00000	181	197	181	36
384	216	100,000	0,00000	180	196	180	42
385	215	100,000	0,00000	179	195	179	42
386	214	100,000	0,00000	178	194	178	43
387	213	100,000	0,00000	136	193	136	8
388	212	100,000	0,00000	135	192	135	8
389	211	100,000	0,00000	177	191	177	12
390	210	100,000	0,00000	176	190	176	12
391	209	100,000	0,00000	175	189	175	12
392	208	100,000	0,00000	174	188	174	12
393	207	100,000	0,00000	173	187	173	12
394	206	100,000	0,00000	150	186	150	12
395	205	100,000	0,00000	149	185	149	12
396	204	100,000	0,00000	167	184	167	16
397	203	100,000	0,00000	166	183	166	16
398	202	100,000	0,00000	172	182	172	37
399	201	100,000	0,00000	171	181	171	37
400	200	100,000	0,00000	170	180	170	43
401	199	100,000	0,00000	169	179	169	43
402	198	100,000	0,00000	168	178	168	44
403	197	100,000	0,00000	155	177	155	13
404	196	100,000	0,00000	154	176	154	13
405	195	100,000	0,00000	153	175	153	13
406	194	100,000	0,00000	152	174	152	13
407	193	100,000	0,00000	151	173	151	13
408	192	100,000	0,00000	165	172	165	38
409	191	100,000	0,00000	164	171	164	38
410	190	100,000	0,00000	163	170	163	44
411	189	100,000	0,00000	162	169	162	44
412	188	100,000	0,00000	161	168	161	45
413	187	100,000	0,00000	148	167	148	17
414	186	100,000	0,00000	147	166	147	17
415	185	100,000	0,00000	160	165	160	39
416	184	100,000	0,00000	159	164	159	39
417	183	100,000	0,00000	158	163	158	45

418	182	100,000	0,00000	157	162	157	45
419	181	100,000	0,00000	156	161	156	46
420	180	100,000	0,00000	146	160	146	40
421	179	100,000	0,00000	145	159	145	40
422	178	100,000	0,00000	144	158	144	46
423	177	100,000	0,00000	143	157	143	46
424	176	100,000	0,00000	142	156	142	47
425	175	100,000	0,00000	134	155	134	14
426	174	100,000	0,00000	133	154	133	14
427	173	100,000	0,00000	132	153	132	14
428	172	100,000	0,00000	131	152	131	14
429	171	100,000	0,00000	130	151	130	14
430	170	100,000	0,00000	104	150	104	13
431	169	100,000	0,00000	103	149	103	13
432	168	100,000	0,00000	129	148	129	18
433	167	100,000	0,00000	128	147	128	18
434	166	100,000	0,00000	127	146	127	41
435	165	100,000	0,00000	126	145	126	41
436	164	100,000	0,00000	125	144	125	47
437	163	100,000	0,00000	124	143	124	47
438	162	100,000	0,00000	123	142	123	48
439	161	100,000	0,00000	116	136	116	9
440	160	100,000	0,00000	115	135	115	9
441	159	100,000	0,00000	114	134	114	15
442	158	100,000	0,00000	113	133	113	15
443	157	100,000	0,00000	112	132	112	15
444	156	100,000	0,00000	111	131	111	15
445	155	100,000	0,00000	110	130	110	15
446	154	100,000	0,00000	102	129	102	19
447	153	100,000	0,00000	101	128	101	19
448	152	100,000	0,00000	109	127	109	42
449	151	100,000	0,00000	108	126	108	42
450	150	100,000	0,00000	107	125	107	48
451	149	100,000	0,00000	106	124	106	48
452	148	100,000	0,00000	105	123	105	49
453	147	100,000	0,00000	95	114	95	16
454	146	100,000	0,00000	94	113	94	16
455	145	100,000	0,00000	93	112	93	16
456	144	100,000	0,00000	92	111	92	16
457	143	100,000	0,00000	91	110	91	16
458	142	100,000	0,00000	100	109	100	43
459	141	100,000	0,00000	99	108	99	43
460	140	100,000	0,00000	98	107	98	49
461	139	100,000	0,00000	97	106	97	49
462	138	100,000	0,00000	96	105	96	50
463	137	100,000	0,00000	86	104	86	14
464	136	100,000	0,00000	85	103	85	14
465	135	100,000	0,00000	84	102	84	20
466	134	100,000	0,00000	83	101	83	20
467	133	100,000	0,00000	82	100	82	44
468	132	100,000	0,00000	81	99	81	44
469	131	100,000	0,00000	80	98	80	50
470	130	100,000	0,00000	79	97	79	50
471	129	100,000	0,00000	78	96	78	51
472	128	100,000	0,00000	61	95	61	17
473	127	100,000	0,00000	60	94	60	17
474	126	100,000	0,00000	59	93	59	17
475	125	100,000	0,00000	58	92	58	17
476	124	100,000	0,00000	57	91	57	17
477	123	100,000	0,00000	29	86	29	15
478	122	100,000	0,00000	28	85	28	15
479	121	100,000	0,00000	56	84	56	21
480	120	100,000	0,00000	55	83	55	21
481	119	100,000	0,00000	77	82	77	45
482	118	100,000	0,00000	76	81	76	45

