



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

ΕΞΕΤΑΣΗ ΤΕΧΝΙΚΩΝ ΑΝΑΛΥΣΗΣ ΧΩΡΟΧΡΟΝΙΚΩΝ
ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΗ ΜΕΛΕΤΗ ΤΗΣ ΚΙΝΗΣΗΣ
ΑΝΘΡΩΠΩΝ ΚΑΙ ΕΦΑΡΜΟΓΗ ΣΕ ΕΜΠΕΙΡΙΚΑ
ΔΕΔΟΜΕΝΑ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΤΟΥ
ΚΩΝΣΤΑΝΤΙΝΟΥ ΚΑΠΕΛΙΟΥ

Επιβλέπων: Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2021

Η σελίδα αυτή είναι σκόπιμα λευκή



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

ΕΞΕΤΑΣΗ ΤΕΧΝΙΚΩΝ ΑΝΑΛΥΣΗΣ ΧΩΡΟΧΡΟΝΙΚΩΝ
ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΗ ΜΕΛΕΤΗ ΤΗΣ ΚΙΝΗΣΗΣ
ΑΝΘΡΩΠΩΝ ΚΑΙ ΕΦΑΡΜΟΓΗ ΣΕ ΕΜΠΕΙΡΙΚΑ
ΔΕΔΟΜΕΝΑ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΤΟΥ
ΚΩΝΣΤΑΝΤΙΝΟΥ ΚΑΠΕΛΙΟΥ

Επιβλέπων: Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 18^η Μαρτίου 2021.

.....
Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

.....
Ιωάννης Ψαρράς
Καθηγητής Ε.Μ.Π.

.....
Χρυσόστομος Δούκας
Αν. Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2021

.....
ΚΩΝΣΤΑΝΤΙΝΟΣ ΚΑΠΕΛΙΟΣ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © ΚΩΝΣΤΑΝΤΙΝΟΣ ΚΑΠΕΛΙΟΣ, 2021

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η ζωή των ανθρώπων είναι άρρηκτα συνδεδεμένη με τη μετακίνηση στο χώρο και το χρόνο. Ανά τους αιώνες, η ζωή και η κοινωνική φύση των ανθρώπων εξαρτάται από τη μετακίνηση σχεδόν σε καθημερινή βάση. Η ανακάλυψη των νόμων που διέπουν την κίνηση ατόμων και πληθυσμών και η πρόβλεψη των μελλοντικών της καταστάσεων έχει μεγάλη αξία και κομβική σημασία για πολλούς επιστημονικούς και οικονομικούς τομείς, όπως την πολεοδομική και συγκοινωνιακή σχεδίαση και την μοντελοποίηση της εξέλιξης των επιδημιών.

Στο πλαίσιο αυτό, γίνεται μια προσπάθεια αποτίμησης των κυριότερων πηγών δεδομένων, μονάδων μέτρησης και τεχνικών ανάλυσης της ανθρώπινης κινητικότητας. Αφού πραγματοποιηθεί η καταγραφή των ανωτέρω, στην παρούσα διπλωματική εργασία πραγματοποιείται μια πειραματική εφαρμογή σε εμπειρικά δεδομένα ατομικής κινητικότητας και ροών πληθυσμών, έχοντας ως στόχο την εξαγωγή συμπερασμάτων και αξιολόγηση των παραγόμενων αποτελεσμάτων.

Σε επίπεδο ατομικής κινητικότητας, αναλύονται δεδομένα μετακίνησης 10 ατόμων στην περιοχή των Χανίων διάρκειας 8 μηνών, ενώ σε επίπεδο ροών πληθυσμών πραγματοποιείται ανάλυση δεδομένων πριν την έναρξη και κατά τη διάρκεια της πανδημίας της COVID-19 στην Ιταλία του 2020. Και στα δύο είδη κινητικότητας εξάγονται βασικές μονάδες μέτρησης της και οπτικοποιούνται τα αποτελέσματα της ανάλυσης. Σε επίπεδο κινήσεων πληθυσμών, δημιουργούνται τεχνητές ροές μέσω του βαρυτικού μοντέλου της ανθρώπινης κίνησης, οι οποίες εξετάζονται και συγκρίνονται με τις πραγματικές.

Λέξεις Κλειδιά: ανθρώπινη κινητικότητα, ατομική κινητικότητα, πηγές δεδομένων ανθρώπινης κινητικότητας, επισήμανση τοποθεσιών, πίνακες πρόελευσης-προορισμού, βαρυτικό μοντέλο

Abstract

Human life is inextricably connected with the movement in space and time. Over the centuries, the life and social nature of people depend on mobility on an almost daily basis. Discovering the laws that govern the movement of people or groups of people and predicting its future states are of great value and major importance for many scientific and economic fields, such as urban and transportation planning and modeling the evolution of infectious diseases.

Within this framework, an attempt is made to evaluate the most common data sources, units of measurement, and techniques of analyzing human mobility. After collecting and writing down the above in this dissertation, an experimental application is made by analyzing empirical data of individual movement and population flows to extract conclusions and evaluate the results.

At the individual mobility level, data on the movement of 10 people in the area of Chania lasting 8 months are analyzed, while at the level of population flows, data analysis is performed before the beginning and during the 2020 COVID-19 pandemic in Italy. In both types of mobility, basic metrics are extracted and the results of the analysis are visualized. At the population flows' level, artificial flows are created by applying the gravity model of human mobility, which are then examined and compared with the actual ones.

Keywords: human mobility, individual mobility, data sources for human mobility, location labeling, origin destination matrices, gravity model

Ευχαριστίες

Θα ήθελα να εκφράσω καταρχάς τις ευχαριστίες μου στον καθηγητή μου κ. Δημήτριο Ασκούνη για την εμπιστοσύνη του στην ανάθεση και επίβλεψη αυτής της διπλωματικής εργασίας. Επίσης, ευχαριστώ θερμά την υποψήφια διδάκτορα κ. Αριάδνη Μιχαλίτση - Ψαρρού για την άριστη συνεργασία και καθοδήγηση καθ'όλη τη διάρκεια της εκπόνησης της διπλωματικής μου εργασίας.

Με την ευκαιρία αυτή θα ήθελα επίσης να εκφράσω την ευγνωμοσύνη μου και τις ευχαριστίες μου προς τους γονείς μου, Κοσμά Γεωργία και Κατέλιο Ιωάννη, στον αδερφό μου Κατέλιο Χρήστο και τους στενούς μου φίλους και συμφοιτητές με τους οποίους μοιράστηκα τα χρόνια των σπουδών μου στο Εθνικό Μετσόβιο Πολυτεχνείο.

Περιεχόμενα

Κατάλογος Σχημάτων	9
Κατάλογος Πινάκων	11
1 Εισαγωγή	12
1.1 Σκοπός	12
1.2 Συνεισφορά	12
1.3 Δομή	13
2 Θεωρητικό Υπόβαθρο	14
2.1 Ιστορική Αναδρομή	14
2.2 Εφαρμογές της μελέτης της ανθρώπινης κινητικότητας	15
2.3 Βασικές μονάδες για τη μέτρηση της ανθρώπινης κινητικότητας	17
2.3.1 Μήκη άλματος (jump lengths)	18
2.3.2 Ακτίνα περιστροφής (RoG)	19
2.3.3 Κ-οστή ακτίνα περιστροφής (k-RoG)	20
2.3.4 Αριθμός επισκεφθεισών τοποθεσιών	21
2.3.5 Μέση τετραγωνική μετατόπιση	22
2.4 Ανακάλυψη γνώσης σε βάσεις δεδομένων	25
2.5 Προστασία της ιδιωτικότητας	27
2.6 Πηγές δεδομένων για την ανάλυση της ανθρώπινης κινητικότητας	28
2.6.1 Ανάλυση πηγών δεδομένων	29
2.6.2 Αξιοποίηση των δεδομένων στην ανάλυση της ανθρώπινης κινητικότητας	33
2.6.3 Ψηφιακό Αποτύπωμα	36

2.6.4	Συγκριτική παρουσίαση πηγών δεδομένων μέτρησης ανθρώπινης κινητικότητας	37
3	Περιγραφή Μεθοδολογιών	42
3.1	Μοντέλα υπολογισμού ατομικής κινητικότητας	43
3.1.1	Κίνηση Μπράουν	44
3.1.2	Πτήσεις Λέβι	44
3.2	Βαρυτικά μοντέλα	46
3.2.1	Αξιοποίηση των βαρυτικών μοντέλων στην ανάλυση ανθρώπινης κινητικότητας	47
3.3	Rule-based Μοντέλα	49
3.4	Μοντέλα βασισμένα σε αλυσίδες Μαρκόβ	52
3.4.1	Τάξης-k μοντέλα Μαρκόβ	55
3.4.2	Τάξης-k μοντέλα Μαρκόβ με εξωτερικές πληροφορίες	58
3.4.3	Κρυφά Μοντέλα Μαρκόβ	59
3.4.4	Αξιοποίηση των μοντέλων Μαρκόβ στην ανάλυση ανθρώπινης κινητικότητας	59
3.5	Νευρωνικά μοντέλα	61
3.5.1	Νευρωνικά μοντέλα με ανάδραση	61
3.5.2	Αξιοποίηση των νευρωνικών μοντέλων στην ανάλυση της ανθρώπινης κινητικότητας	65
3.6	Συγκριτική παρουσίαση μεθοδολογιών, σύνοψη	67
4	Πειραματική Εφαρμογή Μεθοδολογιών Σε Εμπειρικά Δεδομένα	72
4.1	Γλώσσα προγραμματισμού και βιβλιοθήκες	72
4.2	Μοντέλο ατομικής κινητικότητας - Ανάλυση δεδομένων καταγραφών λεπτομερειών κλήσεων (CDR) για μετακινήσεις στην περιοχή των Χανίων	74
4.2.1	Σύνολο δεδομένων (dataset)	76
4.2.2	Προ-επεξεργασία δεδομένων	78
4.2.3	Εξαγωγή μονάδων μέτρησης κινητικότητας	80
4.2.4	Αναγνώριση θέσεων οικίας και παρουσίαση αποτελεσμάτων	87

4.3	Μοντέλο κινήσεων πληθυσμών - Ανάλυση δεδομένων κινητικότητας μεταξύ των επαρχιών της Ιταλίας πριν και κατά τη διάρκεια της πανδημίας COVID-19	91
4.3.1	Σύνολο δεδομένων	92
4.3.2	Προ-επεξεργασία δεδομένων	95
4.3.3	Επίδραση των περιοριστικών μέτρων στην κίνηση μεταξύ επαρχιών	96
4.3.4	Σύγκριση μετακινήσεων πριν και κατά τη διάρκεια της πανδημίας	100
4.3.5	Δημιουργία συνθετικών ροών	105
4.4	Σύνοψη αποτελεσμάτων πειραματικής εφαρμογής	113
4.4.1	Σύνοψη αποτελεσμάτων ατομικής κινητικότητας	113
4.4.2	Σύνοψη αποτελεσμάτων κινήσεων πληθυσμών	114
5	Συμπεράσματα και Μελλοντικές Προεκτάσεις	116
5.1	Συμπεράσματα	116
5.2	Μελλοντικές προεκτάσεις	117
	Βιβλιογραφία	119

Κατάλογος Σχημάτων

2.1	Μέση τετραγωνική μετατόπιση έναντι χρόνου, για ομάδες ατόμων με διαφορετικές ακτίνες περιστροφής [1].	24
2.2	Τα βήματα που συνθέτουν μια διαδικασία ανακάλυψης πληροφορίας (KDD, Knowledge Data Discovery) [2].	27
3.1	Δέντρο επιδημιολογικής εισροής [3].	48
3.2	Συσχέτιση του αριθμού των μοναδικών αναγνωρισθισών τοποθεσιών και του συνολικού αριθμού ταξιδιών [4]	51
3.3	Ένα παράδειγμα αλυσίδας Μαρκόβ [5].	54
3.4	RNN με συνδέσεις ανατροφοδότησης των κρυφών μονάδων [6].	64
3.5	RNN με ανάδραση από την έξοδο (output layer) προς τους νευρώνες κάποιου κρυφού επιπέδου (hidden units). [6].	64
3.6	Χρονικά αναδιπλωμένο νευρωνικό δίκτυο με ανατροφοδότηση των κρυφών μονάδων στον εαυτό τους και με μια έξοδο στο τέλος της ακολουθίας. Το δίκτυο αυτό μπορεί να χρησιμοποιηθεί για να συνοψιστεί και να χαρακτηριστεί μια ακολουθία δεδομένων [6].	65
4.1	Παρουσίαση της ιστοσελίδας MySignals [7].	75
4.2	Ακτίνα περιστροφής ανά αναγνωριστικό συσκευής.	82
4.3	Μέγιστη διανυθείσα απόσταση σε χιλιόμετρα ανά συσκευή.	84
4.4	Λογαριθμική σύγκριση αριθμού διακριτών τοποθεσιών που αναγνωρίστηκαν και συνολικού αριθμού καταγραφών μετακίνησης χρηστών.	87
4.5	Οι ανιχνευθείσες θέσεις οικίας των συμμετεχόντων στην έρευνα MySignals.	89

4.6	Εσφαλμένη αναγνώριση θέσης οικίας ενός χρήστη της έρευνας MySignals.	90
4.7	Σύγκριση των κατοικιών χρηστών που αναγνωρίστηκαν αυτόματα από τον αλγόριθμο (αριστερό σχήμα) σε σχέση με τις πραγματικές τοποθεσίες των σπιτιών τους (δεξί σχήμα). Δεξί σχήμα από [7].	91
4.8	Ροές από και προς την επαρχία Lodi πριν και μετά την ανακοίνωση περιοριστικών μέτρων. Με μπλε χρώμα σημειώνεται το σύνολο των ανθρώπων που έφευγε εκτός της επαρχίας, ενώ με πορτοκαλί το σύνολο των πολιτών που εισέρχονταν σε αυτήν.	98
4.9	Εβδομαδιαίος μέσος όρος ροών μεταξύ των επαρχιών και άθροισμα εβδομαδιαίων νέων κρουσμάτων, για την περίοδο από 20/1/2020 έως και 12/7/2020.	99
4.10	Εβδομαδιαίο άθροισμα νέων κρουσμάτων COVID-19 στην Ιταλία, από 15/06 έως και 29/11.	100
4.11	Μέση κανονικοποιημένη απόσταση Λέβενσταιν (mean NLOD) μετακινήσεων μεταξύ περιοχών. Στο σχήμα, συγκρίνονται οι οχτώ πρώτες βδομάδες του διαστήματος που μελετάται με τις εβδομάδες που ακολουθούν.	103
4.12	Μέση κανονικοποιημένη απόσταση Λέβενσταιν (mean NLOD) μετακινήσεων μεταξύ περιοχών. Στο σχήμα συγκρίνονται οι εβδομάδες από 23/3 έως και 21/6/2020 με τις επόμενές τους.	104
4.13	Μέση κανονικοποιημένη απόσταση Λέβενσταιν (mean NLOD) μετακινήσεων μεταξύ περιοχών. Στο σχήμα συγκρίνονται οι επτά τελευταίες εβδομάδες με τις επόμενές τους.	105
4.14	Ροές άνω των 10.000 ατόμων μεταξύ των επαρχιών της Ιταλίας από 20 έως 26/1/2020.	107
4.15	Οπτικοποίηση των πραγματικών ροών άνω των 10.000 ατόμων μεταξύ των επαρχιών της Ιταλίας από 20-26/1/2020 (αριστερά) και των συνθετικών ροών που δημιουργήθηκαν από το βαρυτικό μοντέλο (δεξιά).	110
4.16	Ροές άνω των 10.000 ατόμων μεταξύ των επαρχιών της Ιταλίας από 30/3 έως 5/4/2020.	111

4.17 Οπτικοποίηση των πραγματικών ροών άνω των 10.000 ατόμων μεταξύ των επαρχιών της Ιταλίας από 30/4-5/4/2020 (αριστερά) και των συνθετικών ροών που δημιουργήθηκαν από το βαρυτικό μοντέλο (δεξιά).	113
---	-----

Κατάλογος Πινάκων

2.1	Συγκριτικός πίνακας των συνόλων δεδομένων (datasets) που συναντώνται συχνότερα στη βιβλιογραφία.	41
3.1	Ακρίβεια αναγνώρισης δραστηριότητας ανά κατηγορία δραστηριότητας της και ο συνολικός αριθμός αναγνωρισθέντων σημείων ενδιαφέροντος [8].	52
3.2	Συγκριτικός πίνακας μοντέλων προσομοίωσης ανθρώπινης κίνησης. . .	68
3.3	Συγκριτικός πίνακας μοντέλων ανάλυσης ανθρώπινης κίνησης.	68
4.1	Τα πεδία του συνόλου δεδομένων και μια σύντομη περιγραφή τους [7]. .	77
4.2	Μοναδικά αναγνωριστικά συσκευών και συνολικός αριθμός καταγραφών ανά συσκευή.	79
4.3	Πίνακας αντιστοίχισης αναγνωριστικού συσκευή με ψευδώνυμο.	80
4.4	K-οστή ακτίνα περιστροφής των μοναδικών συσκευών της έρευνας MySignals σε χιλιόμετρα, για K=2,5 και 10.	83
4.5	Συνολικός αριθμός διακριτών τοποθεσιών που αναγνωρίστηκαν ανά συσκευή.	85
4.6	Εντροπία θέσης ανά συσκευή της έρευνας MySignals.	86
4.7	Δυνητικά αποτελέσματα εκτέλεσης κακόβουλης επίθεσης για αποκάλυψη προσωπικών δεδομένων. Η κλίμακα ρίσκου είναι κανονικοποιημένη.	88
4.8	Ημερομηνίες ανακοίνωσης περιοριστικών μέτρων καθώς και η ημερομηνία ανακοίνωσης της άρσης τους.	97
4.9	Συγκριτικός πίνακας της μέσης απόστασης Λεβενστάιν μεταξύ των μελετώμενων εβδομάδων, από 20/1/2020 έως και 21/06/2020.	102

- 4.10 Μέση κανονικοποιημένη απόσταση Λέβενσταιν μεταξύ των πραγματικών και των συνθετικών δεδομένων που παρήχθησαν για την εβδομάδα 20-26/1/2020. 109
- 4.11 Μέση κανονικοποιημένη απόσταση Λέβενσταιν μεταξύ των πραγματικών και των συνθετικών δεδομένων που παρήχθησαν για την εβδομάδα αυστηρών περιορισμών 30/4-5/4/2020. 112

1. Εισαγωγή

1.1 Σκοπός

Οι πρωταρχικοί στόχοι της διπλωματικής αυτής εργασίας μπορούν να συνοψιστούν σε δύο σημεία. Το πρώτο σημείο αφορά την εξέταση των διάφορων πηγών δεδομένων και τεχνικών ανάλυσης ανθρώπινης κίνησης καθώς και τον προσδιορισμό ενός συγκριτικού πλαισίου αυτών. Ως δεύτερος στόχος της εργασίας, είναι η πειραματική εφαρμογή κάποιων χαρακτηριστικών τεχνικών εξαγωγής συμπερασμάτων ανθρώπινης κίνησης και η ανάλυση των αποτελεσμάτων που προκύπτουν από την εφαρμογή αυτών.

1.2 Συνεισφορά

Η συνεισφορά της διπλωματική εργασίας αποτυπώνεται στα παρακάτω στάδια:

- Εξετάστηκαν οι υπάρχουσες τεχνικές ανάλυσης ανθρώπινης κινητικότητας και εντοπίστηκαν τα χαρακτηριστικά στοιχεία αυτών.
- Σημειώθηκαν επίσης οι διάφορες πηγές δεδομένων για την ανάλυση της κίνησης των ανθρώπων.
- Παρουσιάστηκαν και διατυπώθηκαν οι κύριες μονάδες μέτρησης και χαρακτηρισμού της ανθρώπινης κίνησης.
- Εξετάστηκε η διαδικασία συλλογής πληροφοριών και συνόλων δεδομένων στο πλαίσιο της προστασίας των προσωπικών δεδομένων.
- Πραγματοποιήθηκε πειραματική εφαρμογή τεχνικών ανάλυσης ανθρώπινης κίνησης σε ατομικό επίπεδο όσο και σε επίπεδο ροών πληθυσμού, αξιοποιώντας σύγχρονες τεχνικές ανάλυσης και σύγκρισης αποτελεσμάτων.

1.3 Δομή

Η παρούσα διπλωματική εργασία αποτελείται από έξι κεφάλαια.

Το πρώτο κεφάλαιο συνιστά μια εισαγωγή στο θέμα της διπλωματικής εργασίας, παρουσιάζοντας το πρόβλημα και την προσέγγιση στην επίλυση του.

Το δεύτερο κεφάλαιο αποτελεί το θεωρητικό υπόβαθρο της εργασίας, στο οποίο γίνεται μια σύντομη ιστορική αναδρομή και παρουσιάζονται βασικές έννοιες και θεωρίες της ανθρώπινης κινητικότητας.

Το τρίτο κεφάλαιο σκιαγραφεί τις βασικές μεθοδολογίες πρόβλεψης της ανθρώπινης κινητικότητας. Αναλύονται οι περισσότερες τεχνικές που προκύπτουν στη βιβλιογραφία σε βάθος και στο τέλος του κεφαλαίου συνοψίζονται τα βασικά τους στοιχεία και οι διαφορές τους.

Στο τέταρτο κεφάλαιο γίνεται πειραματική εφαρμογή μεθοδολογιών ανάλυσης ανθρώπινης κίνησης σε ατομικό επίπεδο και σε επίπεδο ροών πληθυσμών.

Στο πέμπτο κεφάλαιο διατυπώνονται οι πιθανές επεκτάσεις και μελλοντικές προοπτικές της εργασίας.

Τέλος, παρατίθεται η βιβλιογραφία που χρησιμοποιήθηκε για την εκπόνηση της διπλωματικής εργασίας.

2. Θεωρητικό Υπόβαθρο

Ανθρώπινη κινητικότητα Ο όρος ανθρώπινη κινητικότητα (human mobility) έχει πολλούς και διαφορετικούς ορισμούς. Μια κοινώς αποδεκτή ερμηνεία είναι πως αναφέρεται στην μετακίνηση ατόμων και ομάδων στο χώρο και το χρόνο.

2.1 Ιστορική Αναδρομή

Η πορεία της ανθρώπινης ιστορίας κατά τη διάρκεια των αιώνων έχει συνδεθεί άμεσα με την ανθρώπινη κινητικότητα. Στην παλαιολιθική εποχή, 70 χιλιάδες χρόνια πριν, ο *homo sapiens* αναζητώντας φιλόξενα και πρόσφορα εδάφη για να εγκατασταθεί μετανάστευσε από την Αφρική στην Ευρώπη. Πιο πρόσφατα, στην εποχή των εξερευνήσεων οι Ευρωπαίοι έφυγαν για να ανακαλύψουν το Νέο Κόσμο. Ακόμη και σήμερα, οι άνθρωποι, ωθούμενοι από κοινωνικοοικονομικές ανάγκες μετακινούνται με αδιάκοπους ρυθμούς. Οι αιτίες για την μετακίνηση των ανθρώπων στην ιστορία δεν ήταν πάντοτε οι ίδιες.

Η ανθρωπότητα ανά τους αιώνες είναι συνυφασμένη με τη μετακίνηση σε καθημερινή βάση για την εξασφάλιση των απαραίτητων ειδών διαβίωσης. Επίσης, η κοινωνική ύπαρξη του ανθρώπου συνδέεται άρρηκτα με την καθημερινή αυτή μετακίνηση. Οι χωροχρονική κλίμακα των καθημερινών μετακινήσεων είναι αρκετά μικρότερη από αυτή των μεταναστευτικών ροών. Ωστόσο, χαρακτηρίζεται από την περιοδικότητα της ανθρώπινης ζωής.

Η μελέτη της ανθρώπινης κινητικότητας άπτεται σε πολλά επιστημονικά πεδία, ωστόσο, η πρώτη επιστήμη που ασχολήθηκε με τη μελέτη της και ανέλυσε δεδομένα παρουσιάζοντας σχετικά θεωρήματα είναι η γεωγραφία [9]. Τον 20ο αιώνα, η θεωρία της ανθρώπινης κινητικότητας συνεισέφερε στην επιστήμη της γεωγραφίας, της κοινωνιολογίας και της οικονομικής θεωρίας.

Τα τελευταία χρόνια, η αυξανόμενη διαθεσιμότητα δεδομένων για τις μετακινήσεις πληθυσμών σε διάφορα επίπεδα ανάλυσης σε συνδυασμό με την άνθηση των γεωγραφικών δεδομένων, οδήγησε στην εισαγωγή πιο εξελιγμένων μαθηματικών μεθόδων όπως τα μοντέλα Μαρκόβ και τα νευρωνικά δίκτυα. Φύλο και

κοινωνικοοικονομικοί παράγοντες πίσω από την κίνηση του πληθυσμού αναλύθηκαν περαιτέρω χάρη σε μεγαλύτερα σύνολα δεδομένων που προέκυψαν από έρευνες και συνεντεύξεις. Επίσης, αναπτύχθηκαν νέα μοντέλα προσομοίωσης της ανθρώπινης κίνησης και σήμερα οι ερευνητές έχουν στη διάθεσή τους εργαλεία που δίνουν τη δυνατότητα παραγωγής τεχνητών συνόλων δεδομένων, τα οποία μπορούν να προσεγγίσουν με μεγάλη ακρίβεια τα πραγματικά [10].

2.2 Εφαρμογές της μελέτης της ανθρώπινης κινητικότητας

Οι εφαρμογές της ανάλυσης της ανθρώπινης κινητικότητας είναι ποικίλες. Οι πιο συχνές εξ αυτών εντοπίζονται στην παρακολούθηση μετάδοσης μολυσματικών ασθενειών, τον εντοπισμό της διάδοσης κακόβουλου λογισμικού, την πολεοδομική σχεδίαση και τη συγκοινωνιακή μηχανική.

Εντοπισμός μετάδοσης μολυσματικών ασθενειών Τα τελευταία χρόνια, όλο και περισσότερες μολυσματικές ασθένειες που πλήττουν την υγεία των πολιτών εμφανίζονται σε παγκόσμιο επίπεδο [11]. Οι προαναφερθείσες ασθένειες, απειλούν και άλλους τομείς της ανθρωπότητας, όπως τις κοινωνικές συναναστροφές και την παγκόσμια οικονομία.

Όπως είναι αναμενόμενο, το ερευνητικό ενδιαφέρον έχει στραφεί στην ανάλυση της μετάδοσης μολυσματικών ασθενειών στο χώρο και το χρόνο. Ο μηχανισμός αυτός προς το παρόν δεν έχει αναλυθεί πλήρως, ωστόσο υπάρχουν ενθαρρυντικά αποτελέσματα από τη μελέτη της ανθρώπινης κινητικότητας. Πράγματι, ποικίλες πρόσφατες έρευνες [12-14] συμπεριλαμβάνουν μηχανισμούς υπολογισμού ανθρώπινης κινητικότητας στα επιδημιολογικά μοντέλα που παρουσιάζουν.

Οι έρευνες καταδεικνύουν άμεση συσχέτιση της μετακίνησης των ανθρώπων και της μετάδοσης ασθενειών, τόσο ως προς την ταχύτητα μετάδοσης όσο και ως προς τη διασπορά τους σε μια γεωγραφική περιοχή. Η κατανόηση του ρόλου της ανθρώπινης κινητικότητας είναι αρκετά χρήσιμη στη σχεδίαση αποτελεσματικών μοντέλων ανακοπής της διάδοσης των ασθενειών. Ένα πρόσφατο παράδειγμα των ανωτέρω, είναι η νόσος Covid-19. Με τη μελέτη της διάδοσης του ιού

χρησιμοποιώντας μοντέλα ανθρώπινης κινητικότητας, διαπιστώθηκε σημαντική μείωση στη διασπορά της νόσου με την εφαρμογή στρατηγικών περιορισμού μετακίνησης πολιτών [15].

Πολεοδομική σχεδίαση Προκειμένου να επιτευχθεί ένας αποτελεσματικός πολεοδομικός σχεδιασμός και να εξυπηρετηθούν οι ανάγκες των πολιτών, είναι ανάγκη να γίνει κατανοητή η χωρική και χρονική διασπορά των ανθρώπων σε μια πόλη. Κάθε αστικός ιστός διαθέτει διάφορες λειτουργικές περιοχές, όπως για παράδειγμα περιοχές εκπαίδευσης, βιομηχανικές, εμπορικές, αναψυχής, κατοικιών και άλλες. Η κάθε περιοχή εξυπηρετεί διαφορετικές λειτουργικές ανάγκες της πόλης.

Χρησιμοποιώντας μοντέλα ανθρώπινης κινητικότητας και ποικίλες πηγές δεδομένων, οι ερευνητές κατασκεύασαν μοντέλα αναγνώρισης λειτουργικών περιοχών πόλεων [16, 17]. Η σημασία ανάπτυξης τέτοιων τεχνικών είναι μεγάλη, αφού σε μία πόλη, τόσο η μελέτη όσο και η κατανόηση της διασποράς των ατόμων στο χώρο και το χρόνο είναι απαραίτητη για κάθε πολεοδομική επένδυση, προκειμένου να επιτευχθεί το βέλτιστο δυνατό αποτέλεσμα. Επιπλέον, η γνώση των λειτουργικών περιοχών μιας πόλης, εκτός απ' τον πολεοδομικό σχεδιασμό, είναι χρήσιμη στην αποτελεσματική αστική οργάνωση. Χαρακτηριστικό παράδειγμα αποτελεί ο σχεδιασμός του συγκοινωνιακού δικτύου.

Οι προαναφερθείσες λειτουργικές περιοχές μπορεί να είναι είτε σχεδιασμένες από αρχιτέκτονες είτε να έχουν διαμορφωθεί βάσει των αναγκών των πολιτών. Ωστόσο, επειδή οι πόλεις αναπτύσσονται και τροποποιούνται συνεχώς, είναι απαραίτητο να πραγματοποιείται αποτελεσματική και τακτική μελέτη της αστικής ανθρώπινης κινητικότητας.

Διαδίκτυο Ενώ εκ πρώτης όψεως η φυσική κίνηση του ανθρώπου και η πλοήγηση στο διαδίκτυο είναι φαινομενικά δυο αρκετά διαφορετικές δραστηριότητες, υπάρχουν θεωρίες που υποστηρίζουν τη συσχέτιση των δυο. Στην εποχή μας, πολλές καθημερινές δραστηριότητες έχουν άμεση εξάρτηση από το διαδίκτυο, όπως για παράδειγμα η τηλεργασία, η καθημερινή ενημέρωση, οι ηλεκτρονικές αγορές και η επικοινωνία. Άνθρωποι όλων των ηλικιακών ομάδων περνούν πλέον ένα μέρος της ημέρας τους στο διαδίκτυο, αφήνοντας ψηφιακά ίχνη τα οποία συλλέγονται από

διαφόρων ειδών trackers. Η πληθώρα των συλλεγόμενων δεδομένων και η ψηφιακή τους φύση δίνει την ευκαιρία σε ερευνητές να τα αναλύσουν και να εξάγουν συμπεράσματα βάσει αυτών.

Πρέπει ωστόσο να σημειωθεί πως συχνά τα συλλεγόμενα δεδομένα περιήγησης τείνουν να εκπροσωπούν και να μεγεθύνουν δραστηριότητες όπως η αναζήτηση πληροφοριών και η επικοινωνία. Πολλές φορές, οι χρήστες του διαδικτύου συνδέονται σε αυτό μόνο και μόνο για να αναζητήσουν κάτι που δε θυμούνται ή να μιλήσουν με άτομα που δε βρίσκονται κοντά. Αν λοιπόν ένας άνθρωπος εμφάνιζε μόνο τις προηγούμενες δυο συμπεριφορές στο διαδίκτυο, θα ήταν εσφαλμένο να θεωρήσουμε πως οι δραστηριότητες αυτές εκφράζουν και την ζωή του εκτός διαδικτύου. Παρόλα αυτά, οι έρευνες δείχνουν ομοιότητες μεταξύ της συμπεριφοράς των ανθρώπων όταν κινούνται φυσικά στο χώρο και των διαδικτυακών τους ενεργειών.

Μέχρι στιγμής, σε σύγκριση με τη φυσική κίνηση, η ψηφιακή δραστηριότητα είναι σχετικά πρόσφατη σε μια εξελικτική κλίμακα και πολλά απομένουν να αποκρυπτογραφηθούν. Είναι λογικό, ωστόσο, να υποτεθεί πως η διαδικτυακή δραστηριότητα καθοδηγείται από τον ίδιο ψυχολογικό μηχανισμό με την φυσική συμπεριφορά, προσαρμοσμένη στο νέο περιβάλλον. Τα τελευταία χρόνια η υπόθεση αυτή έτυχε όλο και περισσότερης επιστημονικής υποστήριξης, τόσο από θεωρητική όσο και από πειραματική άποψη. Ένα παράδειγμα ενίσχυσης της συσχέτισης φυσικής και ψηφιακής συμπεριφοράς προήλθε από την παρατήρηση συμπεριφορών αγοράς φαγητού σε διαδικτυακά καταστήματα. Εκεί, διαπιστώθηκε πως η συμπεριφορά και η πορεία της πλοήγησης των ανθρώπων στο ηλεκτρονικό κατάστημα έμοιαζε με την κίνηση των ανθρώπων σε ένα φυσικό μαγαζί.

2.3 Βασικές μονάδες για τη μέτρηση της ανθρώπινης κινητικότητας

Για τη μέτρηση της ανθρώπινης κινητικότητας και την ποσοτικοποίησή της, στη βιβλιογραφία αξιοποιούνται ορισμένες μονάδες μέτρησης. Μερικές από τις συχνότερα εμφανιζόμενες μονάδες μέτρησης της ανθρώπινης κινητικότητας είναι τα μήκη άλματος (jump lengths), η ακτίνα περιστροφής, ο αριθμός επισκεφθεισών

τοποθεσιών και η μέση τετραγωνική μετατόπιση. Ο τρόπος που υπολογίζονται καθώς και βασικές πληροφορίες για αυτές αναλύονται στην τρέχουσα ενότητα.

2.3.1 Μήκη άλματος (jump lengths)

Ένας βασικός παράγοντας για τη μοντελοποίηση της ανθρώπινης κινητικότητας είναι η απόσταση που διανύει ένα άτομο σε μια δεδομένη χρονική περίοδο. Τα μέτρα απόστασης συχνά εξαρτώνται από τις πηγές των χρησιμοποιούμενων δεδομένων. Για παράδειγμα, οι πρώτες μονάδες μέτρησης ανθρώπινης κινητικότητας αξιοποίησαν πληροφορίες σχετικά με τις χωρικές τροχιές των χαρτονομισμάτων που άλλαζαν χέρια κατά την κυκλοφορία τους στην αγορά.

Συμβατικά, το διάστημα μεταξύ δύο θέσεων εμφάνισης ενός τραπεζογραμματίου στα αναλυόμενα δεδομένα αρχικά θεωρήθηκε ως μισό μήκος (half length) [18]. Με την πάροδο των χρόνων και την άνθηση της τεχνολογίας, οι μετρήσεις ανθρώπινης κινητικότητας προέρχονταν από Call Detail Records (CDRs), τα οποία έδωσαν τη δυνατότητα σε πιο λεπτομερείς αναλύσεις σε σχέση με το παρελθόν.

Κάνοντας χρήση των καταγραφών λεπτομερειών κλήσεων (Call Detail Records, CDRs), οι ερευνητές μπόρεσαν να μετρήσουν με μεγαλύτερη αξιοπιστία τη θέση (και την αντίστοιχη μετατόπιση) των ατόμων με βάση την τοποθεσία των πύργων κινητής τηλεφωνίας που ενεργοποιούνταν κάθε φορά που ένα άτομο πραγματοποιούσε μια κλήση ή έστειλε ένα μήνυμα. Το χρονικό διάστημα μεταξύ δυο κλήσεων μπορούσε επίσης να βοηθήσει στην εκτίμηση του χρόνου παραμονής ενός ατόμου σε μια θέση ή της μετακίνησής του στο χρόνο αυτό. Στο σημείο αυτό να σημειωθεί πως μια τέτοια εκτίμηση δεν είναι πάντοτε ακριβής και αξιόπιστη, καθώς η θέση των ανθρώπων στο διάστημα μεταξύ της έναρξης και λήξης της κλήσης δεν καταχωρείται και συνεπώς δεν υπάρχουν επαρκή στοιχεία για κάποιον υπολογισμό.

Με την πάροδο των ετών, η ανάπτυξη της τεχνολογίας προσέφερε ακόμη πιο αναλυτικά δεδομένα. Τα δεδομένα αυξήθηκαν τόσο ως προς τον όγκο, αλλά και ως προς την καταγραφείσα πληροφορία σε συνδυασμό με την ανάπτυξη τεχνικών ανάλυσής τους. Τα δεδομένα γεωεντοπισμού (Global Positioning System (GPS)) έδωσαν στους ερευνητές τη δυνατότητα υπολογισμού της ανθρώπινης κινητικότητας με μεγαλύτερη ακρίβεια σε σχέση με τις καταγραφές λεπτομερειών κλήσεων. Η αύξηση της ακρίβειας και η μείωση των σφαλμάτων υπολογισμού θέσης πηγάζει από

το γεγονός πως οι συσκευές γεωεντοπισμού καταγράφουν άμεσα τη θέση του χρήστη, ενώ οι καταγραφές λεπτομερειών κλήσεων καταγράφουν τη σχετική θέση του χρήστη ως προς τον κοντινότερο - κατά κύριο λόγο - πύργο τηλεπικοινωνίας.