483	117	100,000	0,00000	75	80	75	51
484	116	100,000	0,00000	74	79	74	51
485	115	100,000	0,00000	73	78	73	52
486	114	100,000	0,00000	72	77	72	46
487	113	100,000	0,00000	71	76	71	46
488	112	100,000	0,00000	70	75	70	52
489	111	100,000	0,00000	69	74	69	52
490	110	100,000	0,00000	68	73	68	53
491	109	100,000	0,00000	54	72	54	47
492	108	100,000	0,00000	53	71	53	47
493	107	100,000	0,00000	67	70	67	53
494	106	100,000	0,00000	66	69	66	53
495	105	100,000	0,00000	65	68	65	54
496	104	100,000	0,00000	64	67	64	54
497	103	100,000	0,00000	63	66	63	54
498	102	100,000	0,00000	62	65	62	55
499	101	100,000	0,00000	52	64	52	55
500	100	100,000	0,00000	51	63	51	55
501	99	100,000	0,00000	50	62	50	56
502	98	100,000	0,00000	41	61	41	18
503	97	100,000	0,00000	40	60	40	18
504	96	100,000	0,00000	39	59	39	18
505	95	100,000	0,00000	38	58	38	18
506	94	100,000	0,00000	37	57	37	18
507	93	100,000	0,00000	36	56	36	22
508	92	100,000	0,00000	35	55	35	22
509	91	100,000	0,00000	49	54	49	48
510	90	100,000	0,00000	48	53	48	48
511	89	100,000	0,00000	47	52	47	56
512	88	100,000	0,00000	46	51	46	56
513	87	100,000	0,00000	45	50	45	57
514	86	100,000	0,00000	34	49	34	49
515	85	100,000	0,00000	33	48	33	49
516	84	100,000	0,00000	44	47	44	57
517	83	100,000	0,00000	43	46	43	57
518	82	100,000	0,00000	42	45	42	58
519	81	100,000	0,00000	32	44	32	58
520	80	100,000	0,00000	31	43	31	58
521	79	100,000	0,00000	30	42	30	59
522	78	100,000	0,00000	20	41	20	19
523	77	100,000	0,00000	19	40	19	19
524	76	100,000	0,00000	18	39	18	19
525	75	100,000	0,00000	17	38	17	19
526	74	100,000	0,00000	16	37	16	19
527	73	100,000	0,00000	27	36	27	23
528	72	100,000	0,00000	26	35	26	23
529	71	100,000	0,00000	25	34	25	50
530	70	100,000	0,00000	24	33	24	50
531	69	100,000	0,00000	23	32	23	59
532	68	100,000	0,00000	22	31	22	59
533	67	100,000	0,00000	21	30	21	60
534	66	100,000	0,00000	15	25	15	51
535	65	100,000	0,00000	14	24	14	51
536	64	100,000	0,00000	13	23	13	60
537	63	100,000	0,00000	12	22	12	60
538	62	100,000	0,00000	11	21	11	61
539	61	100,000	0,00000	10	20	10	20
540	60	100,000	0,00000	9	19	9	20
541	59	100,000	0,00000	8	18	8	20
542	58	100,000	0,00000	7	17	7	20
543	57	100,000	0,00000	6	16	6	20
544	56	100,000	0,00000	5	15	5	52
545	55	100,000	0,00000	4	14	4	52
546	54	100,000	0,00000	3	13	3	61
547	53	100,000	0,00000	2	12	2	61

548	52	100,000	0,00000	1	11	1	62
549	51	96,050	0,28181	340	341	340	6
550	50	96,050	0,28181	470	471	470	2
551	49	96,050	0,28181	137	138	137	14
552	48	96,050	0,28181	120	121	120	6
553	47	96,050	0,28181	118	119	118	9
554	46	96,050	0,28181	88	89	88	14
555	45	96,050	0,28181	29	87	29	23
556	44	96,050	0,28181	6	7	6	40
557	43	96,050	0,28181	1	2	1	123
558	42	96,050	0,28181	476	477	476	2
559	41	96,050	0,28181	474	475	474	2
560	40	96,050	0,28181	472	473	472	2
561	39	96,050	0,28181	448	449	448	3
562	38	96,050	0,28181	141	447	141	9
563	37	96,050	0,28181	139	140	139	14
564	36	96,050	0,28181	116	117	116	14
565	35	96,050	0,28181	10	115	10	29
566	34	96,050	0,28181	27	28	27	38
567	33	96,050	0,28181	5	26	5	75
568	32	96,050	0,28181	8	9	8	40
569	31	96,050	0,28181	3	4	3	113
570	30	96,050	0,28181	417	418	417	2
571	29	96,050	0,28181	415	416	415	4
572	28	96,050	0,28181	342	343	342	6
573	27	92,101	0,56361	90	340	90	12
574	26	92,101	0,56361	120	122	120	8
575	25	92,101	0,56361	476	478	476	3
576	24	92,101	0,56361	448	450	448	4
577	23	88,151	0,84542	470	472	470	4
578	22	88,151	0,84542	137	139	137	28
579	21	88,151	0,84542	29	88	29	37
580	20	88,151	0,84542	6	8	6	80
581	19	88,151	0,84542	1	3	1	236
582	18	88,151	0,84542	116	118	116	23
583	17	88,151	0,84542	5	27	5	113
584	16	88,151	0,84542	415	417	415	6
585	15	84,201	1,12722	474	476	474	5
586	14	84,201	1,12722	141	448	141	13
587	13	84,201	1,12722	90	342	90	18
588	12	80,251	1,40903	10	116	10	52
589	11	76,374	1,68564	137	470	137	32
590	10	76,374	1,68564	1	6	1	316
591	9	74,167	1,84317	141	474	141	18
592	8	74,167	1,84317	29	120	29	45
593	7	71,579	2,02781	5	10	5	165
594	6	68,402	2,25444	29	90	29	63
595	5	62,368	2,68499	137	141	137	50
596	4	58,995	2,92561	1	5	1	481
597	3	52,603	3,38167	29	415	29	69
598	2	31,020	4,92161	1	137	1	531
599	1	0,000	7,13481	1	29	1	600

Πίνακας συγχώνευσης συστάδων στη συσταδοποίηση των μεταβλητών sem_traj_id-avg_speed

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	599	100,000	0,0000	160 594	160	2

2	598	100,000	0,0000	430	434	430	2
3	597	100,000	0,0000	533	543	533	2
4	596	100,000	0,0000	258	572	258	2
5	595	100,000	0,0000	583	587	583	2
6	594	100,000	0,0000	78	90	78	2
7	593	100,000	0,0000	44	202	44	2
8	592	100,000	0,0000	29	72	29	2
9	591	100,000	0,0000	150	234	150	2
10	590	100,000	0,0000	96	104	96	2
11	589	100,000	0,0000	559	563	559	2
12	588	100,000	0,0000	178	186	178	2
13	587	100,000	0,0000	105	479	105	2
14	586	100,000	0,0000	56	430	56	3
15	585	100,000	0,0000	349	353	349	2
16	584	100,000	0,0000	11	284	11	2
17	583	100,000	0,0000	150	558	150	3
18	582	100,000	0,0000	95	553	95	2
19	581	100,000	0,0000	78	548	78	3
20	580	100,000	0,0000	161	167	161	2
21	579	100,000	0,0000	49	50	49	2
22	578	100,000	0,0000	61	441	61	2
23	577	100,000	0,0000	150	200	150	4
24	576	100,000	0,0000	203	533	203	3
25	575	100,000	0,0000	129	600	129	2
26	574	100,000	0,0000	205	554	205	2
27	573	100,000	0,0000	5	218	5	2
28	572	100,000	0,0000	62	310	62	2
29	571	100,000	0,0000	64	65	64	2
30	570	100,000	0,0000	78	544	78	4
31	569	100,000	0,0000	526	532	526	2
32	568	100,000	0,0000	263	577	263	2
33	567	100,000	0,0000	327	343	327	2
34	566	100,000	0,0000	253	559	253	3
35	565	100,000	0,0000	77	294	77	2
36	564	100,000	0,0000	29	172	29	3
37	563	100,000	0,0000	266	583	266	3
38	562	100,000	0,0000	462	520	462	2
39	561	100,000	0,0000	96	315	96	3
40	560	100,000	0,0000	455	515	455	2
41	559	100,000	0,0000	469	525	469	2
42	558	100,000	0,0000	105	109	105	3
43	557	100,000	0,0000	15	45	15	2
44	556	100,000	0,0000	78	489	78	5
45	555	100,000	0,0000	29	358	29	4
46	554	100,000	0,0000	205	582	205	3
47	553	100,000	0,0000	289	290	289	2
48	552	100,000	0,0000	392	508	392	2
49	551	100,000	0,0000	5	526	5	4
50	550	100,000	0,0000	36	488	36	2
51	549	100,000	0,0000	229	599	229	2
52	548	100,000	0,0000	178	241	178	3
53	547	100,000	0,0000	49	295	49	3
54	546	100,000	0,0000	397	429	397	2
55	545	100,000	0,0000	10	20	10	2
56	544	100,000	0,0000	64	67	64	3
57	543	100,000	0,0000	150	483	150	5
58	542	100,000	0,0000	177	258	177	3
59	541	100,000	0,0000	78	507	78	6
60	540	100,000	0,0000	161	367	161	3
61	539	100,000	0,0000	56	105	56	6
62	538	100,000	0,0000	11	15	11	4
63	537	100,000	0,0000	61	122	61	3
64	536	100,000	0,0000	203	305	203	4
65	535	100,000	0,0000	266	418	266	4
66	534	100,000	0,0000	77	96	77	5

67	533	100,000	0,0000	62	129	62	4
68	532	100,000	0,0000	392	510	392	3
69	531	100,000	0,0000	29	253	29	7
70	530	100,000	0,0000	178	299	178	4
71	529	100,000	0,0000	160	289	160	4
72	528	100,000	0,0000	44	302	44	3
73	527	100,000	0,0000	327	455	327	4
74	526	100,000	0,0000	5	11	5	8
75	525	100,000	0,0000	49	56	49	9
76	524	100,000	0,0000	36	229	36	4
77	523	100,000	0,0000	95	136	95	3
78	522	100,000	0,0000	263	385	263	3
79	521	100,000	0,0000	64	78	64	9
80	520	100,000	0,0000	61	155	61	4
81	519	100,000	0,0000	177	380	177	4
82	518	100,000	0,0000	205	244	205	4
83	517	100,000	0,0000	77	392	77	8
84	516	100,000	0,0000	266	349	266	6
85	515	100,000	0,0000	326	397	326	3
86	514	100,000	0,0000	49	150	49	14
87	513	100,000	0,0000	29	161	29	10
88	512	100,000	0,0000	5	203	5	12
89	511	100,000	0,0000	62	64	62	13
90	510	100,000	0,0000	36	160	36	8
91	509	100,000	0,0000	141	263	141	4
92	508	100,000	0,0000	61	193	61	5
93	507	100,000	0,0000	318	518	318	2
94	506	100,000	0,0000	77	348	77	9
95	505	100,000	0,0000	279	327	279	5
96	504	100,000	0,0000	49	178	49	18
97	503	100,000	0,0000	44	205	44	7
98	502	100,000	0,0000	95	177	95	7
99	501	100,000	0,0000	10	61	10	7
100	500	100,000	0,0000	29	62	29	23
101	499	100,000	0,0000	402	505	402	2
102	498	100,000	0,0000	36	77	36	17
103	497	100,000	0,0000	5	49	5	30
104	496	100,000	0,0000	95	462	95	9
105	495	100,000	0,0000	141	469	141	6
106	494	100,000	0,0000	197	453	197	2
107	493	100,000	0,0000	10	41	10	8
108	492	100,000	0,0000	29	279	29	28
109	491	100,000	0,0000	36	266	36	23
110	490	100,000	0,0000	361	499	361	2
111	489	100,000	0,0000	308	339	308	2
112	488	100,000	0,0000	141	450	141	7
113	487	100,000	0,0000	5	44	5	37
114	486	100,000	0,0000	10	326	10	11
115	485	100,000	0,0000	3	541	3	2
116	484	100,000	0,0000	5	29	5	65
117	483	100,000	0,0000	267	331	267	2
118	482	100,000	0,0000	10	95	10	20
119	481	100,000	0,0000	75	223	75	2
120	480	100,000	0,0000	356	493	356	2
121	479	100,000	0,0000	5	36	5	88
122	478	100,000	0,0000	23	292	23	2
123	477	100,000	0,0000	88	351	88	2
124	476	100,000	0,0000	123	354	123	2
125	475	100,000	0,0000	42	511	42	2
126	474	100,000	0,0000	27	287	27	2
127	473	100,000	0,0000	285	592	285	2
128	472	100,000	0,0000	100	197	100	3
129	471	100,000	0,0000	130	516	130	2
130	470	100,000	0,0000	416	590	416	2
131	469	100,000	0,0000	146	398	146	2