Ανεξάρτητα από την διατύπωση και τους περιορισμούς αυτής, το μήκος άλματος, που συνήθως συμβολίζεται ως Δr , ορίζεται ως την ευκλείδεια απόσταση μεταξύ του $r(t)$ και $r(t + dt)$. Τα $r(t)$ και $r(t+dt)$ αντιστοιχούν σε καταγραφείσες τοποθεσίες μεταξύ διαστημάτων t και $t + dt$. Ιδιαίτερο ενδιαφέρον παρουσιάζει η κατανομή του Δr σε έναν πληθυσμό, δηλαδή ο υπολογισμός της πιθανότητας ενός τυχαίου μέλους του πληθυσμού να ταξιδέψει σε μια απόσταση r από την τοποθεσία προέλευσής του σε ένα χρονικό διάστημα dt .

Κατά τη μοντελοποίηση της ανθρώπινης κινητικότητας, γίνεται συχνά χρήση της συνάρτησης κατανομής πιθανότητας (Probability Distribution Function, PDF) του μήκους άλματος, δηλαδή η $P(\Delta r)$. Με άλλα λόγια, η $P(\Delta r)$ μπορεί να διατυπωθεί ως η πιθανότητα καταγραφής μιας μετατόπισης Δr σε βήμα dt .

2.3.2 Ακτίνα περιστροφής (RoG)

Η υπο-διάχυτη συμπεριφορά των ανθρώπων σε συγκεκριμένες κλίμακες υποδηλώνει ότι τείνουν να ξεφεύγουν από μια χαρακτηριστική απόσταση μακριά από τις αρχικές τους θέσεις. Αυτή η απόσταση μπορεί να μετρηθεί με τη λεγόμενη ακτίνα περιστροφής (Radius of Gyration, RoG), r_g , που ορίζεται ως η μέση τετραγωνική ρίζα (RMS, root mean square) της απόστασης ενός συνόλου σημείων από έναν άξονα. Η ακτίνα περιστροφής αντλεί το θεωρητικό της υπόβαθρο από τις επιστήμες της φυσικής και της μηχανικής.

Μια εξίσωση υπολογισμού της ακτίνας περιστροφής που αξιοποιείται συχνά στη βιβλιογραφία είναι η ακόλουθη:

$$r_g = \sqrt{\frac{1}{n} \sum_{i=1}^N (r_i - r_0)^2},$$

όπου r_i οι συντεταγμένες των μεμονωμένων σημείων (ή μετρήσεις της θέσης) και r_0 το διάνυσμα θέσης του κέντρου μάζας του συνόλου σημείων, ενώ το r_{cm} ορίζεται ως $r_{cm} = \sum_{i=1}^N \frac{r_i}{N}$. Στα πλαίσια της ανθρώπινης κινητικότητας, η ακτίνα περιστροφής δύναται να χαρακτηρίσει την τυπική απόσταση ενός ατόμου από το κέντρο μάζας της

τροχιάς του. Τέλος, η ακτίνα περιστροφής δεν ταυτίζεται με τη μέση τετραγωνική ρίζα μετατόπισης, καθώς η τελευταία αναφέρεται σε μια χρονική στιγμή προέλευσης.

2.3.3 K-οστή ακτίνα περιστροφής (k-RoG)

Η ακτίνα περιστροφής (Radius of Gyration, RoG) χρησιμοποιείται στο χαρακτηρισμό της τυπικής απόστασης που ταξιδεύει ένα άτομο και εξαρτάται τόσο από την αμοιβαία απόσταση των τοποθεσιών που επισκέπτεται όσο και από τον χρόνο που αφιερώνει (ή τον συνολικό αριθμό επισκέψεων) σε κάθε τοποθεσία. Ωστόσο, η ακτίνα της περιστροφής δεν λαμβάνει υπόψιν της τη συχνότητα επίσκεψης ενός ανθρώπου σε κάθε τοποθεσία στον καθορισμό της χαρακτηριστικής του κινητικότητας.

Ένα άτομο που δαπανά την πλειοψηφία του χρόνου του σε τοποθεσίες μεγάλης επισκεψιμότητας, όπως λόγου χάρη το σπίτι και το χώρο εργασίας του, τότε θα έχει μεγάλη ακτίνα περιστροφής r_g εάν αυτές οι δύο τοποθεσίες τυχαίνει να απέχουν πολύ μεταξύ τους. Αντίθετως, ακόμη και αν οι τοποθεσίες με τις περισσότερες επισκέψεις είναι κοντά η μια στην άλλη, ενδέχεται να διαπιστωθούν μεγάλες αποκλίσεις στην ακτίνα περιστροφής r_g εάν το άτομο επισκέπτεται πολλές απομακρυσμένες τοποθεσίες.

Για να περιοριστεί η επιρροή των φαινομένων αυτών, υπάρχει δυνατότητα να μελετηθεί η επίδραση της συχνότητας των επισκεφθεισών τοποθεσιών στη χαρακτηριστική απόσταση που διανύθηκε. Αυτό μπορεί να συμβεί από τη μελέτη της κ-οστής ακτίνας περιστροφής (k-Radius of Gyration, k-RoG), που ορίζεται ως την ακτίνα περιστροφής των κ-τοποθεσιών που επισκέπτεται συχνότερα ένα άτομο [19]. Η τροποποιημένη αυτή μορφή της ακτίνας περιστροφής μπορεί να διατυπωθεί ως εξής:

$$r_g^{(k)} = \sqrt{\frac{1}{N_k} \sum_{j=1}^k n_j (r_j - r_{cm}^{(k)})^2}$$

όπου N_k το άθροισμα των συνολικών επισκέψεων στις κ πιο συχνές τοποθεσίες, $r_{cm}^{(k)}$ το κέντρο μάζας των τοποθεσιών αυτών και n_j ο αριθμός των επισκέψεων στην j-οστή συχνότερη επισκεφθείσα τοποθεσία.

2.3.4 Αριθμός επισκεφθεισών τοποθεσιών

Η συνιστώσα αυτή προκύπτει από το γεγονός ότι οι άνθρωποι τείνουν να επισκέπτονται ορισμένες περιοχές πιο συχνά σε σχέση με άλλες. Στις κινήσεις μίας ημέρας ή μιας εβδομάδας, οι τοποθεσίες μπορούν να διακριθούν ως προς τη σημασία τους και την συχνότητα ή περιοδικότητα που το άτομο τις επισκέπτεται.

Οι άνθρωποι έχουν την τάση να επιστρέφουν στο σπίτι τους σε καθημερινή βάση και επομένως οι περισσότερες καθημερινές και εβδομαδιαίες τροχιές συνήθως ξεκινούν και τελειώνουν στην ίδια τοποθεσία. Επίσης, κάποιες τοποθεσίες, όπως το σπίτι ενός ανθρώπου, ο χώρος εργασίας ή τα προτιμώμενα εστιατόρια λαμβάνουν πολύ περισσότερες επισκέψεις σε σχέση με άλλα μέρη που επισκέπτεται σε ένα σχετικά εύλογο διάστημα παρατήρησης.

Μια μέθοδος διάκρισης της σημασίας μιας τοποθεσίας είναι η κατάταξή της ως προς τη βαρύτητά της. Για παράδειγμα, η τοποθεσία με τις περισσότερες επισκέψεις θα έχει την κατάταξη 1, το σχολείο, ο χώρος εργασίας ένα αγαπημένο κατάστημα μπορεί να έχει την κατάταξη 2 ή 3. Σε μια έρευνα [20], διαπιστώθηκε ότι η συχνότητα επίσκεψης ακολουθεί το νόμο του Zipf, σύμφωνα με τον οποίο η πιθανότητα εύρεσης ενός χρήστη σε μια θέση είναι $P(L) \sim \frac{1}{L}$.

Μια άλλη μέθοδος διάκρισης μεταξύ τοποθεσιών είναι η κατασκευή του μοτίβου κινητικότητας κάθε ατόμου σε ένα δίκτυο. Σε μια έρευνα [21], οι Schneider et al. ανέλυσαν δίκτυα καθημερινών ταξιδιών χρησιμοποιώντας δεδομένα δύο πόλεων που προήλθαν από έρευνες, καθώς και ανώνυμα δεδομένα περιήγησης κινητής τηλεφωνίας. Στη συνέχεια, κατασκεύασαν δίκτυα καθημερινής κινητικότητας για κάθε άτομο που συμμετείχε στην έρευνα. Τα δίκτυα αυτά περιορίστηκαν μόνο στις εργάσιμες ημέρες και το εύρος τους ήταν μια ημέρα. Κάθε δίκτυο μετακινήσεων ξεκινούσε και τελείωνε στο ίδιο σημείο, το οποίο ήταν η οικία του εκάστοτε ατόμου. Ως "οικία", καθορίστηκε το σημείο στο οποίο οι χρήστες βρίσκονταν συχνότερα μεταξύ 03:00 και 03:30 το πρωί.

Οι ερευνητές συμπέραναν πως περίπου το 90% των καταγεγραμμένων ταξιδιών των χρηστών μπορεί να περιγραφεί με μόνο 17 καθημερινά δίκτυα. Τα καθημερινά δίκτυα τα οποία εμφανίζονταν με συχνότητα άνω του 0.5% ονομάστηκαν μοτίβα, ενώ το θεωρητικό υπόβαθρο των μοτίβων αντλήθηκε από τη θεωρία δικτύων.

Το γεγονός λοιπόν ότι τα καθημερινά μοτίβα της ανθρώπινης κινητικότητας

μπορούσαν να κατασκευαστούν από μόλις 17 δίκτυα ταξιδιού, υποδηλώνει ότι αυτά τα μοτίβα αντιπροσωπεύουν τους υποσυνείδητους κανόνες και τις ρουτίνες που καθορίζουν τις καθημερινές κινήσεις των ανθρώπων. Φαίνεται λοιπόν πως η ανάλυση των μοτίβων αυτών, η ανεύρεση νέων μοτίβων και ο εντοπισμός των κανόνων που διέπουν την ανθρώπινη μετακίνηση μπορεί να συνδράμει στην ακριβή μοντελοποίηση της προσομοίωσης της κίνησης των ανθρώπων.

2.3.5 Μέση τετραγωνική μετατόπιση

Στη στατιστική μηχανική, η μέση τετραγωνική μετατόπιση (Mean Squared Displacement, MSD, επίσης μέση τετραγωνική διακύμανση) είναι η μέτρηση της απόκλισης της θέσης ενός σωματιδίου ως προς κάποια θέση αναφοράς με την πάροδο του χρόνου. Χρησιμοποιείται κατά κύριο λόγο ως μέτρηση της χωρικής έκτασης της τυχαίας κίνησης και μπορεί να αναπαραστήσει την κίνηση ενός ανθρώπου που κινείται τυχαία στο χώρο.

Στη βιβλιογραφία, η μέση τετραγωνική μετατόπιση ορίζεται ως εξής:

$$MSD(t) = \langle \langle (r_t - r_0)^2 \rangle \rangle \equiv \langle \Delta r(t)^2 \rangle$$

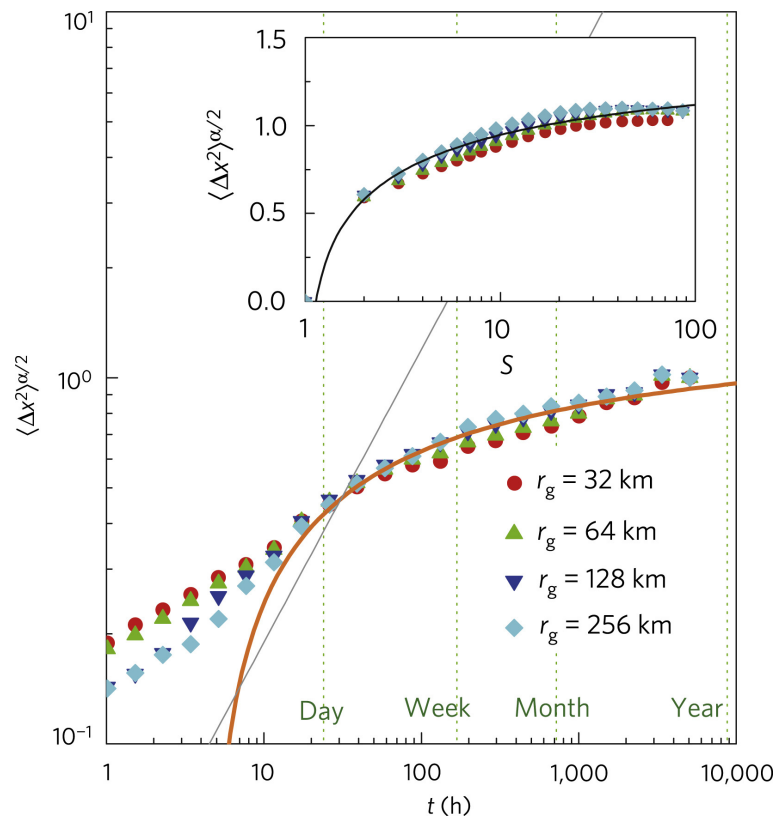
όπου r_0 είναι ένα διάνυσμα της προέλευσης του ατόμου ως προς κάποιο σημείο αναφοράς, ή με άλλα λόγια, η θέση από την οποία ξεκινά μια συγκεκριμένη τροχιά. Το $r(t)$ υπολογίζει την επόμενη θέση του ατόμου σε συνάρτηση με το χρόνο. Η κλιμάκωση της μέσης τετραγωνικής μετατόπισης αποτελεί μια μέτρηση του τύπου διασποράς των ατόμων σε σχέση με την αφετηρία τους σε ένα ταξίδι. Επιπλέον, η μέση τετραγωνική μετατόπιση (MSD) αντιστοιχεί στη μέση τετραγωνική απόσταση από την τοποθεσία προέλευσης μετά από ένα χρονικό διάστημα t . Γενικότερα, εάν οι ατομικές τροχιές κινητικότητας ακολουθούν τον λεγόμενο Συνεχούς Χρόνου Τυχαίο Περίπατο (Continuous Time Random Walk, CTRW), η μέση τετραγωνική μετατόπιση λαμβάνει την ακόλουθη μορφή:

$$\langle \Delta r(t) \rangle \sim t^\nu,$$

για το οποίο ισχύει $\nu = \frac{2\alpha}{\beta}$, όπου α και β οι εκθέτες του χρόνου αναμονής, δηλαδή το χρονικό διάστημα μεταξύ δύο διαδοχικών αλμάτων.

Πιο συγκεκριμένα, για τυχαίο περίπατο (κανονική διάχυση) θέτουμε στην εξίσωση της Μέσης Τετραγωνικής Μετατόπισης όπου $v=1$. Πρέπει να σημειωθεί πως ο Συνεχούς Χρόνου Τυχαίος Περίπατος (CTRW) δεν είναι μια απολύτως ρεαλιστική αναπαράσταση του τρόπου με τον οποίο πραγματικά κινούνται οι άνθρωποι. Ειδικότερα, η τυχαία κίνηση του ανθρώπου στο χώρο τείνει να απομακρύνεται από την ρίζα της τροχιάς του σχετικά γρήγορα. Η απομάκρυνση αυτή διευρύνεται, όσο αυξάνεται το διάστημα που παρέρχεται μεταξύ των κινήσεων. Όπως είναι αναμενόμενο, οι εμπειρικές μετρήσεις δείχνουν ότι οι άνθρωποι έχουν την τάση - κατά μέσο όρο - να επιστρέφουν στο σπίτι τους σε καθημερινή βάση. Επιπλέον, τα μέσα μετακίνησης περιορίζουν το μέγιστο μήκος άλματος που μπορεί να πραγματοποιηθεί. Έχει λοιπόν παρατηρηθεί πως οι συναρτήσεις πυκνότητας πιθανότητας έχουν εκθετικές αποκοπές. Με απλά λόγια, αυτό σημαίνει πως οι διακυμάνσεις τους είναι πεπερασμένες. Ως εκ τούτου, στο ασυμπτωτικό όριο ($t \rightarrow \infty$), η κλιμάκωση της μέσης τετραγωνικής μετατόπισης θα συγκλίνει ασυμπτωτικά με αυτήν της κίνησης Μπράουν, η οποία θα αναλυθεί στο επόμενο κεφάλαιο.

Η ασυμφωνία της πραγματικής κίνησης στο χώρο και της προσομοίωσής της μέσω του Συνεχούς Χρόνου Τυχαίου Περιπάτου, ελήφθη υπόψη στην έρευνα των Song et al. [1]. Οι ερευνητές ανέλυσαν τόσο πυκνά όσο και αραιά δεδομένα τοποθεσίας. Αυτό σημαίνει πως συνδύασαν τόσο εγγραφές κινητών τηλεφώνων που περιείχαν ένα στίγμα τοποθεσίας κάθε φορά που πραγματοποιούνταν μια κλήση, όσο και δεδομένα τοποθεσίας που καταγράφονταν από τους παρόχους κινητής τηλεφωνίας ανά διαστήματα μιας ώρας. Από τις μετρήσεις τους κατέληξαν στο συμπέρασμα πως η ανθρώπινη κινητικότητα ακολουθεί μια εξαιρετικά αργή διαδικασία διάχυσης, την οποία μάλιστα χαρακτήρισαν βραδύτερη της λογαριθμικής. Η παρατήρηση αυτή φαίνεται και στο σχήμα 2.1. Η γκριζα γραμμή του σχήματος αντιπροσωπεύει την αναλυτική πρόβλεψη του Συνεχούς Χρόνου Τυχαίου Περιπάτου (CTRW, Continuous Time Random Walk). Η πορτοκαλί γραμμή με τη σειρά της, αντιπροσωπεύει την αναλυτική πρόβλεψη της ασυμπτωτικής συμπεριφοράς $MSD(t) \sim (\log \log(t))^{\frac{2}{\beta}}$, η οποία αντικατοπτρίζει με μεγαλύτερη ακρίβεια το μοτίβο κίνησης των ανθρώπων.



Σχήμα 2.1: Μέση τετραγωνική μετατόπιση έναντι χρόνου, για ομάδες ατόμων με διαφορετικές ακτίνες περιστροφής [1].