132	468	100,000	0,0000	22	231	22	2
133	467	100,000	0,0000	82	530	82	2
134	466	100,000	0,0000	421	423	421	2
135	465	100,000	0,0000	21	86	21	2
136	464	100,000	0,0000	120	460	120	2
137	463	100,000	0,0000	23	578	23	3
138	462	100,000	0,0000	32	196	32	2
139	461	100,000	0,0000	30	100	30	4
140	460	100,000	0,0000	1	365	1	2
141	459	100,000	0,0000	214	484	214	2
142	458	100,000	0,0000	3	168	3	3
143	457	100,000	0,0000	144	206	144	2
144	456	100,000	0,0000	308	341	308	3
145	455	100,000	0,0000	364	579	364	2
146	454	100,000	0,0000	187	456	187	2
147	453	100,000	0,0000	88	125	88	3
148	452	100,000	0,0001	93	551	93	2
149	451	100,000	0,0001	70	242	70	2
150	450	100,000	0,0001	156	346	156	2
151	449	100,000	0,0001	359	546	359	2
152	448	100,000	0,0001	1	537	1	3
153	447	100,000	0,0001	98	313	98	2
154	446	100,000	0,0001	47	501	47	2
155	445	99,999	0,0001	8	132	8	2
156	444	99,999	0,0001	84	588	84	2
157	443	99,999	0,0001	393	427	393	2
158	442	99,999	0,0001	75	597	75	3
159	441	99,999	0,0001	141	463	141	8
160	440	99,999	0,0001	219	344	219	2
161	439	99,999	0,0001	261	383	261	2
162	438	99,999	0,0001	182	406	182	2
163	437	99,999	0,0001	80	361	80	3
164	436	99,999	0,0001	89	128	89	2
165	435	99,999	0,0001	23	311	23	4
166	434	99,999	0,0001	30	146	30	6
167	433	99,999	0,0001	91	153	91	2
168	432	99,999	0,0001	88	585	88	4
169	431	99,999	0,0001	42	280	42	3
170	430	99,999	0,0001	251	356	251	3
171	429	99,999	0,0001	130	187	130	4
172	428	99,999	0,0001	34	595	34	2
173	427	99,999	0,0001	232	275	232	2
174	426	99,999	0,0001	144	285	144	4
175	425	99,999	0,0001	303	359	303	3
176	424	99,999	0,0001	93	316	93	3
177	423	99,999	0,0001	416	495	416	3
178	422	99,999	0,0001	245	335	245	2
179	421	99,999	0,0001	3	170	3	4
180	420	99,999	0,0002	39	393	39	3
181	419	99,999	0,0002	378	566	378	2
182	418	99,999	0,0002	219	535	219	3
183	417	99,999	0,0002	386	388	386	2
184	416	99,999	0,0002	115	569	115	2
185	415	99,999	0,0002	370	372	370	2
186	414	99,999	0,0002	189	254	189	2
187	413	99,999	0,0002	10	368	10	21
188	412	99,999	0,0002	118	437	118	2
189	411	99,999	0,0002	141	442	141	9
190	410	99,999	0,0002	5	123	5	90
191	409	99,999	0,0002	337	513	337	2
192	408	99,999	0,0002	142	156	142	3
193	407	99,998	0,0002	27	273	27	3
194	406	99,998	0,0002	264	390	264	2
195	405	99,998	0,0002	1	88	1	7
196	404	99,998	0,0002	21	251	21	5

197	403	99,998	0,0002	6	374	6	2
198	402	99,998	0,0002	23	75	23	7
199	401	99,998	0,0002	82	503	82	3
200	400	99,998	0,0002	37	175	37	2
201	399	99,998	0,0002	80	195	80	4
202	398	99,998	0,0002	34	107	34	3
203	397	99,998	0,0002	235	528	235	2
204	396	99,998	0,0002	57	93	57	4
205	395	99,998	0,0003	16	421	16	3
206	394	99,998	0,0003	3	84	3	6
207	393	99,998	0,0003	91	130	91	6
208	392	99,998	0,0003	165	414	165	2
209	391	99,998	0,0003	247	416	247	4
210	390	99,998	0,0003	102	308	102	4
211	389	99,998	0,0003	68	230	68	2
212	388	99,998	0,0003	137	139	137	2
213	387	99,998	0,0003	25	451	25	2
214	386	99,998	0,0003	225	277	225	2
215	385	99,998	0,0003	444	467	444	2
216	384	99,998	0,0003	30	144	30	10
217	383	99,998	0,0003	116	151	116	2
218	382	99,997	0,0003	214	267	214	4
219	381	99,997	0,0003	491	539	491	2
220	380	99,997	0,0004	8	39	8	5
221	379	99,997	0,0004	173	435	173	2
222	378	99,997	0,0004	402	556	402	3
223	377	99,997	0,0004	248	352	248	2
224	376	99,997	0,0004	23	481	23	8
225	375	99,997	0,0004	134	458	134	2
226	374	99,997	0,0004	142	232	142	5
227	373	99,997	0,0005	57	564	57	5
228	372	99,997	0,0005	47	363	47	3
229	371	99,996	0,0005	3	32	3	8
230	370	99,996	0,0005	27	306	27	4
231	369	99,996	0,0005	376	395	376	2
232	368	99,996	0,0005	21	297	21	6
233	367	99,996	0,0005	18	110	18	2
234	366	99,996	0,0005	521	523	521	2
235	365	99,996	0,0005	342	364	342	3
236	364	99,996	0,0005	52	163	52	2
237	363	99,996	0,0005	219	303	219	6
238	362	99,996	0,0005	370	419	370	3
239	361	99,996	0,0006	98	198	98	3
240	360	99,996	0,0006	10	91	10	27
241	359	99,996	0,0006	249	282	249	2
242	358	99,996	0,0006	1	30	1	17
243	357	99,996	0,0006	70	237	70	3
244	356	99,996	0,0006	470	478	470	2
245	355	99,996	0,0006	158	216	158	2
246	354	99,995	0,0006	381	446	381	2
247	353	99,995	0,0006	425	568	425	2
248	352	99,995	0,0007	182	300	182	3
249	351	99,995	0,0007	296	330	296	2
250	350	99,994	0,0007	34	235	34	5
251	349	99,994	0,0008	102	245	102	6
252	348	99,994	0,0008	214	264	214	6
253	347	99,994	0,0009	145	185	145	2
254	346	99,994	0,0009	47	180	47	4
255	345	99,993	0,0009	301	494	301	2
256	344	99,993	0,0009	23	337	23	10
257	343	99,993	0,0009	259	444	259	3
258	342	99,993	0,0009	71	304	71	2
259	341	99,993	0,0009	173	378	173	4
260	340	99,993	0,0009	80	491	80	6
261	339	99,993	0,0009	6	57	6	7

262	338	99,993	0,0009	333	386	333	3
263	337	99,993	0,0010	142	247	142	9
264	336	99,993	0,0010	127	208	127	2
265	335	99,993	0,0010	120	370	120	5
266	334	99,993	0,0010	1	5	1	107
267	333	99,992	0,0010	16	376	16	5
268	332	99,992	0,0011	192	321	192	2
269	331	99,992	0,0011	63	181	63	2
270	330	99,992	0,0011	18	318	18	4
271	329	99,992	0,0011	3	219	3	14
272	328	99,992	0,0011	108	291	108	2
273	327	99,991	0,0011	269	486	269	2
274	326	99,991	0,0011	8	10	8	32
275	325	99,991	0,0012	521	573	521	3
276	324	99,991	0,0012	13	580	13	2
277	323	99,991	0,0012	184	227	184	2
278	322	99,991	0,0012	272	527	272	2
279	321	99,991	0,0012	27	165	27	6
280	320	99,991	0,0012	338	560	338	2
281	319	99,991	0,0013	21	42	21	9
282	318	99,990	0,0014	164	405	164	2
283	317	99,989	0,0014	415	589	415	2
284	316	99,989	0,0014	116	134	116	4
285	315	99,989	0,0015	360	413	360	2
286	314	99,989	0,0015	73	225	73	3
287	313	99,988	0,0016	329	561	329	2
288	312	99,988	0,0016	70	201	70	4
289	311	99,988	0,0017	394	567	394	2
290	310	99,987	0,0017	138	574	138	2
291	309	99,987	0,0017	276	542	276	2
292	308	99,987	0,0018	137	141	137	11
293	307	99,987	0,0018	26	487	26	2
294	306	99,986	0,0018	80	82	80	9
295	305	99,986	0,0018	158	249	158	4
296	304	99,986	0,0019	16	189	16	7
297	303	99,986	0,0019	98	432	98	4
298	302	99,986	0,0019	71	209	71	3
299	301	99,986	0,0019	47	182	47	7
300	300	99,985	0,0020	37	439	37	3
301	299	99,985	0,0020	113	115	113	3
302	298	99,985	0,0021	470	474	470	3
303	297	99,984	0,0021	23	142	23	19
304	296	99,984	0,0022	52	68	52	4
305	295	99,984	0,0022	217	366	217	2
306	294	99,983	0,0023	119	192	119	3
307	293	99,983	0,0023	35	89	35	3
308	292	99,983	0,0023	27	102	27	12
309	291	99,983	0,0023	34	214	34	11
310	290	99,982	0,0025	6	8	6	39
311	289	99,981	0,0025	1	3	1	121
312	288	99,981	0,0025	457	517	457	2
313	287	99,981	0,0025	14	506	14	2
314	286	99,980	0,0027	259	465	259	4
315	285	99,980	0,0027	162	238	162	2
316	284	99,980	0,0027	381	521	381	5
317	283	99,979	0,0028	73	410	73	4
318	282	99,979	0,0028	25	80	25	11
319	281	99,979	0,0028	126	179	126	2
320	280	99,978	0,0029	59	256	59	2
321	279	99,978	0,0030	166	415	166	3
322	278	99,978	0,0030	213	301	213	3
323	277	99,977	0,0030	396	428	396	2
324	276	99,977	0,0031	350	362	350	2
325	275	99,975	0,0033	47	402	47	10
326	274	99,975	0,0033	108	221	108	3

327	273	99,975	0,0033	71	250	71	4
328	272	99,975	0,0034	127	184	127	4
329	271	99,975	0,0034	106	164	106	3
330	270	99,975	0,0034	116	173	116	8
331	269	99,974	0,0035	270	417	270	2
332	268	99,974	0,0035	281	538	281	2
333	267	99,974	0,0035	169	454	169	2
334	266	99,974	0,0035	19	519	19	2
335	265	99,973	0,0036	16	120	16	12
336	264	99,973	0,0036	438	440	438	2
337	263	99,971	0,0039	21	23	21	28
338	262	99,971	0,0040	269	329	269	4
339	261	99,970	0,0041	498	500	498	2
340	260	99,969	0,0041	492	509	492	2
341	259	99,969	0,0042	228	557	228	2
342	258	99,969	0,0042	158	333	158	7
343	257	99,969	0,0042	211	360	211	3
344	256	99,968	0,0042	98	497	98	5
345	255	99,968	0,0044	2	591	2	2
346	254	99,967	0,0044	18	37	18	7
347	253	99,966	0,0045	320	570	320	2
348	252	99,966	0,0046	66	207	66	2
349	251	99,965	0,0047	22	496	22	3
350	250	99,965	0,0047	293	336	293	2
351	249	99,965	0,0047	28	485	28	2
352	248	99,964	0,0048	212	404	212	2
353	247	99,963	0,0049	27	34	27	23
354	246	99,963	0,0050	188	565	188	2
355	245	99,962	0,0051	43	540	43	2
356	244	99,962	0,0051	113	571	113	4
357	243	99,962	0,0051	259	448	259	5
358	242	99,961	0,0053	79	217	79	3
359	241	99,960	0,0053	52	70	52	8
360	240	99,960	0,0053	400	408	400	2
361	239	99,960	0,0053	14	502	14	3
362	238	99,959	0,0054	137	381	137	16
363	237	99,958	0,0056	213	240	213	4
364	236	99,957	0,0057	342	480	342	4
365	235	99,956	0,0059	94	457	94	3
366	234	99,955	0,0060	9	379	9	2
367	233	99,955	0,0061	124	433	124	2
368	232	99,955	0,0061	111	550	111	2
369	231	99,955	0,0061	319	459	319	2
370	230	99,954	0,0061	183	452	183	2
371	229	99,954	0,0061	24	504	24	2
372	228	99,953	0,0063	307	598	307	2
373	227	99,953	0,0064	6	16	6	51
374	226	99,951	0,0066	470	472	470	4
375	225	99,950	0,0067	236	340	236	2
376	224	99,949	0,0069	482	562	482	2
377	223	99,949	0,0069	118	549	118	3
378	222	99,949	0,0069	312	401	312	2
379	221	99,948	0,0069	152	325	152	2
380	220	99,948	0,0069	1	21	1	149
381	219	99,948	0,0070	324	425	324	3
382	218	99,948	0,0070	257	424	257	2
383	217	99,947	0,0070	26	281	26	4
384	216	99,945	0,0074	25	47	25	21
385	215	99,945	0,0074	411	555	411	2
386	214	99,944	0,0074	220	334	220	2
387	213	99,944	0,0075	138	449	138	3
388	212	99,944	0,0075	211	389	211	4
389	211	99,943	0,0076	147	596	147	2
390	210	99,943	0,0076	314	431	314	2
391	209	99,942	0,0078	157	347	157	2