Ένα κοινό πρόβλημα που προκύπτει κατά τον υπολογισμό της μέσης τετραγωνικής μετατόπισης, είναι ότι η τοποθεσία προέλευσης ενός ατόμου είναι συχνά ασαφής. Μια μέθοδος για να ξεπεραστεί αυτό το ζήτημα είναι να ληφθεί ο μέσος όρος των τιμών της μέσης τετραγωνικής μετατόπισης. Ο μέσος όρος υπολογίζεται μεταβάλλοντας την προέλευση των τοποθεσιών που επισκέπτεται ένα άτομο. Μια τέτοια προσέγγιση υιοθετήθηκε στο [22], όπου αναλύθηκαν τα γεωγραφικά ψηφιακά αποτυπώματα ατόμων σε διάφορες τοποθεσίες. Οι τοποθεσίες αυτές περιελάμβαναν μεταξύ άλλων μια πανεπιστημιούπολη της Νέας Υόρκης, τη Ντίσνεϋ Γουόρντ και μια έκθεση στην Αμερική. Μέσω της διαδικασίας ανάλυσης της μέσης τετραγωνικής μετατόπισης αποκαλύφθηκαν δύο τύποι χρονικής συμπεριφοράς. Έως ένα συγκεκριμένο χρονικό διάστημα παρατήρησης, οι συμμετέχοντες μετακινούνταν υπερ-διάχυτα (super-diffusively), όπως αναμενόταν σε μια διαδικασία Συνεχούς Χρόνου Τυχαίου Περιπάτου. Για περιόδους μεγαλύτερες του λεπτού, τα άτομα μετακινούνταν υπο-διάχυτα (sub-diffusively).

Σε περιπτώσεις όπως αυτή που περιγράφηκε παραπάνω, η έννοια της τάξης

μεγέθους καθίσταται σημαντική. Οι συγγραφείς προσπάθησαν να αιτιολογήσουν τη συγκεκριμένη συμπεριφορά υποστηρίζοντας πως σε μικρές χρονικές κλίμακες η κατανομή του μήκους άλματος (jump length) ακολουθεί μια κατανομή ισχύος. Σε μεγαλύτερες κλίμακες, τα μήκη άλματος είναι πιο ομοιογενή και ακολουθούν μια κατανομή Γκάους.

2.4 Ανακάλυψη γνώσης σε βάσεις δεδομένων

Σε πολλά επιστημονικά πεδία, τα δεδομένα που συλλέγονται, συσσωρεύονται με διαρκώς αυξανόμενους ρυθμούς. Τα δεδομένα αυτά έχουν διάφορες μορφές, από επιχειρηματικές συναλλαγές, δεδομένα αισθητήρων και πειραμάτων μέχρι οπτικοακουστικό υλικό. Υπάρχει λοιπόν επείγουσα ανάγκη ανάπτυξης συστημάτων, θεωριών και εργαλείων που βοηθούν τους ανθρώπους στην εξαγωγή χρήσιμων πληροφοριών (γνώσης, Knowledge) από τον όγκο των διαθέσιμων δεδομένων.

Οι θεωρίες και τα εργαλεία αυτά αποτελούν το αντικείμενο της ανακάλυψης γνώσης σε βάσεις δεδομένων (Data Mining, Knowledge Discovery in Databases). Επιπλέον, αναφέρονται στην εξαγωγή έμμεσων, προηγουμένως άγνωστων και ενδεχομένως χρήσιμων πληροφοριών από δεδομένα ήδη αποθηκευμένα σε βάσεις δεδομένων.

Σύμφωνα με την έρευνα [2] των Fayyad et al., η διαδικασία ανακάλυψης πληροφορίας βασίζεται σε μεθόδους εύρεσης δεδομένων ή μηχανικής μάθησης. Μάλιστα, ακολουθεί τα παρακάτω βήματα:

Βήματα διαδικασίας ανακάλυψης πληροφορίας

1. Κατανόηση του

- I Περιβάλλοντος εφαρμογής
- II Σχετικής προηγούμενης γνώσης
- III Στόχων για τον τελικό χρήστη

2. Δημιουργία ενός επιθυμητού συνόλου δεδομένων - στόχου: Επιλογή ενός συνόλου δεδομένων ή εστίαση σε ένα υποσύνολο μεταβλητών ή δειγμάτων

δεδομένων, στα οποία θα πραγματοποιηθεί η διαδικασία ανακάλυψης πληροφοριών (Creating a target data set)

3. Οργάνωση και προεπεξεργασία δεδομένων (Data Cleaning and Preprocessing).

I Αφαίρεση θορύβου ή ακραίων τιμών

II Συλλογή απαραίτητης πληροφορίας για τη μοντελοποίηση ή συνυπολογισμός του θορύβου

III Κατάστρωση στρατηγικών διαχείρισης ελλιπών δεδομένων

IV Συνυπολογισμός πληροφοριών αλληλουχίας χρόνου και γνωστών αλλαγών.

4. Περιορισμός δεδομένων και προβολή.

I Εύρεση χρήσιμων λειτουργιών για την αναπαράσταση των δεδομένων ανάλογα με τον στόχο της εργασίας.

II Χρήση μεθόδων μείωσης διαστάσεων ή μετασχηματισμού είτε για τη μείωση του πραγματικού αριθμού των υπό εξέταση μεταβλητών είτε για την εύρεση αμετάβλητων αναπαραστάσεων για τα δεδομένα.

5. Επιλογή της εργασίας εξόρυξης δεδομένων. (data mining task)

I Επιλογή στόχου της διαδικασίας ανακάλυψης γνώσης σε βάσεις δεδομένων, όπως για παράδειγμα ταξινόμηση, regression, ομαδοποίηση κ.α.

6. Επιλογή αλγορίθμου (ή αλγορίθμων) εξόρυξης δεδομένων. (data mining algorithms)

I Επιλογή μεθόδων που θα χρησιμοποιηθούν για την αναζήτηση μοτίβων στα δεδομένα.

II Απόφαση για το ποια μοντέλα και ποιες παράμετροι μπορούν να είναι κατάλληλα στην εκάστοτε περίπτωση.

III Αντιστοίχιση μιας συγκεκριμένης μεθόδου εξόρυξης δεδομένων με τα συνολικά κριτήρια της διαδικασίας ανακάλυψης αυτών.

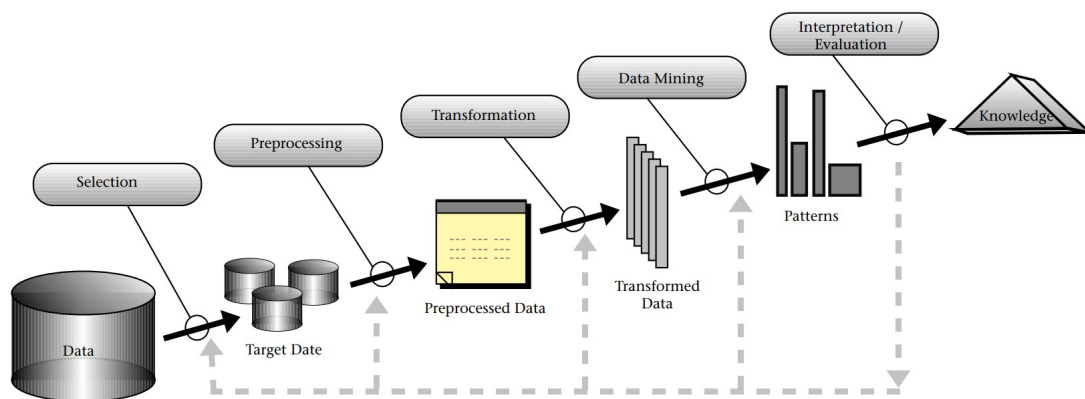
7. Εξόρυξη δεδομένων.

Η Αναζήτηση μοτίβων ενδιαφέροντος σε μια συγκεκριμένη αντιπροσωπευτική μορφή ή ένα σύνολο παραστάσεων όπως κανόνες ταξινόμησης, δέντρα, παλινδρόμηση, ομαδοποίηση και άλλα.

8. Ερμηνεία αποτελεσμάτων εξόρυξης μοτίβων.

9. Εφαρμογή αποκτηθείσας γνώσης και αναζήτηση σχετικών ενεργειών βασισμένων σε αυτή.

Τα παραπάνω βήματα αναπαριστώνται και σχηματικά στο σχήμα 2.2.



Σχήμα 2.2: Τα βήματα που συνθέτουν μια διαδικασία ανακάλυψης πληροφορίας (KDD, Knowledge Data Discovery) [2].

2.5 Προστασία της ιδιωτικότητας

Η ιδιωτικότητα είναι το δικαίωμα ενός ατόμου ή μιας ομάδας, να αποφασίζουν από μόνοι τους για το πότε, το πως και μέχρι ποιο σημείο οι πληροφορίες που αφορούν τους ίδιους, θα διαβιβάζονται σε άλλους. Αφορά λοιπόν την προστασία της προσωπικότητας και του ιδιωτικού βίου. Η ιδιωτικότητα αποτελεί ένα αναγνωρισμένο θεμελιώδες ανθρώπινο δικαίωμα.

Η ιδιωτικότητα αντλεί τις ρίζες της από την αρχαία φιλοσοφία. Ο Αριστοτέλης διέκρινε δυο σφαίρες της ανθρώπινης ύπαρξης: την δημόσια- που σχετιζόταν με την πολιτική ζωή- και την ιδιωτική.

Παρά τη σημασία της, η προστασία της ιδιωτικότητας στηρίζεται σε άτυπους μηχανισμούς. Για παράδειγμα, για ένα μεγάλο χρονικό διάστημα η ατομική μετακίνηση ήταν αδύνατο να εντοπιστεί και να καταγραφεί σε μεγάλη κλίμακα. Ωστόσο, με την πρόοδο της τεχνολογίας, και ειδικότερα με την ανάπτυξη του διαδικτύου και της τηλεφωνίας, τίθεται υπό αμφισβήτηση η θεμελιώδης έννοια της ιδιωτικότητας.

Μια έρευνα [23] μελέτησε για δεκαπέντε μήνες δεδομένα ανθρώπινης κινητικότητας ενάμισι εκατομμυρίου ανθρώπων και διαπίστωσε πως η κίνηση των ατόμων από μόνη της είναι αρκετή για να τους ξεχωρίσει και να τους ταυτοποιήσει. Στην πραγματικότητα, σε ένα σύνολο δεδομένων όπου η θέση ενός ατόμου καθορίζεται ανά τακτά χρονικά διαστήματα και μπορεί να προσδιοριστεί χωρικά με κεραίες τηλεπικοινωνίας, μόλις τέσσερα χωροχρονικά σημεία είναι αρκετά για να εξακριβωθεί η ταυτότητα του με 95% βεβαιότητα.

Ως εκ τούτου, είναι φανερό πως ακόμη και τα γενικευμένα σύνολα δεδομένων παρέχουν μικρή ανωνυμία στους ανθρώπους των οποίων τα στοιχεία έχουν συλλεχθεί. Διαπιστώνεται λοιπόν πως πρόσβαση σε αυτές τις πληροφορίες ενδέχεται να σημαίνει και περιορισμό της ιδιωτικότητας και της ατομικής ελευθερίας. Σε μεγάλη κλίμακα, η προστασία των δεδομένων ανθρώπινης κινητικότητας οφείλει να αποτελέσει προτεραιότητα στο σχεδιασμό πλαισίων και ρυθμιστικών κανονισμών.

2.6 Πηγές δεδομένων για την ανάλυση της ανθρώπινης κινητικότητας

Στην ενότητα αυτή αναλύονται οι κυριότερες πηγές δεδομένων για την ανάλυση και πρόβλεψη της ανθρώπινης κινητικότητας. Παρουσιάζεται επίσης η αξιοποίηση των πηγών αυτών στη βιβλιογραφία και γίνεται αναφορά στο ψηφιακό αποτύπωμα των ανθρώπων των οποίων τα στοιχεία μελετώνται, συμπληρώνοντας τα ηθικά ζητήματα που αναλύθηκαν στην προηγούμενη ενότητα. Στο τέλος της ενότητας, παρουσιάζεται μια συγκριτική ανασκόπηση των πηγών δεδομένων και των

ιδιαίτερών τους χαρακτηριστικών.

2.6.1 Ανάλυση πηγών δεδομένων

Για την ανάλυση της ανθρώπινης κίνησης και την πρόβλεψη των μελλοντικών της καταστάσεων, οι συχνότερες πηγές δεδομένων στη βιβλιογραφία είναι οι επιτόπιες έρευνες, οι λεπτομέρειες καταγραφής κλήσεων (CDRs), τα δεδομένα WiFi, τα δεδομένα GiS και τα δεδομένα που εξάγονται από τα κοινωνικά δίκτυα.

Επιτόπιες έρευνες Οι επιτόπιες έρευνες συνήθως πραγματοποιούνται με ερωτηματολόγια. Οι έρευνες αυτές αντλούν δεδομένα και πληροφορίες όπως το μέσο μετακίνησης, ο λόγος μετακίνησης, η θέση προέλευσης και προορισμού και άλλα. Οι επιτόπιες έρευνες μπορούν να συλλέξουν πληροφορίες που είναι δύσκολο να συλλεχθούν από άλλες πηγές δεδομένων, ωστόσο έχουν εγγενείς δυσκολίες και μειονεκτήματα. Το κύριο μειονέκτημα τους είναι το κόστος και η αδυναμία γρήγορης συλλογής πολλών πληροφοριών για ένα ικανό στατιστικό δείγμα πληθυσμού. Ως εκ τούτου, οι επιτόπιες έρευνες συχνά έχουν πολύ μικρότερο αριθμό ερωτηθέντων σε σύγκριση με μια απογραφή. Επιπλέον, κατά κύριο λόγο καλύπτουν μικρότερες περιοχές, όπως μια μικρή πόλη ή συνοικίες αυτής, και όχι μια ολόκληρη περιφέρεια ή ένα κράτος.

Οι επιτόπιες έρευνες πλεονεκτούν στην ανάλυση της ανθρώπινης κινητικότητας έναντι των απογραφών, καθώς εκτός των δημογραφικών στοιχείων συμπεριλαμβάνουν και πληροφορίες όπως λόγου χάρη ο σκοπός μιας μετακίνησης και το μεταφορικό μέσο που χρησιμοποιήθηκε για τη μετακίνηση αυτή.

Παράλληλα, τα δεδομένα απογραφής παρέχουν μόνο γενικά στοιχεία μετανάστευσης πληθυσμού. Η απογραφή δε, αποκαλύπτει λίγες πληροφορίες σχετικά με το πού οι άνθρωποι περνούν το χρόνο τους, το πού και το πώς αυτοί ταξιδεύουν. Συμπεραίνουμε λοιπόν πως και οι δύο αυτές πηγές δεδομένων δεν έχουν την ικανότητα να παρέχουν μια δυναμική εικόνα της ανθρώπινης κινητικότητας [24].

Λεπτομέρειες καταγραφής κλήσεων (CDRs) Οι λεπτομέρειες καταγραφής κλήσεων (Call Detail Record (CDR)) είναι καταγραφές δεδομένων που παράγονται από μια τηλεφωνική ανταλλαγή και καταγράφουν τις λεπτομέρειες της τηλεπικοινωνιακής

συναλλαγής που περνά μέσω μιας συσκευής. Οι λεπτομέρειες καταγραφής κλήσεων περιέχουν διάφορες πληροφορίες που μπορούν να ταυτοποιήσουν μια κλήση, χωρίς ωστόσο να καταγράφεται το περιεχόμενό αυτής. Τα δεδομένα που δύναται να συλλεχθούν ποικίλλουν. Μερικά από αυτά είναι:

- Πληροφορίες για τους συμμετέχοντες στην κλήση (προέλευση – προορισμός)
- Ημερομηνία και ώρα έναρξης της κλήσης
- Διάρκεια κλήσης
- Λογαριασμός χρέωσης
- Αναγνωριστικά id των εξοπλισμών που μεσολαβούν και διεκπεραιώνουν την κλήση.
- Την κατάσταση της κλήσης (επιτυχής, αποτυχημένη κ.α.)
- Τύπος της κλήσης (SMS, Voice, MMS κ.α.)
- Σφάλματα που προέκυψαν κατά τη διάρκεια της κλήσης

Πέραν της αξίας τους στην ανάλυση της ανθρώπινης κινητικότητας και των μοντέλων συμπεριφοράς, οι λεπτομέρειες καταγραφής κλήσεων έχουν κομβική σημασία για τους τηλεπικοινωνιακούς παρόχους, καθώς βάσει αυτών χρεώνουν τις υπηρεσίες τους στους καταναλωτές. Επίσης, βοηθούν το έργο της δικαιοσύνης, αφού διευκολύνουν τον εντοπισμό των υπόπτων και των διασυνδέσεών τους.

Δεδομένα WiFi Οι καταγραφές των σημείων πρόσβασης WiFi (WiFi Access Points) έχουν χρησιμοποιηθεί επίσης σε πολλές έρευνες για την ανάλυση της ανθρώπινης κινητικότητας. Οι καταγραφές αυτές περιέχουν διάφορες πληροφορίες, όπως το μοναδικό αναγνωριστικό της συσκευής (Unique ID, διεύθυνση MAC), την ώρα σύνδεσης ενός χρήστη στο δίκτυο WiFi και τη διάρκεια της συνεδρίας. Επίσης, μπορεί να περιλαμβάνουν πληροφορίες όπως την τοποθεσία του χρήστη, την ισχύ του σήματος ανά το χρόνο κ.α. Ένα πλεονέκτημα των καταγραφών WiFi σε σχέση με άλλες καταγραφές, είναι πως μπορούν να συμπληρώσουν δεδομένα που υπό άλλες συνθήκες δεν θα μπορούσαν να καταγραφούν. Για παράδειγμα, σε εσωτερικούς ή

υπόγειους χώρους συχνά το σήμα τηλεπικοινωνίας είναι ασθενές, με αποτέλεσμα οι κινητές συσκευές να μην μπορούν να πραγματοποιήσουν κλήσεις ή να αποστείλουν γραπτά μηνύματα. Εκεί λοιπόν το κενό καταγραφής δύναται να συμπληρωθεί από δεδομένα καταγραφής WiFi, μιας και συνήθως υπάρχουν διαθέσιμα σημεία πρόσβασης σε ιδιωτικούς και δημόσιους χώρους.

Συστήματα γεωγραφικών πληροφοριών (GIS) Τα συστήματα γεωγραφικών πληροφοριών (Geographic Information Systems) είναι εννοιολογικά πλαίσια που δίνουν την δυνατότητα συλλογής και ανάλυσης χωρικών και γεωγραφικών δεδομένων.

Τα δεδομένα GIS είναι χρήσιμα σε πολλές τεχνολογίες και σε αυτά βασίζονται οι υπηρεσίες γεωεντοπισμού. Επίσης, είναι πλέον άρρηκτα συνδεδεμένα και με άλλες εφαρμογές και επιστήμες, όπως τη μηχανική, τα logistics και τις τηλεπικοινωνίες.

Τα συστήματα γεωγραφικών πληροφοριών (GIS) παρέχουν τη δυνατότητα συσχέτισης προηγουμένως ασύνδετων πληροφοριών. Κάθε τοποθεσία και έκταση γης, μπορεί να μεταφραστεί στο σύστημα GIS. Τα δεδομένα αυτά περιλαμβάνουν την ημερομηνία και ώρα καταγραφής, τις συντεταγμένες μήκους (x), γεωγραφικού πλάτους (y) και υψομέτρου (z).