392	208	99,937	0,0084	14	226	14	4
393	207	99,937	0,0084	248	387	248	3
394	206	99,937	0,0084	345	593	345	2
395	205	99,933	0,0089	127	158	127	11
396	204	99,932	0,0091	162	529	162	3
397	203	99,931	0,0092	119	323	119	4
398	202	99,931	0,0093	498	531	498	3
399	201	99,930	0,0093	18	116	18	15
400	200	99,930	0,0093	350	403	350	3
401	199	99,930	0,0094	13	52	13	10
402	198	99,929	0,0095	512	514	512	2
403	197	99,928	0,0097	159	283	159	2
404	196	99,927	0,0097	87	355	87	2
405	195	99,926	0,0099	55	204	55	2
406	194	99,921	0,0106	79	309	79	4
407	193	99,919	0,0108	97	265	97	2
408	192	99,918	0,0110	228	298	228	3
409	191	99,916	0,0112	92	371	92	2
410	190	99,916	0,0113	2	338	2	4
411	189	99,915	0,0114	60	377	60	2
412	188	99,910	0,0121	126	270	126	4
413	187	99,910	0,0121	143	224	143	2
414	186	99,908	0,0123	121	317	121	2
415	185	99,908	0,0124	22	63	22	5
416	184	99,907	0,0124	1	27	1	172
417	183	99,907	0,0124	243	342	243	5
418	182	99,906	0,0125	357	536	357	2
419	181	99,905	0,0127	4	145	4	3
420	180	99,905	0,0127	443	524	443	2
421	179	99,901	0,0132	137	259	137	21
422	178	99,900	0,0134	239	271	239	2
423	177	99,900	0,0134	48	307	48	3
424	176	99,897	0,0139	73	98	73	9
425	175	99,896	0,0139	40	176	40	2
426	174	99,895	0,0141	215	328	215	2
427	173	99,894	0,0142	312	407	312	3
428	172	99,893	0,0143	261	575	261	3
429	171	99,893	0,0143	13	269	13	14
430	170	99,892	0,0145	166	233	166	4
431	169	99,892	0,0145	9	133	9	3
432	168	99,892	0,0145	191	320	191	3
433	167	99,888	0,0150	183	391	183	3
434	166	99,887	0,0151	59	118	59	5
435	165	99,887	0,0151	199	412	199	2
436	164	99,886	0,0153	470	476	470	5
437	163	99,884	0,0155	33	293	33	3
438	162	99,884	0,0155	54	210	54	2
439	161	99,884	0,0156	255	369	255	2
440	160	99,883	0,0157	26	28	26	6
441	159	99,878	0,0164	547	581	547	2
442	158	99,877	0,0165	12	490	12	2
443	157	99,875	0,0167	6	18	6	66
444	156	99,874	0,0169	46	147	46	3
445	155	99,871	0,0172	25	127	25	32
446	154	99,868	0,0177	99	274	99	2
447	153	99,859	0,0189	31	211	31	5
448	152	99,855	0,0194	159	276	159	4
449	151	99,854	0,0196	222	492	222	3
450	150	99,854	0,0196	71	545	71	5
451	149	99,842	0,0212	422	436	422	2
452	148	99,839	0,0215	135	461	135	2
453	147	99,831	0,0227	152	190	152	3
454	146	99,818	0,0244	58	119	58	5
455	145	99,813	0,0251	14	24	14	6
456	144	99,809	0,0256	191	324	191	6

457	143	99,806	0,0260	55	246	55	3
458	142	99,805	0,0261	73	212	73	11
459	141	99,800	0,0269	248	350	248	6
460	140	99,799	0,0269	4	272	4	5
461	139	99,798	0,0271	7	94	7	4
462	138	99,794	0,0277	60	131	60	3
463	137	99,779	0,0296	314	332	314	3
464	136	99,777	0,0298	124	166	124	6
465	135	99,774	0,0302	85	357	85	3
466	134	99,774	0,0302	162	215	162	5
467	133	99,771	0,0306	473	475	473	2
468	132	99,768	0,0311	194	243	194	6
469	131	99,767	0,0312	106	312	106	6
470	130	99,765	0,0315	183	236	183	5
471	129	99,765	0,0315	1	25	1	204
472	128	99,757	0,0326	53	103	53	2
473	127	99,753	0,0330	31	411	31	7
474	126	99,750	0,0336	35	228	35	6
475	125	99,749	0,0336	17	257	17	3
476	124	99,748	0,0337	114	322	114	2
477	123	99,745	0,0341	87	126	87	6
478	122	99,738	0,0351	426	438	426	3
479	121	99,736	0,0353	6	59	6	71
480	120	99,736	0,0354	97	345	97	4
481	119	99,734	0,0356	137	261	137	24
482	118	99,729	0,0364	471	477	471	2
483	117	99,725	0,0369	296	482	296	4
484	116	99,709	0,0390	43	159	43	6
485	115	99,702	0,0400	213	222	213	7
486	114	99,694	0,0410	48	157	48	5
487	113	99,694	0,0411	9	420	9	4
488	112	99,684	0,0424	66	71	66	7
489	111	99,678	0,0431	239	400	239	4
490	110	99,675	0,0436	12	69	12	3
491	109	99,674	0,0437	101	143	101	3
492	108	99,668	0,0445	373	422	373	3
493	107	99,666	0,0447	112	191	112	7
494	106	99,664	0,0450	33	108	33	6
495	105	99,662	0,0452	55	584	55	4
496	104	99,656	0,0461	13	73	13	25
497	103	99,654	0,0463	447	466	447	2
498	102	99,653	0,0466	399	409	399	2
499	101	99,635	0,0489	260	384	260	2
500	100	99,633	0,0492	113	152	113	7
501	99	99,633	0,0492	19	121	19	4
502	98	99,625	0,0503	26	534	26	7
503	97	99,615	0,0516	14	498	14	9
504	96	99,614	0,0517	464	468	464	2
505	95	99,613	0,0519	252	286	252	2
506	94	99,610	0,0523	135	319	135	4
507	93	99,609	0,0523	138	443	138	5
508	92	99,608	0,0526	22	194	22	11
509	91	99,587	0,0553	54	199	54	4
510	90	99,557	0,0594	87	220	87	8
511	89	99,556	0,0595	7	92	7	6
512	88	99,514	0,0652	314	547	314	5
513	87	99,504	0,0665	46	149	46	4
514	86	99,492	0,0681	40	174	40	3
515	85	99,484	0,0692	99	512	99	4
516	84	99,459	0,0725	169	278	169	3
517	83	99,454	0,0731	79	183	79	9
518	82	99,452	0,0735	171	586	171	2
519	81	99,427	0,0768	9	255	9	6
520	80	99,421	0,0776	58	188	58	7
521	79	99,409	0,0792	140	445	140	2

522	78	99,405	0,0797	76	268	76	2
523	77	99,401	0,0802	17	552	17	4
524	76	99,397	0,0808	106	248	106	12
525	75	99,379	0,0832	1	13	1	229
526	74	99,346	0,0877	53	81	53	3
527	73	99,339	0,0886	6	112	6	78
528	72	99,324	0,0905	2	43	2	10
529	71	99,307	0,0929	22	124	22	17
530	70	99,277	0,0969	4	35	4	11
531	69	99,258	0,0994	48	97	48	9
532	68	99,247	0,1009	162	213	162	12
533	67	99,244	0,1013	26	85	26	10
534	66	99,213	0,1055	19	426	19	7
535	65	99,199	0,1073	12	66	12	10
536	64	99,161	0,1125	31	399	31	9
537	63	99,134	0,1160	60	135	60	7
538	62	99,119	0,1181	46	83	46	5
539	61	99,076	0,1238	54	148	54	5
540	60	99,064	0,1255	113	396	113	9
541	59	99,035	0,1294	17	111	17	6
542	58	98,949	0,1409	14	79	14	18
543	57	98,943	0,1417	87	296	87	12
544	56	98,857	0,1533	1	239	1	233
545	55	98,846	0,1547	447	464	447	4
546	54	98,834	0,1562	101	252	101	5
547	53	98,788	0,1625	471	473	471	4
548	52	98,745	0,1682	33	106	33	18
549	51	98,682	0,1766	260	382	260	3
550	50	98,656	0,1801	38	40	38	4
551	49	98,654	0,1804	58	373	58	10
552	48	98,570	0,1917	55	76	55	6
553	47	98,569	0,1918	6	114	6	80
554	46	98,568	0,1919	138	576	138	6
555	45	98,547	0,1948	162	314	162	17
556	44	98,534	0,1964	7	9	7	12
557	43	98,524	0,1978	48	99	48	13
558	42	98,454	0,2073	171	288	171	3
559	41	98,381	0,2170	154	394	154	3
560	40	98,218	0,2388	2	26	2	20
561	39	98,080	0,2573	14	22	14	35
562	38	97,989	0,2695	4	87	4	23
563	37	97,888	0,2830	53	101	53	8
564	36	97,724	0,3051	140	447	140	6
565	35	97,710	0,3070	12	33	12	28
566	34	97,660	0,3136	1	54	1	238
567	33	97,463	0,3401	7	19	7	19
568	32	97,423	0,3453	58	113	58	19
569	31	97,211	0,3738	46	48	46	18
570	30	96,908	0,4144	138	522	138	7
571	29	96,829	0,4250	17	38	17	10
572	28	96,788	0,4304	169	171	169	6
573	27	96,625	0,4524	260	262	260	4
574	26	96,554	0,4619	14	162	14	52
575	25	96,426	0,4790	12	31	12	37
576	24	96,168	0,5136	7	60	7	26
577	23	95,870	0,5536	2	4	2	43
578	22	95,693	0,5773	154	375	154	4
579	21	95,561	0,5949	53	55	53	14
580	20	94,075	0,7941	138	140	138	13
581	19	93,334	0,8935	137	260	137	28
582	18	93,277	0,9011	1	12	1	275
583	17	93,195	0,9120	7	58	7	45
584	16	92,965	0,9429	6	17	6	90
585	15	92,943	0,9458	2	46	2	61
586	14	91,664	1,1173	53	169	53	20

587	13	89,703	1,3801	117	154	117	5
588	12	89,559	1,3994	1	14	1	327
589	11	87,936	1,6168	138	471	138	17
590	10	87,240	1,7102	137	470	137	33
591	9	86,156	1,8555	53	74	53	21
592	8	85,282	1,9727	6	7	6	135
593	7	82,044	2,4065	1	2	1	388
594	6	80,139	2,6620	53	117	53	26
595	5	80,022	2,6776	137	138	137	50
596	4	79,004	2,8141	1	6	1	523
597	3	65,743	4,5914	1	53	1	549
598	2	53,951	6,1718	1	137	1	599
599	1	0,000	13,4027	1	51	1	600