Ανάλυση δεδομένων κοινωνικών δικτύων Η ανάλυση των δεδομένων κοινωνικών δικτύων (social media analytics) είναι η διαδικασία συλλογής και ανάλυσης δεδομένων από τα κοινωνικά δίκτυα. Συχνά στη βιβλιογραφία, πηγές τέτοιων δεδομένων είναι μεγάλες πλατφόρμες κοινωνικών δικτύων, όπως το Facebook, το Instagram, το Twitter και το LinkedIn. Τα αποτελέσματα της ανάλυσης αποτελούν χρήσιμα εργαλεία παρακολούθησης και μέτρησης ενδιαφέροντος για προϊόντα, υπηρεσίες και εταιρείες από άτομα που ασχολούνται με το μάρκετινγκ ή την οικονομία γενικότερα. Εκτός της ανάλυσης των δεδομένων για εμπορικούς σκοπούς, η ανάλυση των δεδομένων κοινωνικών δικτύων φαίνεται εξαιρετικά χρήσιμη στον προσδιορισμό της ανθρώπινης κινητικότητας.

Σήμερα, οι αναρτήσεις των χρηστών στα μέσα κοινωνικής δικτύωσης (φωτογραφίες, κείμενά και βίντεο) μπορούν να επισημανθούν με γεωγραφική ετικέτα, δηλαδή να συσχετιστούν με μια γεωγραφική θέση και χρονική στιγμή.

Έχοντας στη διάθεσή μας τις αναρτήσεις αυτές, είναι δυνατή η ανακατασκευή της πορείας των χρηστών ακολουθώντας απλά την αλληλουχία των δημοσιεύσεών τους, με την προϋπόθεση πως οι τελευταίες περιλαμβάνουν χωροχρονικές πληροφορίες (όπως τη χρονική στιγμή που έγινε η δημοσίευση).

Ορισμένες πλατφόρμες, όπως λόγω χάρη το Twitter παρέχουν δύο τύπους γεωγραφικής ετικέτας (geo-tag):

- Την ακριβή γεωγραφική τοποθεσία της ανάρτησης (δηλαδή το γεωγραφικό πλάτος και μήκος της τοποθεσίας που έγινε η ανάρτηση. Σημείωση: η μορφή αυτή καταργήθηκε πρόσφατα)
- Μια προκαθορισμένη τοποθεσία (πόλη, περιοχή ή εστιατόριο) που προτείνει η πλατφόρμα στο χρήστη κατά τη σύνταξη της ανάρτησης

Σε άλλες πλατφόρμες, όπως το Foursquare και στο Facebook, οι χρήστες μπορούν να κάνουν check-in σε προκαθορισμένες τοποθεσίες που ονομάζονται "venues", δηλαδή σημεία ενδιαφέροντος που παρέχουν πληροφορίες σχετικά με κοινωνικά, πολιτιστικά και δομικά στοιχεία μιας γεωγραφικής περιοχής (π.χ. πόλεις, καταστήματα, μουσεία). Ένας χώρος (venue) σχετίζεται με μια φυσική τοποθεσία (ζεύγος γεωγραφικού πλάτους και μήκους) και πληροφορίες κειμένου (περιγραφή του μέρους ή των δραστηριοτήτων που σχετίζονται με το μέρος) και μπορεί να ακολουθήσει μια ιεραρχική κατηγοριοποίηση που παρέχει διαφορετικά επίπεδα λεπτομέρειας σχετικά με τις δραστηριότητες (π.χ. φαγητό, ασιατικό εστιατόριο, ελληνική ταβέρνα κ.α.).

Στο Yelp, όποτε ο χρήστης δημοσιεύει ένα σχόλιο ή μια φωτογραφία, μπορεί να το συσχετίσει με μια συγκεκριμένη γεωγραφική τοποθεσία επιχείρησης ή χώρου. Ομοίως, οι χρήστες του Flickr μπορούν να δημοσιεύουν και να προσθέτουν γεωγραφικές ετικέτες σε φωτογραφίες, οι οποίες μπορεί επίσης να σχετίζονται με προκαθορισμένες γεωγραφικές συντεταγμένες ή τοποθεσίες.

Σε γενικές γραμμές, μια εγγραφή με γεωγραφική ετικέτα καταγράφει το αναγνωριστικό του χρήστη που τη δημοσιεύει, το είδος της δημοσίευσης (όπως ανάρτηση, tweet, φωτογραφία, βίντεο), τη χρονική στιγμή δημοσίευσης (timestamp) και ανάλογα με την πλατφόρμα, το αναγνωριστικό ή την κατηγορία του χώρου ή της τοποθεσίας σε μορφή συμβολοσειράς (π.χ. Άγαλμα της Ελευθερίας, Νέα Υόρκη).

Επίσης, μπορεί να περιέχει ένα ζεύγος γεωγραφικών συντεταγμένων (γεωγραφικού πλάτους και μήκους).

Στις περισσότερες πλατφόρμες κοινωνικών μέσων, οι γεωγραφικές αναρτήσεις μπορούν να ληφθούν μέσω της διεπαφής προγραμματισμού εφαρμογής (application programming interface, API) τους. Πρέπει ωστόσο να σημειωθεί πως τίθενται περιορισμοί στον αριθμό των αναρτήσεων και των ερωτημάτων που μπορούν να πραγματοποιηθούν ημερησίως ή απαιτούνται διαπιστευτήρια και εξουσιοδοτήσεις από την πλατφόρμα προκειμένου οι χρήστες να μπορέσουν να κατεβάσουν τα δεδομένα.

Η τοποθεσία των χρηστών είναι διαθέσιμη μόνο όταν δημοσιεύουν κάτι ή κάνουν check-in σε έναν χώρο. Η συχνότητα των επισημάνσεων διαφέρει από χρήστη σε χρήστη, ενώ συχνά οι τοποθεσίες επισημαίνονται με αραιό ρυθμό, προκαλώντας ζητήματα ανεπαρκών δεδομένων. Το σημαντικό πλεονέκτημα της αξιοποίησης δεδομένων κοινωνικών μέσων είναι ότι προσφέρουν έναν αντικειμενικό ορισμό της θέσης των χρηστών, διευκολύνοντας κατά συνέπεια την επεξεργασία των αναλυόμενων δεδομένων.

2.6.2 Αξιοποίηση των δεδομένων στην ανάλυση της ανθρώπινης κινητικότητας

Στη βιβλιογραφία, τα GIS, οι καταγραφές λεπτομερειών κλήσεων και τα δεδομένα που πηγάζουν από τα κοινωνικά δίκτυα αποτελούν συχνά τα αναλυόμενα δεδομένα για την εξαγωγή συμπερασμάτων ανθρώπινης κινητικότητας. Η ακρίβεια των ανωτέρω δεδομένων δεν είναι πάντοτε η ίδια. Για παράδειγμα, τα CDRs ανανεώνονται κάθε φορά που κάποιος χρήστης πραγματοποιεί μια κλήση ή αποστέλλει ένα μήνυμα [25]. Επειδή ακριβώς περιέχουν δεδομένα όπως την τοποθεσία (ο τηλεφωνικός πύργος που διεκπεραίωσε την κλήση) και τη διάρκεια της κλήσης, μπορούν να προσδώσουν πολύ χρήσιμες πληροφορίες για έναν ερευνητή. Ένα ακόμη πλεονέκτημα της χρήσης CDRs ως πηγής πληροφοριών, είναι ότι αποτελούν μια ανεξάντλητη πηγή μελέτης για μεγάλες χρονικές περιόδους ενώ παράλληλα μπορούν να συμπεριλάβουν μεγάλες γεωγραφικές περιοχές. Τέλος, είναι μια ανέξοδη τεχνική μελέτης, καθώς τα δεδομένα της συλλέγονται αυτόματα για

όλους τους χρήστες ενός δικτύου.

Επιτόπιες έρευνες Ένα αντιπροσωπευτικό παράδειγμα επιτόπιας έρευνας έγινε στην περιοχή του Σικάγου, μεταξύ 2007 και 2008 [26]. Η έρευνα αυτή περιείχε, μεταξύ άλλων στοιχείων, πληροφορίες σχετικά με τις ταξιδιωτικές δραστηριότητες των μελών μίας οικογένειας. Συνολικά, συλλέχθηκαν δεδομένα μετακίνησης μιας ή δυο ημερών για 10,552 νοικοκυριά. Καταγράφηκαν λοιπόν λεπτομερώς πληροφορίες για τις μετακινήσεις των ανθρώπων, όπως ο σκοπός, ο τρόπος μετακίνησης καθώς και οι ώρες αναχώρησης και άφιξης στον προορισμό. Επίσης, συλλέχθηκαν και πληροφορίες για τυχόν δημόσιες συγκοινωνίες που χρησιμοποιήθηκαν κατά τη μετακίνηση αυτή. Τέτοιες πληροφορίες περιελάμβαναν το αντίτιμο της μετακίνησης, το είδος της συγκοινωνίας που χρησιμοποιήθηκε και το αναγνωριστικό αυτής, τη στάση επιβίβασης και αποβίβασης, ενώ συμπεριλήφθηκε και η απόσταση από τον τελικό προορισμό. Μια παρόμοια έρευνα πραγματοποιήθηκε από τους Liang et al. στο [27]. Στη μελέτη αυτή αναλύθηκαν στοιχεία 46.000 διαδρομών μεταξύ 2.017 σημείων του Λος Άντζελες.

Καταλήγοντας, αξίζει να επισημανθεί πως οι επιτόπιες έρευνες μικρής κλίμακας συχνά συνοδεύονται από άλλες πηγές δεδομένων, όπως για παράδειγμα δεδομένα GPS, προκειμένου να προσδιοριστεί με μεγαλύτερη ακρίβεια η θέση των ατόμων [21].

CDRs και δεδομένα GPS Υπάρχουν πολλές έρευνες που αξιοποιούν δεδομένα λεπτομερειών καταγραφών κλήσεων (CDRs) και γεωοεντοπισμού (GPS) προκειμένου να εξάγουν συμπεράσματα ανθρώπινης κίνησης και να προβλέψουν μελλοντικές της καταστάσεις. Για παράδειγμα, στην [28] οι Jiang et al. εξήγαγαν μοτίβα ανθρώπινης κινητικότητας από δεδομένα CDR προκειμένου να κατανοήσουν τη συμπεριφορά των ανθρώπων κατά τη μετακίνηση. Επίσης μερίμνησαν στον τομέα της αστικής και συγκοινωνιακής σχεδίασης, παρέχοντας εργαλεία και αλγόριθμους για εφαρμογές μεγάλου όγκου δεδομένων (big data).

Όσον αφορά τα δεδομένα GPS, ποικίλλουν οι έρευνες που τα αξιοποιούν. Οι Yuan et al. [16] παρουσίασαν έναν αλγόριθμο ανακάλυψης περιοχών ανακάλυψης περιοχών με που εξυπηρετούν διαφορετικές λειτουργίες και ανάγκες μιας πόλης.

Αξιοποίησαν δεδομένα γεωεντοπισμού κινητικότητας μεταξύ των περιοχών μιας πόλης καθώς και δεδομένα σημείων ενδιαφέροντος. Επίσης, ο αλγόριθμος χώρισε την πόλη σε τμήματα και υπολόγισε τη δημοτικότητα κάθε περιοχής. Έτσι κατάφεραν να επισημάνουν τις διάφορες περιοχές των πόλεων ως προς τον σκοπό που εξυπηρετούν. Ο αλγόριθμος αυτός εξετάστηκε και αναγνώρισε με μεγάλη ακρίβεια περιοχές του Πεκίνου, βασιζόμενος σε δεδομένα GPS σημείων ενδιαφέροντος του Πεκίνου και διαδρομών ταξί μεταξύ 2010 και 2011.

Ιδιαίτερο ενδιαφέρον παρουσιάζει η έρευνα των Williams et al. [29], όπου προτείνεται μία μέθοδος συνδυασμού δεδομένων CDR και GIS. Στην έρευνα, εκφράζονται αμφιβολίες ως προς την απλή ανάλυση δεδομένων καταγραφών κλήσεων (CDR). Μάλιστα σημειώνεται πως συχνά, λόγω φόρτου δικτύου, οι κλήσεις ανακατευθύνονται από τον κοντινότερο τηλεφωνικό πύργο σε έναν με χαμηλότερη απασχόληση. Έτσι καταγράφονται ψευδείς τοποθεσίες των ανθρώπων, με αποτέλεσμα να αναγνωρίζονται εσφαλμένα ως κινήσεις. Επίσης, νεόδημοι τηλεπικοινωνιακοί πύργοι αλλάζουν τη χωρική πυκνότητα του δικτύου. Η προτεινόμενη μέθοδος συνδυάζει τα δεδομένα CDR με τα δεδομένα GIS και υποδιαιρεί μια περιοχή σε γεωγραφικά τετράγωνα 5 χιλιομέτρων. Ο συνδυασμός αυτός φαίνεται πως ενισχύει την ακρίβεια ανάλυσης της ανθρώπινης κινητικότητας και αποτελεί μια καινοτόμα προσέγγιση στην μελέτη της.

Δεδομένα WiFi Σε μια έρευνα, οι Traunmueller et al. [30] χρησιμοποίησαν καταγραφές WiFi και δεδομένα τοποθεσίας προκειμένου να αναλύσουν διαδρομές ανθρώπινης κίνησης σε πόλεις. Συγκεκριμένα, ανέπτυξαν ένα μοντέλο που αξιοποιεί δεδομένα WiFi μεγάλης κλίμακας για την εξαγωγή συμπερασμάτων ανθρώπινης κινητικότητας σε πυκνά αστικά περιβάλλοντα. Άντλησαν δεδομένα από ένα δημόσιο δίκτυο WiFi με 54 σημεία πρόσβασης στη Νέα Υόρκη και συνέλεξαν περισσότερες από τριάντα εκατομμύρια καταγραφές που προήλθαν από οχτακόσιες χιλιάδες μοναδικές συσκευές. Με την ανάλυση των χωρικών δικτύων προσδιορίστηκαν βασικές μονάδες μέτρησης της ανθρώπινης κινητικότητας, εξήχθησαν συμπεράσματα για αυτήν και υπογραμμίστηκαν οι προοπτικές της αξιοποίησης αντίστοιχων δεδομένων σε μελλοντικές αναλύσεις.

Δεδομένα κοινωνικών δικτύων Σε μια έρευνα [31], αναλύθηκαν δημοσιεύσεις στο Twitter με γεωγραφική ετικέτα (geo-tagged tweets) που προήλθαν από τις Ηνωμένες Πολιτείες της Αμερικής. Από τις δημοσιεύσεις αυτές, εξήχθησαν βασικές μονάδες μέτρησης της ανθρώπινης κινητικότητας, καθώς και μερικές ετικέτες όπως η κατοικία και τα κέντρα δραστηριότητας (activity centers). Επίσης, διερευνήθηκαν τα χωροχρονικά χαρακτηριστικά κατανομής και κινητικότητας των χρηστών του Twitter μιας περιοχής. Συμπερασματικά, είναι εμφανές πως τα κοινωνικά δίκτυα παρέχουν πληροφορίες υψηλού ενδιαφέροντος για την επιστήμη της ανθρώπινης κινητικότητας και η ανάλυσή τους προσφέρει οφέλη σε αρκετούς τομείς της οικονομίας.

2.6.3 Ψηφιακό Αποτύπωμα

Το ποσοστό διείσδυσης των κινητών τηλεφώνων αγγίζει το 96% παγκοσμίως [32], γεγονός που τα καθιστά ως μια αρκετά αξιόπιστη πηγή μέτρησης της ανθρώπινης κινητικότητας. Ωστόσο, αυτό δεν ισχύει πάντοτε καθώς στα δεδομένα καταγραφής κινητών τηλεφώνων δεν αντικατοπτρίζονται όλες οι κινήσεις ενός πληθυσμού. Για παράδειγμα, σε αρκετές τοποθεσίες απαγορεύεται η χρήση κινητών τηλεφώνων, στις οποίες οι ιδιοκτήτες τους οφείλουν να τα έχουν απενεργοποιημένα. Τέτοιοι χώροι, μεταξύ άλλων, είναι οι κινηματογράφοι, τα σχολεία ή οι δικαστικές αίθουσες. Μπορεί λοιπόν η καταγεγραμμένη δραστηριότητα ενός χρήστη να διακοπεί για μια περίοδο της ημέρας. Στην περίπτωση αυτή, η φαινομενική έλλειψη δραστηριότητας μπορεί να προσεγγιστεί και να συμπληρωθεί συγκρίνοντας τα χαρακτηριστικά των καταγραφών κινητών τηλεφώνων αμέσως πριν και μετά τη διακοπή.

Τα ίχνη που αφήνουν οι χρήστες από της ηλεκτρονικές τους συσκευές συνιστούν το **ψηφιακό τους αποτύπωμα**. Ένα ακόμη πλεονέκτημα της αξιοποίησης του μεγάλου ποσοστού κατοχής κινητών τηλεφώνων, είναι πως οι ιδιοκτήτες τους μετακινούνται μαζί με τις συσκευές τους καθ' όλη της διάρκειας της ημέρας. Αυτό συμβαίνει γιατί οι άνθρωποι συνηθίζουν να μην τα αποχωρίζονται όταν ταξιδεύουν ή κατά τη διάρκεια δραστηριοτήτων στις οποίες δε τα χρησιμοποιούν. Μάλιστα, μια έρευνα κατέδειξε πως το 87% των ανθρώπων στις Ηνωμένες Πολιτείες της Αμερικής χρησιμοποιούν ενεργά τις κινητές τους συσκευές κατά τη διάρκεια της μετακίνησής τους από και προς την εργασία τους [33].

Προκύπτει πως με την αξιοποίηση κατάλληλης τεχνολογίας αισθητήρων, τα κινητά τηλέφωνα - ή ακόμη καλύτερα τα smartphones - μπορούν να αποτελέσουν μια αξιόπιστη πηγή πληροφόρησης για την κίνηση των ανθρώπων καθ' όλη τη διάρκεια της ημέρας. Επιπλέον, το πολύ υψηλό ποσοστό κατοχής κινητών τηλεφώνων και smartphones δίνει τη δυνατότητα διεξαγωγής αξιόπιστων μετρήσεων που συμπεριλαμβάνουν ομοιόμορφα μεγάλες μερίδες πληθυσμού με προσεγγιστικά μεγάλη κάλυψη των ιδιαίτερων χαρακτηριστικών του πληθυσμού αυτού.

Συμπερασματικά, το ψηφιακό αποτύπωμα των χρηστών αρκεί για να τους ταυτοποιήσει με μεγάλη ακρίβεια, όπως θα έκανε και το φυσικό τους αποτύπωμα. Συνεπώς οι πηγές δεδομένων που εμπεριέχουν ευαίσθητα προσωπικά δεδομένα οφείλουν να εφαρμόζουν αλγορίθμους τεχνητού θορύβου και να μεριμνούν για την ανωνυμία των χρηστών. Η διασφάλιση της ιδιωτικότητας και της προστασίας των προσωπικών δεδομένων προϋποθέτει την προσεκτική μεταχείριση δεδομένων που εμπεριέχουν τα ψηφιακά αποτυπώματα των ανθρώπων.

2.6.4 Συγκριτική παρουσίαση πηγών δεδομένων μέτρησης ανθρώπινης κινητικότητας

Στα πλαίσια της ανάλυσης δεδομένων για τον υπολογισμό της ανθρώπινης κινητικότητας, καταγράφηκαν και συλλέχθηκαν ποικίλα δεδομένα αστικής κινητικότητας ανθρώπων, λόγω της σημαντικής αύξησης των αισθητήρων και των σχετικών τεχνολογιών. Σημαντική ώθηση επίσης, έδωσε η ύπαρξη υπερυπολογιστών, ικανών να αναλύσουν αλγόριθμους με μεγάλη πολυπλοκότητα σε πυκνά datasets, ανάλυση που υπό άλλες συνθήκες θα ήταν αδύνατο να πραγματοποιηθεί.

Τα κύρια σύνολα δεδομένων ανθρώπινης κινητικότητας που καταγράφονται μπορούν να ομαδοποιηθούν ως προς το σημείο συλλογής πληροφορίας στις ακόλουθες τρεις κατηγορίες [34]:

1. Δεδομένα που καταγράφουν μια σχετική τοποθεσία ως προς κάποια σημεία (Bluetooth ή WiFi Κεραίες τηλεπικοινωνίας κ.α.)
2. Δεδομένα που καταχωρούν απευθείας πληροφορίες GPS από μεμονωμένες συσκευές

3. Δεδομένα εγγραφών στιγμάτων γεωεντοπισμού (GPS) που καταγράφονται από μέσα μεταφοράς, όπως λεωφορεία, τρόλεϊ και ταξί.

Δεδομένα που καταγράφουν σχετική τοποθεσία Τα σύνολα δεδομένων σχετικής τοποθεσίας συλλέγουν τις πληροφορίες τοποθεσίας μιας κινητής συσκευής, όπως για παράδειγμα ενός κινητού τηλεφώνου, ως προς κάποιο σημείο του οποίου η θέση συνήθως είναι γνωστή. Οι πληροφορίες αυτές δεν είναι άμεσες, αλλά έμμεσες και αναφέρονται στο σημείο πρόσβασης που έχει συνδεθεί η συσκευή. Τα σημεία αυτά μπορεί να είναι κεραιές τηλεπικοινωνίας, WiFi hotspots κ.α.

Εάν γνωρίζουμε τη θέση του σημείου πρόσβασης, μπορούμε να συμπεράνουμε με κάποιο σφάλμα την τοποθεσία του ατόμου. Τα δεδομένα αυτά μπορούν επίσης να περιέχουν πληροφορίες σχετικές με τα κοινωνικά δίκτυα ενός ατόμου. Για παράδειγμα, δύο άτομα μπορούν να θεωρηθούν φίλοι ή γνωστοί εάν παραμένουν στην ίδια τοποθεσία για μεγάλο χρονικό διάστημα, ενώ αν διανυκτερεύουν στο ίδιο σημείο να θεωρηθούν συγγάμοι ή συγγενείς. Το μειονέκτημα είναι ότι η λεπτομέρεια των συνόλων δεδομένων σημείου πρόσβασης είναι συνήθως χαμηλή, καθώς η ακρίβεια θέσης κεραιάς τηλεπικοινωνιών είναι συνήθως $\pm 2-3$ χιλιόμετρα, αισθητά χειρότερη από την ακρίβεια θέσης γεωεντοπισμού (GPS).

Δεδομένα που περιγράφουν ακριβή θέση Τα εξατομικευμένα σύνολα δεδομένων κινητικότητας περιέχουν τη γεωγραφική θέση κάθε ατόμου που διαθέτει μια συσκευή καταγραφής. Τα δεδομένα αυτά θεωρούνται συχνά τα βέλτιστα για την μοντελοποίηση της ανθρώπινης κινητικότητας, ωστόσο διαθέτουν κάποια μειονεκτήματα. Οι καταγραφές αυτές εμπεριέχουν πληροφορίες που εμπίπτουν στην κατηγορία των ευαίσθητων προσωπικών δεδομένων. Τόσο η νομοθεσία του εκάστοτε κράτους όσο και οι προτιμήσεις απορρήτου των ατόμων, δυσχεραίνουν τη διαδικασία συλλογής δεδομένων με αποτέλεσμα να προκύπτουν ποικίλα εμπόδια σε καταγραφές ευρείας κλίμακας.

Οι συλλογές πληροφοριών που περιγράφουν την ακριβή θέση μπορούν να μας καταδείξουν την καθημερινή ρουτίνα των ανθρώπων. Οι συνήθειες αυτές μπορεί να περιλαμβάνουν τον τόπο διαμονής, εργασίας ή σπουδών, προτιμώμενων διαδρομών, εστιατορίων και άλλων. Είναι λοιπόν προφανές, ότι οι καταγραφές αυτές θέτουν σε

κίνδυνο την ιδιωτικότητα των συμμετεχόντων και συνεπώς πολλές φορές προστίθεται τεχνητός θόρυβος στις πληροφορίες ή εφαρμόζονται σε αυτές αλγόριθμοι που διασφαλίζουν την ανωνυμία των χρηστών [35,36].

Ένα παράδειγμα δεδομένων που περιγράφουν την ακριβή θέση, μπορεί να θεωρηθεί το σύνολο πληροφοριών που καταγράφει ένας αισθητήρας ενσωματωμένος σε ταξί. Τα σύνολα δεδομένων ταξί συνήθως περιέχουν τις παρακάτω πληροφορίες: αναγνωριστικό ταξί, χρονική σήμανση και θέση ταξί (γεωγραφικό μήκος, γεωγραφικό πλάτος). Στα μοτίβα κινητικότητας ταξί, οι οδηγοί μετακινούνται είτε για να παραλάβουν και να αφήσουν πελάτες ή για να σταθμεύσουν σε χώρους αναμονής περιμένοντας να υποδεχτούν νέους πελάτες. Έτσι, η τοποθεσία παραλαβής και παράδοσης μπορεί να θεωρηθεί ως η προέλευση και ο προορισμός του εκάστοτε ταξιδιού για έναν επιβάτη.

Δεδομένα που συνοψίζουν θέσεις πολλών ατόμων Τα δεδομένα που συνοψίζουν τις θέσεις πολλών ατόμων αναφέρονται συχνά σε συλλογές πληροφοριών από μέσα μαζικής μεταφοράς. Τέτοιες πληροφορίες μπορεί να είναι δεδομένα κίνησης λεωφορείων, τρόλεϊ ή τα δεδομένα κίνησης του μετρό. Στα προαναφερθέντα μέσα μεταφοράς, μπορούν να μετακινούνται ταυτόχρονα περισσότεροι του ενός άνθρωποι και κατά συνέπεια να προσδίδουν μια πιο γενική εικόνα της ανθρώπινης κινητικότητας. Πέραν του γεωγραφικού στίγματος των οχημάτων, τα τελευταία χρόνια, έχει πραγματοποιηθεί η άφιξη των ψηφιακών καρτών επιβίβασης στα μέσα μεταφοράς. Με αυτό τον τρόπο, οι ερευνητές έχουν στη διάθεσή τους ακόμη περισσότερες πληροφορίες για την ανθρώπινη κινητικότητα, αφού μπορούν σε κάθε στιγμή να προσδιορίσουν όχι μόνο το σύνολο των επιβατών στο εκάστοτε μέσο αλλά και τα ατομικά μοτίβα κινητικότητας των επιβατών.

Συγκριτικός πίνακας Στον πίνακα 2.1 καταγράφονται οι συχνότερες τεχνολογίες συνόλων δεδομένων που απαντώνται στη βιβλιογραφία. Οι τεχνολογίες αυτές συγκρίνονται ως προς τον τύπο καταγραφής της θέσης του χρήστη (άμεση, έμμεση), ως προς τη συχνότητα δειγματοληψίας και τη δυσκολία απόκτησης πληροφοριών και συνόλων δεδομένων. Επιπλέον, όπως σε κάθε περίπτωση σημειώνονται τα πλεονεκτήματα και τα μειονεκτήματά αυτών.

Ως προς την ευκολία απόκτησης πληροφοριών, στον πίνακα 2.1 διακρίνονται οι καταγραφές λεπτομερειών κλήσεων (CDRs). Αυτό οφείλεται στο γεγονός πως οι καταγραφές αυτές δημιουργούνται αυτόματα, κάθε φορά που κάποιος χρήστης πραγματοποιεί μια κλήση ή αποστέλλει ένα μήνυμα. Τα δεδομένα αυτά αποθηκεύονται ή δέχονται επεξεργασία από τους παρόχους τηλεπικοινωνιακών υπηρεσιών.

Έχει διαπιστωθεί πως οι πληροφορίες τοποθεσίας που εξάγονται από τα δίκτυα κινητής τηλεφωνίας, είναι εξαιρετικά σημαντικές για τη μελέτη της ανθρώπινης κινητικότητας [37]. Οι πληροφορίες αυτής της κατηγορίας περιλαμβάνουν τις καταγραφές λεπτομερειών κλήσεων και δεδομένα παγκόσμιου συστήματος εντοπισμού θέσης (GPS). Το σημαντικό τους πλεονέκτημα είναι πως συλλέγουν παθητικά (passively) πληροφορίες τοποθεσίας, ενόσω οι χρήστες εκμεταλλεύονται τις υπηρεσίες του παρόχου τηλεπικοινωνίας. Η παθητική καταγραφή συνεπάγεται και χαμηλό κόστος υλοποίησης και απόκτησης της πληροφορίας.

Τα τελευταία χρόνια, οι χρήστες και οι σχετικές νομοθεσίες έχουν περιορίσει την διαθεσιμότητα δεδομένων γεωεντοπισμού. Η συλλογή δεδομένων GPS αυξάνει την κατανάλωση ενέργειας των τεχνολογικών συσκευών και κατά συνέπεια πολλοί χρήστες την απενεργοποιούν, ενώ επίσης τίθενται ζητήματα προστασίας της ιδιωτικότητας. Έτσι, σε πολλές έρευνες προκειμένου να συλλεχθούν δεδομένα γεωεντοπισμού επιλέγεται μια περιορισμένη ομάδα εθελοντών με παρόμοιες συνήθειες διαβίωσης.

Ένα ακόμη μειονέκτημα των δεδομένων γεωεντοπισμού είναι πως δεν είναι εύκολο να καταγραφούν σε εσωτερικούς ή υπόγειους χώρους και συχνά εμφανίζονται ψευδώς άλματα θέσης. Όσον αφορά τις καταγραφές λεπτομερειών κλήσεων (CDR), καταγράφουν τις ταυτότητες του πύργου τηλεπικοινωνίας όταν οι κινητές συσκευές καλούν ή δέχονται μια κλήση ή γραπτό μήνυμα (SMS). Αυτό τις καθιστά αραιές στο χρόνο και απρόβλεπτες.

Τα ανωτέρω μειονεκτήματα των δεδομένων GPS και των καταγραφών λεπτομερειών κλήσεων περιορίζουν το εύρος των εφαρμογών τους για τη μελέτη της ανθρώπινης κινητικότητας ενός πληθυσμού σε περιοχές με μεγάλη πυκνότητα κτιρίων, όπως λ.χ. μια πόλη. Πρόσφατα, η πρόσβαση των πολιτών σε συσκευές που υποστηρίζουν δεδομένα υψηλότερης ταχύτητας (αρχικά LTE, στη συνέχεια 4g και

Τεχνολογία	Καταγραφή Θέσης	Δειγματοληψία	Απόκτηση πληροφ.	Πλεονεκτήματα	Μειονεκτήματα
Ερωτηματολόγια και Τοπικές έρευνες	Άμεση	-	Δύσκολη	Μπορούν να περιέχουν στοχευμένες πληροφορίες	-Ακρίβη και χρονοβόρα υλοποίηση -Η ακρίβεια των μετρήσεων εξαρτάται από τα άτομα που διεξάγουν τις έρευνες
Καταγραφές WiFi	Άμεση	Υψηλή	Μέτρια	Μεγάλη ακρίβεια Ενεργειακή απόδοση	-Μικρή περιοχή κάλυψης -Μεγάλο κόστος τοποθέτησης σημείων πρόσβασης στο δίκτυο
GPS	Άμεση, Έμμεση	Χαμηλή (1 sec)	Μέτρια	-Μεγάλη ακρίβεια -Μπορεί να αξιοποιηθεί στο διαχωρισμό μεταξύ ειδών μετακίνησης (πεζή, σε όχημα κ.α.) - Δημιουργούνται αυτόματα, καθιστώντας τα ως την πλέον οικονομική λύση	-Αυξημένες ανάγκες ενέργειας σε σχέση με άλλες μεθόδους -Ακρίβη -Αδυναμία καταγραφής σε κλειστούς χώρους
Καταγραφές Λεπτομερειών Κλήσεων (CDRs)	Έμμεση	Χαμηλή	Εύκολη	- Μπορούν να καταγράψουν μεγάλη μερίδα πληθυσμού	Μεγάλο μεσοδιάστημα καταγραφών Μικρή ακρίβεια θέσης
Ενεργός τηλεπικοινωνιακός εντοπισμός χρηστών (LTE)	Άμεση	Υψηλή	Εύκολη	Μεγαλύτερη ακρίβεια σε σχέση με τις καταγραφές λεπτομερειών κλήσεων Δεύτερη οικονομικότερη υλοποίηση	-Κοστοβόρα λύση σε σχέση με τις καταγραφές λεπτομερειών κλήσεων -Ο όγκος των δεδομένων δυσχεραίνει τη διαχείρισή τους

Πίνακας 2.1: Συγκριτικός πίνακας των συνόλων δεδομένων (datasets) που συναντώνται συχνότερα στη βιβλιογραφία.

πλέον 5g) δίνει τη δυνατότητα σε ερευνητές να εκμεταλλευτούν την σύνθετη πληροφορία τους και τον όγκο των δεδομένων που προκύπτουν προκειμένου να μπορούν να προβλέψουν με ακόμη μεγαλύτερη ακρίβεια την κίνηση των ανθρώπων στο χώρο και το χρόνο. Αξίζει ωστόσο να σημειωθεί, πως η διερεύνηση της κινητικότητας των χρηστών με χρήστες μεγάλης κλίμακας σε μια πόλη, επιτρέπει την υλοποίηση προηγμένων εφαρμογών με επίκεντρο τον άνθρωπο καθώς δίνει τη δυνατότητα σε μια έξυπνη πόλη να συνεργάζεται πιο αποτελεσματικά και ενεργά με τους πολίτες της.

3. Περιγραφή Μεθοδολογιών

Εισαγωγή Τα μοντέλα ανθρώπινης κινητικότητας μπορούν να διαχωριστούν ως προς τη στοχοθεσία τους σε δυο κατηγορίες:

- Μοντέλα που επιδιώκουν την αναπαράσταση ατομικών μοτίβων κινητικότητας.
- Μοντέλα που αποσκοπούν στην αναπαράσταση μετακινήσεων μιας μερίδας πληθυσμού.

Ανεξαρτήτως του εύρους της αναζήτησης, και οι δυο κατηγορίες μεθοδολογιών διαθέτουν ένα ορισμένο εύρος. Λαμβάνουν δηλαδή κάποιους χωρικούς και χρονικούς περιορισμούς στην ανάλυση που πραγματοποιούν. Το εύρος ανάλυσης μπορεί να κυμαίνεται από λίγα μέτρα ή μια μικρή περιοχή έως ένα γεωγραφικό διαμέρισμα ή μια μεγάλη χώρα. Επίσης, το διάστημα αναζήτησης δύναται να περιορίζεται σε μερικές ώρες ή λεπτά έως μήνες και ολόκληρα έτη. Ανάλογα με το εύρος της αναζήτησης, τα μοντέλα προσέγγισης και ανάλυσης της ανθρώπινης κίνησης πρέπει να προσαρμόζονται σε αυτήν. Γίνεται λοιπόν αντιληπτό πως κάθε κατηγορία ανθρώπινης κινητικότητας και ανάλυσης διαθέτει μοντέλα που μπορούν να την αναπαραστήσουν καλύτερα.

Για παράδειγμα, η τυχαιότητα της ατομικής κίνησης λόγω της ατομικής βούλησης και της ευμεταβλητότητας της ανθρώπινης ζωής απαιτεί κάποιο βαθμό στοχαστικότητας στα μοντέλα που την αναλύουν. Για το λόγο αυτό τα μοντέλα ανάλυσης ανθρώπινης κινητικότητας σε ατομικό επίπεδο χρησιμοποιούν τυχαίους περιπάτους και κίνηση Brown για να την αναπαραστήσουν. Πρέπει ωστόσο να επισημανθεί πως σε μελέτες διαπιστώθηκε, μέσω της ανάλυσης της κίνησης σε ατομικό επίπεδο, ότι οι ατομικές τροχιές, αν και διαθέτουν ένα βαθμό τυχαιότητας, διαθέτουν επίσης μεγάλο βαθμό κανονικότητας και προβλεψιμότητας. Μια έρευνα που έριξε φως στην στοχαστικότητα της κίνησης των ανθρώπων ήταν η [38], στην οποία οι Song et al. ανέλυσαν δεδομένα τηλεφωνίας πενήντα χιλιάδων ατόμων διάρκειας τριών μηνών. Υπολογίζοντας την εντροπία της πορείας του κάθε ανθρώπου, διαπίστωσαν 93% πιθανότητα πρόβλεψης των κινήσεων των ανθρώπων που ανέλυσαν. Παρά τις σημαντικές διαφορές στα μοτίβα κινητικότητας μεταξύ των

ανθρώπων, παρουσιάζεται χαμηλή μεταβλητότητα στην προβλεψιμότητά τους. Αυτή η επαναληψιμότητα παρατηρείται ανεξάρτητα της καθημερινής απόστασης που αυτοί διανύουν. Εν κατακλείδι, η ανθρώπινη κίνηση απέχει αρκετά από την τυχαία κίνηση. Η επαναλαμβανόμενη και περιοδική της φύση μπορεί να αξιοποιηθεί για την πρόβλεψη των μελλοντικών της καταστάσεων και την κατασκευή ρεαλιστικών μοντέλων ανθρώπινης κινητικότητας.