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

Βιβλιογραφία

Ελληνική

- “Στατιστικές Μέθοδοι Εξόρυξης Δεδομένων”, Πελέκης Ν., Θεοδωρίδης Ι., Πειραιάς 2012
- “Πρόβλεψη σε Ημερολόγια Κίνησης με Χρήση Νευρωνικών Δικτύων”, ΧΟΝΔΡΟΔΗΜΑ ΕΥΑΓΓΕΛΙΑ, μεταπτυχιακή διατριβή, Πειραιάς 2013

Ξένα

- ANOOP KUMAR JAIN, SATYAM MAHESWARI, Survey of Recent Clustering Techniques in Data Mining, *International Archive of Applied Sciences and Technology*, June 2012Q pp 68- 75
- Hierarchical Clustering, DAVID M. BLEI, Princeton University February 28,2008
- Hierarchical Clustering, Cambridge University Press, April 1, 2009, online edition © 2009, Cambridge UP
- MICRO NANNI, DINO PEDRESCHI “Time focused clustering of trajectories of moving objects”, November 2006, volume 27, pp 267- 289, Springer
- ALVARES- BOGORNY “A Clustering- Based Approach for Discovering Interesting Places in Trajectories”, December 2008, Porto Alegre
- EN. APRIT GUPTA, ER. ANKIT GUPTA, ER. AMIT MISHRA, *Research paper on cluster techniques of data variations/ International Journal of Advance Technology and Engineering Research/ Vol. 1, Issue 1, November 2011*
- PARENT et al., *Semantic Trajectories Modeling and Analysis*, 2009-2012 Programme, Coordination Action type Project MODAP (Mobility, Data Mining, and Privacy) <http://www.modap.org>
- XIE K., DENG K. AND ZHOU X. 2009. From trajectories to activities: a spatio-temporal join approach. *GIS-LBSN 2009*: 25-32
- TIAKAS E., PAPADOPOULOS A.N., NANOPOULOS A., MANOLOPOULOS Y., STOJANOVIC D. AND DJORDJEVIC-KAJAN S. 2009. Searching for similar trajectories in spatial networks. *The Journal of Systems and Software*, 82(5), May 2009
- SPINSANTI L., CELLI F. AND RENSO C. 2010. Where you stop is who you are: understanding peoples' activities. in *5th BMI'10, User Behaviour Modelling*, Karlsruhe, Germany
- SPACCAPIETRA S., PARENT C., DAMIANI M.L., MACEDO J.A., PORTO F. AND VANGENOT C. 2008. A conceptual view on trajectories. *Data & Knowledge Engineering* 65: 126–146
- PELEKIS N., KOPANAKIS I., KOTSIFAKOS E.E., FRENTZOS E. AND THEODORIDIS Y. 2011b Clustering uncertain trajectories. *Knowledge and Information Systems* 28(1): 117- 147
- PANAGIOTAKIS C., PELEKIS N., KOPANAKIS I., RAMASSO E. AND THEODORIDIS Y. 2012. Segmentation and Sampling of Moving Object Trajectories based on Representativeness. *IEEE Transactions on Knowledge and Data Engineering*

- PALMA A.T., BOGORNY V., KUIJPERS B. AND ALVARES L.O. 2008. A clustering-based approach for discovering interesting places in trajectories. *ACMSAC*, pp. 863–868, New York, NY, USA. ACM Press
- PELEKIS N., ANDRIENKO G., ANDRIENKO N., KOPANAKIS I., MARKETOS G., THEODORIDIS Y. 2011a Visually Exploring Movement Data via Similarity-based Analysis. *Journal of Intelligent Information Systems*: 129-140
- LEE J.-G., HAN J. AND WHANG K.-Y. 2007. Trajectory clustering: a partition-and-group framework. In *Proc. SIGMOD Conf.*, 593–604. ACM
- KALNIS P., MAMOULIS N. AND BAKIRAS S. 2005. On discovering moving clusters in spatio-temporal data. *Proc. of 9th Int. Symposium on Spatial and Temporal Databases*, pages 364--381. Springer
- BAGLIONI M., DE MACEDO J.A.F., RENSO C., TRASARTI R. AND WACHOWICZ M. 2012. How you move reveals who you are: understanding human behavior by analyzing trajectory data. *Knowledge and Information System Journal (KAIS), special issue on Behavior Computing*
- ASHBROOK D. AND STARNER T. 2003. Using GPS to learn significant locations and predict movement across multiple users, *Personal Ubiquitous Computing*, 7, 275–286
- ANDRIENKO G., ANDRIENKO N. AND WROBEL S. 2007a. Visual Analytics Tools for Analysis of Movement Data, *ACM SIGKDD Explorations*, 2007, v.9 (2): 38-46
- KETTERLIN, “*Clustering Sequences of Complex Objects*”, 1997, American Association for Artificial Intelligence (www.aaai.org)
- RINZIVILLO S., PEDRESCHI D., NANNI M., GIANNOTTI F., ANDRIENKO N. AND ANDRIENKO G. 2008. Visually-driven analysis of movement data by progressive clustering, *Information Visualization*, 7(3/4): 225-239
- ROCHA J.A.M., TIMES V.C., OLIVEIRA G., ALVARES L.O. AND BOGORNY V. 2010. DB-SMoT: A direction-based spatio-temporal clustering method. *IEEE Conf. of Intelligent Systems 2010*, 114-119
- ZHENG, Y., ZHANG, L., MA, Z., XIE, X., AND MA, W-Y. 2011. Recommending Friends and Locations Based on Individual Location History, *Journal ACM Transactions on the Web (TWEB)*, 5(1): 1559-1131
- Minitab Tutorial, [www.csyue.nccu.edu.tw/ch/Multivariate_Analysis\(Minitab\).pdf](http://www.csyue.nccu.edu.tw/ch/Multivariate_Analysis(Minitab).pdf)