Στο επίπεδο της αναπαράστασης μετακινήσεων μεγαλύτερης μερίδας πληθυσμού, τα μοντέλα που περιγράφουν τη συνολική κινητικότητα πολλών ατόμων στοχεύουν στην κατασκευή των πινάκων προέλευσης - προορισμού (Origin-Destination matrices). Ορίζοντας ένα διάστημα αναζήτησης, κατασκευάζουν τους πίνακες προέλευσης - προορισμού εκτιμώντας το μέσο αριθμό ταξιδιωτών μεταξύ δύο χωρικών περιοχών, όπως για παράδειγμα μεταξύ δυο γειτονικών δήμων. Το χρονικό διάστημα που ορίζεται στις έρευνες αυτές μπορεί να διαρκεί από μία ημέρα, σε περιπτώσεις ροών μετακίνησης, έως και έτη, σε ροές μετανάστευσης.

Στη συνέχεια θα αναλυθούν τα μοντέλα ανάλυσης ανθρώπινης κινητικότητας που παρουσιάζονται συχνότερα στη βιβλιογραφία. Στο κεφάλαιο αυτό θα δοθεί μεγαλύτερη έμφαση σε τεχνικές που θα εξεταστούν πειραματικά στο κεφάλαιο 4. Στο τέλος του κεφαλαίου συντάχθηκε ένας συνοπτικός πίνακας των μοντέλων που καταγράφηκαν, προκειμένου να δοθεί μια οπτική και συγκεντρωτική απεικόνιση τους όσον αφορά το είδος και το πεδίο εφαρμογής.

3.1 Μοντέλα υπολογισμού ατομικής κινητικότητας

Το μοντέλα της ατομικής κινητικότητας χαρακτηρίζονται κυρίως από τα χωροχρονικά μοτίβα κινητικότητας των ατόμων. Για τη μέτρηση της ανθρώπινης κινητικότητας και την ποσοτικοποίησή της, χρησιμοποιούνται κοινώς κάποιες μετρικές, όπως η ακτίνα περιστροφής, η κίνηση Μπράουν και τα μήκη άλματος (jump lengths). Στην παρούσα ενότητα θα αναλυθεί η κίνηση Μπράουν (Brownian motion) και οι πτήσεις Λέβι (Levy flights). Τα μήκη άλματος και η ακτίνα περιστροφής δε θα αναλυθούν, μιας και το θεωρητικό τους υπόβαθρο καταγράφηκε στην ενότητα 2.3.

3.1.1 Κίνηση Μπράουν

Η κίνηση του Μπράουν (Brownian Motion) είναι η τυχαία κίνηση των σωματιδίων που αιωρούνται σε ένα υγρό ή ένα αέριο. Η κίνηση αυτή αποτελείται συνήθως από τυχαίες διακυμάνσεις στη θέση ενός σωματιδίου εντός ενός υγρού πεδίου, ακολουθούμενες από μετακινήσεις σε κάποιο άλλο πεδίο. Κάθε μετακίνηση συνδέεται με μία ή περισσότερες διακυμάνσεις στο νέο κλειστό πεδίο.

Η κινητική ενέργεια των μοριακών κινήσεων της κίνησης Μπράουν, σε συνδυασμό με τις μοριακές περιστροφές και δονήσεις, αποτελούν το θερμικό συστατικό της εσωτερικής ενέργειας ενός υγρού ή αερίου (θεώρημα Equipartition).

Η κίνηση περιγράφηκε για πρώτη φορά από τον Ρόμπερτ Μπράουν, του οποίου έλαβε και τιμητικά το όνομα, το 1827 όταν μελετούσε κόκκους γύρης στο νερό. Το 1905 ο Άλμπερτ Αϊνστάιν ασχολήθηκε με την κίνηση Μπράουν, και υποστήριξε πως το αίτιο της κίνησης αυτής εντοπίζεται στις συγκρούσεις των κόκκων με τα μόρια του νερού [39].

Ας υποθέσουμε ότι ένα σωματίδιο μπορεί να κάνει βήματα μήκους $1/\sqrt{k}$ αριστερά ή δεξιά με ίση πιθανότητα και ότι μετά από χρόνο t το σωματίδιο έχει πραγματοποιήσει $N = tk$ βήματα. Για ένα συγκεκριμένο k , η μετατόπιση του σωματιδίου στο χρόνο t δίνεται από τη σχέση

$$X_k(t) = \frac{1}{\sqrt{k}} \sum_{i=1}^{tk} \Delta X_i,$$

Λαμβάνοντας το όριο $k \rightarrow \infty$, η κίνηση του Μπράουν προκύπτει ως απόρροια του Κεντρικού Οριακού Θεωρήματος (Central Limit Theorem, CLT). Στην πραγματικότητα, στο άπειρο k το $X_k(t)$ τείνει στο $X(t)$, του οποίου η συνάρτηση πυκνότητας πιθανότητας είναι η ακόλουθη:

$$P(x, t) = \frac{1}{\sqrt{2\pi t\sigma^2}} e^{-\frac{(x-\mu kt)^2}{2t\sigma^2}}.$$

Όπου $\mu = \langle \Delta X \rangle$ ο μέσος και $\sigma^2 = \langle \Delta X^2 \rangle$ και η διακύμανση της τυχαίας διαδρομής μετατόπισης.

3.1.2 Πτήσεις Λέβι

Οι πτήσεις Λέβι (Levy Flights) αποτελούν μια κατηγορία τυχαίων περιπάτων στην οποία δεν μπορεί να γίνει χρήση του κεντρικού οριακού θεωρήματος. Αποτελούνται από μια

σειρά μικρών μετατοπίσεων, διασκορπισμένων περιστασιακά από μια πολύ μεγάλη μετατόπιση.

Τυπικά, οι πτήσεις Λέβι ορίζονται ως το άθροισμα των ανεξάρτητων τυχαία κατανεμημένων μεταβλητών των οποίων η συνάρτηση πυκνότητας πιθανότητας για ένα άλμα έχει μια κατανομή της ακόλουθης μορφής:

$$f(\Delta x) \sim \frac{1}{\Delta x^{1+\beta}},$$

για τη οποία ισχύει $0 < \beta < 2$.

Τα μήκη άλματος (jump lengths) των πτήσεων Λέβι ενδέχεται να είναι μεγάλα, όμως συμβαίνουν εντός της ίδιας χρονικής περιόδου. Έτσι, οι πτήσεις Λέβι μπορούν να θεωρηθούν ως μια πρόχειρη προσέγγιση της ανθρώπινης κίνησης.

Οι πτήσεις Λέβι αντλούν το θεωρητικό τους υπόβαθρο από τα μαθηματικά της θεωρίας του χάους (Chaos Theory) και αξιοποιούνται σε στοχαστικές μετρήσεις και προσομοιώσεις για τυχαία ή ψευδο-τυχαία φυσικά φαινόμενα. Παραδείγματα επιστημονικών πεδίων εφαρμογής των πτήσεων Λέβι περιλαμβάνουν, μεταξύ άλλων, την ανάλυση σεισμικών δεδομένων, τα οικονομικά, την κρυπτογραφία και τη βιολογία.

Οι πτήσεις Λέβι στη βιβλιογραφία Στην έρευνά τους, οι Rhee et al. [22] παρατήρησαν πως τα μοτίβα ανθρώπινης κίνησης έχουν παρόμοια συμπεριφορά με αυτή των Levy flights. Μάλιστα, το συμπέρασμα αυτό τους προκάλεσε εντύπωση καθώς οι άνθρωποι περίπατοι δεν είναι τυχαίοι, παρόλα αυτά εμφανίζουν αρκετές ομοιότητες με αυτές των τυχαίων περιπάτων. Η ανάλυσή τους, συμπεριέλαβε 226 ημερήσια ίχνη γεωεντοπισμού GPS που συλλέχθηκαν από 101 εθελοντές σε πέντε διαφορετικές τοποθεσίες. Κατασκεύασαν ένα truncated μοντέλο κινητικότητας περιπάτου Levy, (Truncated Levy-walk mobility model, TLW) που προσομοιώνει τη στατιστική συμπεριφορά που παρατηρήθηκε στην ανάλυσή τους.

Σε μια άλλη έρευνα, οι Kang et al. [40] προσπάθησαν να υπολογίσουν την επιρροή των πόλεων στην ανθρώπινη κίνηση. Υπέθεσαν πως οι πόλεις, με το μέγεθος και την πυκνότητα της αστικής δόμησης, επηρεάζουν σε κάποιο βαθμό την κίνηση των ανθρώπων. Συνέλεξαν δεδομένα οχτώ πόλεων της βορειοδυτικής Κίνας και εξήγαγαν μοτίβα ατομικής κινητικότητας. Προκειμένου να εξάγουν συμπεράσματα για το

ποσοστό αστικοποίησης μιας περιοχής, θεώρησαν ως κριτήριο την χωρική πυκνότητα των τηλεπικοινωνιακών πύργων. Τα αποτελέσματα της έρευνάς τους έδειξαν πως η ενδο-αστική μετακίνηση (Intra-urban mobility) ακολουθεί την εκθετική κατανομή. Οι σταθερές της εκθετικής αυτής κατανομής διαφέρουν από πόλη σε πόλη και εξαρτώνται από το μέγεθός της και την πολεοδομική σχεδίαση. Τέλος, πραγματοποίησαν μια ανάλυση μέσω προσομοίωσης. Κατασκεύασαν ένα μοντέλο Μόντε Κάρλο βασισμένο σε πτήσεις Levy προκειμένου να εξετάσουν και να επιβεβαιώσουν τη σχέση μεταξύ ενδο-αστικής μετακίνησης και πολεοδομίας. Τα αποτελέσματα της έρευνάς τους έδειξαν πως το μέγεθος και η δομή μιας πόλης επηρεάζουν σημαντικά την ατομική κινητικότητα εντός του αστικού ιστού. Οι μεγάλες πόλεις προκαλούν μεγαλύτερες και πιο χρονοβόρες μετακινήσεις, ενώ το ακανόνιστό τους σχήμα περιορίζει την βραχεία μετακίνηση των πολιτών.

3.2 Βαρυτικά μοντέλα

Οι μεθοδολογίες που βασίζονται στο νόμο βαρύτητας αναλύθηκαν αρχικά στη θεωρία συγκοινωνιών [41]. Τα βαρυτικά μοντέλα υποθέτουν πως η ροή μετακίνησης w_{ij} από την οικία στην εργασία και αντίστροφα μεταξύ του τμήματος πληθυσμού i (με συνολικό πληθυσμό N_i) και του τμήματος πληθυσμού j (με συνολικό πληθυσμό N_j) λαμβάνει την παρακάτω μορφή:

$$w_{ij} = C \frac{N_i^\alpha N_j^\gamma}{f(d_{ij})},$$

όπου το C αποτελεί μια σταθερά αναλογικότητας, ενώ τα α και γ προσαρμόζουν την εξάρτηση των ροών μετακίνησης ανάλογα με τα μεγέθη των τμημάτων πληθυσμού. Τέλος, το $f(d_{ij})$ είναι συνάρτηση της απόστασης. Τα βαρυτικά μοντέλα συνήθως υποθέτουν τη συμπεριφορά της συνάρτησης $f(d_{ij})$ αναλογική της εκθετικής ή βασισμένη σε κάποια δύναμη.

3.2.1 Αξιοποίηση των βαρυτικών μοντέλων στην ανάλυση ανθρώπινης κινητικότητας

Τα βαρυτικά μοντέλα ανάλυσης ανθρώπινης κινητικότητας εμφανίζονται συχνά στη βιβλιογραφία. Ένα αντιπροσωπευτικό παράδειγμα εφαρμογής των βαρυτικών μοντέλων είναι η μελέτη [3], στην οποία οι Balcan et al., κάνοντας χρήση βαρυτικών μοντέλων, αναζήτησαν την επιρροή της προσθήκης μικρής εμβέλειας δικτύου επαγγελματικών μετακινήσεων σε ένα παγκόσμιο επιδημιολογικό μοντέλο που περιλαμβάνει όλες τις αεροπορικές μετακινήσεις μεταξύ 3.362 αεροδρομίων. Η προσθήκη αυτή επέτρεψε τη διάκριση μεταξύ σύντομων και μεγάλων ροών ανθρώπινης κινητικότητας.

Πιο συγκεκριμένα, οι Balcan et al. μελέτησαν τις ροές μετακίνησης ανθρώπων από και προς τη δουλειά τους (commuting flows) σε παγκόσμια κλίμακα. Έτσι, κατάφεραν να προσεγγίσουν υπολογιστικά τη **μετάδοση ασθενειών** στον πληθυσμό. Προκειμένου να εντοπίσουν τη συσχέτιση της μικρής κλίμακας μετακίνησης (από την οικία στην εργασία) και των πολύωρων πτήσεων στην επιδημιολογική εξέλιξη, επέκτειναν την έρευνά τους σε παγκόσμιο επίπεδο. Μάλιστα, ανέλυσαν στοιχεία κινητικότητας 29 χωρών. Επιπλέον, κατασκεύασαν ένα βαρυτικό μοντέλο ικανό να περιγράψει τα μοτίβα κινητικότητας από και προς την εργασία των ανθρώπων, συμπεριλαμβάνοντας διαδρομές μετακίνησης (commuting) έως και τριακοσίων χιλιομέτρων. Επίσης, ενσωμάτωσαν σε ένα επιδημιολογικό μοντέλο, μια χωροχρονική τεχνική για την αξιολόγηση της επίδρασης της ανθρώπινης κινητικότητας στη δυναμική μιας ασθένειας.

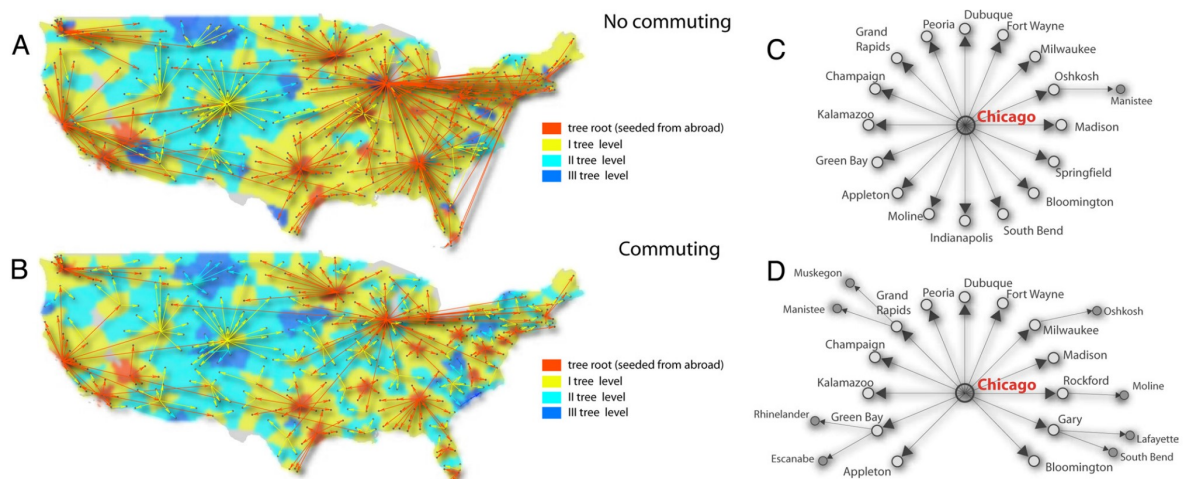
Συνολικά, ανέλυσαν πάνω από 10^4 ροές μετακινούμενων πολιτών παγκοσμίως μεταξύ τμημάτων πληθυσμών. Οι ροές αυτές ορίστηκαν μέσω μιας αποσύνθεσης Voronoi. Διαπιστώθηκε λοιπόν, πως η καλύτερη εφαρμογή της βαρυτικής θεωρίας στην υπό εξέταση περίπτωση, μπορούσε να επιτευχθεί με μια συνάρτηση της μορφής:

$$w_{ij} = C \frac{N_i^\alpha N_j^\gamma}{f(d_{ij}/\kappa)}$$

Χρησιμοποιήθηκε δηλαδή μια εκθετική συνάρτηση της μορφής $f(d_{ij}) = \exp(d_{ij}/r)$, όπου r το χαρακτηριστικό μήκος που ορίζει την αποσύνθεση των ροών μετακίνησης.

Οι εκθέτες α και γ για αποστάσεις (distances, d) μικρότερες ή ίσες των 300 χιλιομέτρων υπολογίστηκαν περίπου ίσοι με 0.46 και 0.64 αντίστοιχα. Για αποστάσεις (d) μεγαλύτερες των 300 χιλιομέτρων βρέθηκαν να είναι περίπου ίσοι με 0.35 και 0.37. Οι ερευνητές υπέθεσαν πως η μεγάλη διαφορά των εκθετών τόσο σε μικρές όσο και σε μεγάλες κλίμακες, πιθανότατα οφείλεται στην επίδραση της απόστασης μεταξύ οικίας και εργασίας. Η επίδραση αυτή φαίνεται να μην ισχύει στις μεγαλύτερες κλίμακες.

Μια οπτική απεικόνιση των ανωτέρω φαίνεται και στο σχήμα 3.1, όπου τα Α και Β αναπαριστούν γεωγραφικά το επιδημιολογικό δέντρο των ηπειρωτικών περιοχών των Ηνωμένων Πολιτειών της Αμερικής. Πιο συγκεκριμένα, το σχήμα Α λαμβάνει υπόψιν μονάχα τις αερομεταφορές, ενώ το σχήμα Β συμπεριλαμβάνει και τις μεταφορές από και προς την εργασία (commuting). Το κόκκινο χρώμα αναπαριστά τις πόλεις-ρίζες της ασθένειας, δηλαδή τις πρώτες πόλεις που είχαν μολυνθεί από το εξωτερικό.



Σχήμα 3.1: Δέντρο επιδημιολογικής εισροής [3].

Όσο απομακρυνόμαστε από τις ρίζες του δέντρου τα χρώματα σταδιακά αλλάζουν από κόκκινο σε κίτρινο και έπειτα σε σκούρο μπλε. Τα βέλη που αντιπροσωπεύουν τις ακμές του δέντρου χρωματίζονται με το χρώμα του κόμβου-γονέα. Τα σχήματα C και D παρέχουν επίσης μια σχηματική αναπαράσταση του επιδημιολογικού δέντρου του Σικάγου. Στο σχήμα C προσμετρώνται μόνο οι πτήσεις, ενώ στο σχήμα D συμπεριλαμβάνεται και η εναέρια κυκλοφορία στις μετακινήσεις από και προς την εργασία. Όπως καταδεικνύεται και στα δύο παραδείγματα, η ροή εξάπλωσης ενισχύεται και μεγεθύνεται από τις αερομεταφορές. Άλλωστε, οι αερομεταφορές

είναι υπεύθυνες κατά κύριο λόγο για την εισροή νέων ιών στην επικράτεια μιας χώρας. Στην έρευνα ωστόσο παρατηρήθηκε πως η πορεία εξάπλωσης διακόπτεται από την εισαγωγή σε αυτήν των ροών εργασιακής μετακίνησης, καθώς ο αριθμός των εστιών γύρω από τους κόμβους των κεντρικών και δευτερευόντων αερολιμένων στους δευτερεύοντες κόμβους αυξάνεται.

Οι ερευνητές διαπίστωσαν στους υπολογισμούς τους πως ο αντίκτυπος της επιδημίας δεν μεταβάλλεται σημαντικά με την προσθήκη της μικρής εμβέλειας μετακινήσεων. Επίσης, παρατηρήθηκε πως η αλληλεπίδραση της μικρής και μεγάλης εμβέλειας κινητικότητας λειτουργεί μόνο στην έναρξη μιας επιδημίας. Όσο η επιδημία αυξάνει τη δυναμική της, οι μετακινήσεις αυτές επιδρούν προοδευτικά όλο και λιγότερο στην εξέλιξή της, αφού τότε ο κύριος πληθυσμός αποτελεί την κινητήρια πηγή μόλυνσης. Επισήμαναν επίσης, πως στην περίπτωση μιας πανδημίας όπου η άμεση λήψη αποφάσεων είναι κρίσιμης σημασίας, θα ήταν πολύ χρήσιμη η αξιοποίηση ενός μοντέλου ανάλυσής της με χαμηλή υπολογιστική πολυπλοκότητα.

Εν κατακλείδι, οι μετακινήσεις μικρής και μεγάλης εμβέλειας συμβάλλουν στα πρώτα στάδια εξάπλωσης μιας επιδημίας. Εάν η υπολογιστική πολυπλοκότητα και ο χρόνος εξαγωγής συμπερασμάτων είναι κρίσιμα μεγέθη για τη μέγιστη αποτελεσματικότητα των αναλύσεων, οι μικρής εμβέλειας μετακινήσεις θα μπορούσαν να παραλειφθούν από τα μελλοντικά βαρυτικά μοντέλα. Έτσι, συμπεράσματα υψηλής αξιοπιστίας θα μπορούσαν να εξαχθούν πιο γρήγορα, χωρίς να γίνουν σημαντικές υποχωρήσεις στην ακρίβεια της ανάλυσης.

3.3 Rule-based Μοντέλα

Στη συγκοινωνιακή θεωρία, η εύρεση των λόγων για τους οποίους οι άνθρωποι μετακινούνται έχει ύψιστη σημασία. Ο στόχος της μετακίνησης συχνά καθορίζει και τις επιλογές μετακίνησης των ανθρώπων. Οι επιλογές μετακίνησης των ανθρώπων ποικίλουν, τόσο ως προς το είδος του μέσου που επιλέγεται, όσο και ως προς την ώρα αναχώρησης και τη βιασύνη τους.

Η πρόοδος στην τεχνολογία αισθητήρων έχει καταστήσει τα καταγραφικά GPS και τα smartphones δημοφιλή εργαλεία για τις έρευνες ανάλυσης ανθρώπινης κινητικότητας. Μια από τις προκλήσεις της εξαγωγής δεδομένων, είναι η ορθή

ταυτοποίηση και αναγνώριση των εκάστοτε δραστηριοτήτων. Για τους ανωτέρω λόγους έχουν αναπτυχθεί μοντέλα αναγνώρισης και επισήμανσης δραστηριοτήτων, ικανά να συνδυάσουν δεδομένα κίνησης και συμφραζόμενα προκειμένου να καταδείξουν το είδος της δραστηριότητας των ατόμων.

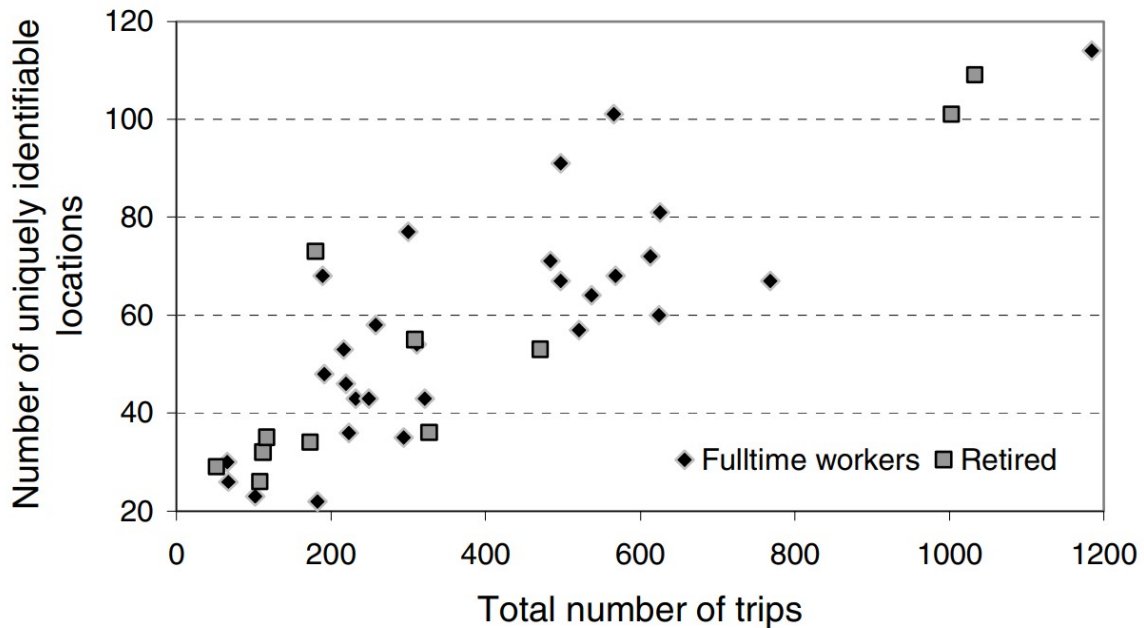
Για την αναγνώριση μιας δραστηριότητας αξιοποιούνται κατά κύριο λόγο δεδομένα GIS, όπως σημεία ενδιαφέροντος (Points of Interest, POI) σε συνδυασμό με χωροχρονικές πληροφορίες, ενώ επιστρατεύονται αλγόριθμοι βασισμένοι σε κανόνες (rule-based algorithms). Οι αλγόριθμοι αυτοί εξειδικεύονται στην ανάλυση δεδομένων που περιέχουν ένα μείγμα αριθμητικών χαρακτηριστικών και συμφραζομένων.

Οι δραστηριότητες που αναγνωρίζονται συνήθως ομαδοποιούνται σε κατηγορίες. Τέτοιες κατηγορίες είναι η παραμονή στην οικία, η εργασία, οι επισκέψεις σε καταστήματα, η περιήγηση σε αξιοθέατα κ.α.

Εφαρμογές των μοντέλων αναγνώρισης δραστηριοτήτων Στη βιβλιογραφία, εμφανίζονται συχνά μοντέλα αναγνώρισης δραστηριοτήτων. Σε μια έρευνα, οι Achhausen et al. [4] ανέπτυξαν μια προσέγγιση βασισμένη σε κανόνες για την αναγνώριση δραστηριοτήτων χρησιμοποιώντας δεδομένα GIS της Σουηδίας. Η τεχνική τους αξιοποίησε έναν αλγόριθμο βασισμένο σε κανόνες και οδήγησε στην αναγνώριση δραστηριοτήτων μέσω των σημείων οικίας, εργασίας και σημείων ενδιαφέροντος. Συγκεκριμένα, αναλύθηκαν δεδομένα GIS σε συνδυασμό με συσκευές GPS που εγκαταστάθηκαν σε εκατοντάδες οχήματα σε τρεις Σουηδικές πόλεις. Η διάρκεια της έρευνας ήταν δυο έτη. Συνολικά καταγράφηκαν 240.435 μετακινήσεις στην περιοχή μελέτης και σχεδόν πενήντα χιλιάδες ημέρες κίνησης οχημάτων. Η έρευνα παρουσίασε μια μεθοδολογία, την "Mobidrive", για την αναγνώριση του τύπου δραστηριότητας και σκοπού μετακίνησης καθώς και τον εντοπισμό απωλειών ή σφαλμάτων καταγραφής.

Ένα ενδιαφέρον εύρημα της έρευνας ήταν η συσχέτιση του αριθμού των μοναδικών τοποθεσιών που επισκέπτονταν οι άνθρωποι με το συνολικό αριθμό των διαδρομών που πραγματοποιούσαν. Παρατηρήθηκε λοιπόν, πως όσο ο συνολικός αριθμός των διαδρομών που πραγματοποιούσαν οι οδηγοί αυξανόταν, τόσο ανοδική ήταν και η πορεία του συνολικού αριθμού των νέων μοναδικών τοποθεσιών. Μια πιθανή εξήγηση που δόθηκε για την παρατήρηση αυτή ήταν πως οφείλεται στην τάση του ανθρώπου

να αναζητεί νέα ερεθίσματα και ποικιλία στη ζωή του (βλ. σχήμα 3.2).



Σχήμα 3.2: Συσχέτιση του αριθμού των μοναδικών αναγνωρισθισών τοποθεσιών και του συνολικού αριθμού ταξιδιών [4]

Σε μια άλλη έρευνα, οι Furletti et al. [8] πρότειναν μια στοχαστική προσέγγιση αυτόματης αναγνώρισης του στόχου δραστηριότητας βάσει της απόστασης ενός σημείου ενδιαφέροντος από την καταγεγραμμένη θέση των χρηστών. Πιο συγκεκριμένα, ο αλγόριθμος συνδύαζε την κατηγορία του σημείου ενδιαφέροντος με ένα μέτρο πιθανότητας βασισμένο στο νόμο βαρύτητας, και ήταν ικανός να συμπεράνει τη δραστηριότητα που πραγματοποιήθηκε.

Προκειμένου να αξιολογηθεί το μοντέλο τους, δοκιμάστηκε πάνω σε μια πραγματική μελέτη. Η μελέτη αυτή αποτελούνταν από διάφορες διαδρομές αυτοκινήτων, οι οποίες είχαν ήδη επισημανθεί από τους οδηγούς των οχημάτων αυτών. Με την εφαρμογή του, το μοντέλο επισήμανε αυτόματα το είδος των δραστηριοτήτων που πραγματοποιούσαν οι χρήστες. Στη συνέχεια, οι ερευνητές συνέκριναν τα αποτελέσματα της τεχνικής με τις επισημάνσεις των οδηγών και κατέληξαν σε ορισμένα ενθαρρυντικά αποτελέσματα. Το μοντέλο φάνηκε να έχει ιδιαίτερα μεγάλη επιτυχία στην αναγνώριση της κατηγορίας δραστηριότητας "φαγητό", με επιτυχία αναγνώρισης 83%. Τα αποτελέσματα της εφαρμογής του μοντέλου ως προς άλλες κατηγορίες δραστηριότητας φαίνονται στον πίνακα 3.1. Η

στήλη Α του πίνακα παρουσιάζει την ακρίβεια ταξινόμησης ανά δραστηριότητα, ενώ η Β τον συνολικό αριθμό σημείων ενδιαφέροντος της δραστηριότητας που αναγνωρίστηκαν στην έρευνα.

Κατηγορία Δραστηριότητας	(Α) Ακρίβεια	(Β) Συνολικά ΡΟΙ
Υπηρεσίες	34%	2057
Φαγητό	83%	832
Καθημερινές αγορές	0%	17
Αγορές	23%	939
Εκπαίδευση	3%	173
Ελεύθερος Χρόνος	49%	727

Πίνακας 3.1: Ακρίβεια αναγνώρισης δραστηριότητας ανά κατηγορία δραστηριότητας της και ο συνολικός αριθμός αναγνωρισθέντων σημείων ενδιαφέροντος [8].

3.4 Μοντέλα βασισμένα σε αλυσίδες Μαρκόβ

Η υπάρχουσα βιβλιογραφία μοντέλων Μαρκόβ επικεντρώνεται σε δύο είδη τεχνικών πρόβλεψης, με τους αντίστοιχους προβλέποντες (predictors):

1. Χωροχρονική πρόβλεψη
2. Πρόβλεψη επόμενης θέσης

Η πρώτη τεχνική αναφέρεται στην πρόβλεψη της θέσης ενός χρήστη σε μια συγκεκριμένη χρονική στιγμή, ενώ η δεύτερη στην πρόβλεψη της αμέσως επόμενης θέσης που θα λάβει ο χρήστης, έχοντας γνώση της τρέχουσας. Η κάθε τεχνική έχει πλεονεκτήματά και μειονεκτήματά, ενώ επίσης η ικανότητα πρόβλεψης των μελλοντικών θέσεων των χρηστών επηρεάζεται από τις συνήθειες και τα ιδιαίτερα χαρακτηριστικά των ατόμων που μελετώνται [38]. Προκύπτει λοιπόν το συμπέρασμα, πως κάποιο μοντέλο πρόβλεψης ενδέχεται να έχει καλή απόδοση για ένα σύνολο χρηστών, ενώ ταυτόχρονα μπορεί να αδυνατεί να προβλέψει με ικανοποιητική ακρίβεια ένα άλλο σύνολο.

Τα μοντέλα που βασίζονται σε αλυσίδες Μαρκόβ μπορούν να λάβουν υπόψιν τους τη συμπεριφορά των χρηστών, υπό την προϋπόθεση πως κατά την ανάλυση και την πρόβλεψη χρησιμοποιούνται κατάλληλοι predictors (χωροχρονικοί ή επόμενης θέσης).

Ιδιότητα Μαρκόβ Στη θεωρία πιθανοτήτων και τη στατιστική, ο όρος Μαρκόβ property [42] αναφέρεται σε μια στοχαστική διαδικασία χωρίς μνήμη. Η στοχαστική αυτή διαδικασία έλαβε το όνομά της από τον Ρώσο μαθηματικό Andrey Markov.

Μια στοχαστική διαδικασία έχει την ιδιότητα Μαρκόβ εάν η υπό όρους κατανομή πιθανοτήτων των μελλοντικών καταστάσεων της διαδικασίας εξαρτάται μόνο από την παρούσα κατάσταση και όχι από την ακολουθία των γεγονότων που προηγήθηκαν. Μία διαδικασία λοιπόν που έχει αυτή την ιδιότητα, ονομάζεται διαδικασία Μαρκόβ. Ο όρος ισχυρή ιδιότητα Μαρκόβ (strong Markov property) είναι παρόμοιος με την ιδιότητα Μαρκόβ, με τη μόνη διαφορά ότι η έννοια του "παρόντος" ορίζεται με όρους μιας τυχαίας μεταβλητής γνωστής και ως "χρόνος διακοπής".

Ένα τυχαίο πεδίο Μαρκόβ ελεγκτείνει αυτήν την ιδιότητα σε δύο ή περισσότερες διαστάσεις ή τυχαίες μεταβλητές που ορίζονται για ένα διασυνδεδεμένο δίκτυο αντικειμένων. Μια στοχαστική διαδικασία διακριτού χρόνου που ικανοποιεί την ιδιότητα Μαρκόβ είναι γνωστή και ως αλυσίδα Μαρκόβ.

Μαθηματική διατύπωση

Έστω (Ω, \mathcal{F}, P) ένας πιθανοχώρος με φιλτράρισμα $(\mathcal{F}_s, s \in I)$ για κάποιο δείκτη I , και έστω (S, \mathcal{S}) ένας μετρήσιμος χώρος. Η (S, \mathcal{S}) στοχαστική διαδικασία $X = \{X_t : \Omega \rightarrow S\}_{t \in I}$ προσαρμοσμένη στο φιλτράρισμα λέγεται πως διαθέτει την ιδιότητα Μαρκόβ αν, για κάθε $A \in \mathcal{S}$ και κάθε $s, t \in I$ με $s < t$ ισχύει:

$$P(X_t \in A | \mathcal{F}_s) = P(X_t \in A | X_s)$$

Ισχυρή ιδιότητα Μαρκόβ

Έστω ότι $X = (X_t : t \geq 0)$ μια στοχαστική διαδικασία σε πιθανοχώρο (Ω, \mathcal{F}, P) με φυσικό φιλτράρισμα $\{\mathcal{F}_t\}_{t \geq 0}$. Τότε, για κάθε διακριτό χρόνο τ στο Ω , μπορούμε να ορίσουμε

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : \forall t \geq 0, \{\tau \leq t\} \cap A \in \mathcal{F}_t\}.$$

Τότε λέμε πως το X διαθέτει ισχυρή ιδιότητα Μαρκόβ. Έστω τ στο διάστημα $\{t < \infty\}$, τότε προκύπτει ότι για κάθε X_τ με $t \geq 0$ ότι το $X_{\tau+t}$ είναι ανεξάρτητο του \mathcal{F}_t

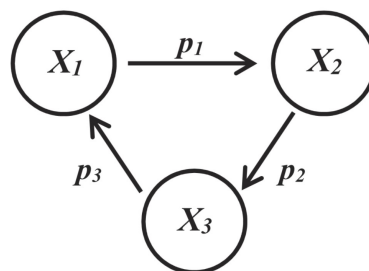
Αλυσίδες Μαρκόβ Οι αλυσίδες Μαρκόβ είναι ένα στοχαστικό μοντέλο που περιγράφει την αλληλουχία πιθανών γεγονότων στα οποία η πιθανότητα πραγματοποίησης κάθε γεγονότος εξαρτάται μόνο από την τιμή που είχε στο αμέσως προηγούμενο στάδιο. Πρόκειται για ένα ευρέως γνωστό μαθηματικό μοντέλο ανάλυσης διαδοχικών δεδομένων που εφαρμόζεται σε πολλούς επιστημονικούς κλάδους, όπως τα οικονομικά και την ανθρώπινη κινητικότητα.

Έστω διαδοχικά δεδομένα μιας ακολουθίας τυχαίων μεταβλητών $X_1, X_2, X_3, \dots, X_n$ [5] με την ιδιότητα Μαρκόβ. Δεδομένης της παρούσας κατάστασης, οι μελλοντικές και προηγούμενες καταστάσεις είναι ανεξάρτητες [43].

Μια αλυσίδα Μαρκόβ μπορεί διατυπωθεί μαθηματικά ως εξής:

$$P_r(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P_r(X_{n+1} = x_{n+1} | X_n = x_n)$$

Όπου οι πιθανές τιμές του X_i για διαμόρφωση ενός συνόλου μετρήσεων S ονομάζεται ο σταθερός χώρος της αλυσίδας. Στη βιβλιογραφία, συχνά η αλυσίδα Μαρκόβ αναπαριστάται από έναν κατευθυνόμενο γράφο, τα άκρα του οποίου επισημαίνονται από την πιθανότητα μετάβασης από μια κατάσταση σε μια άλλη. Το σχήμα 3.3 παρουσιάζει ένα παράδειγμα μιας κλασικής αλυσίδας Μαρκόβ. Στην εικονιζόμενη αλυσίδα Μαρκόβ μπορούν να διακριθούν τρεις καταστάσεις του κατευθυνόμενου γράφου. Η πιθανότητα μετάβασης από την κατάσταση X_1 στη X_2 είναι p_1 , από τη X_2 στη X_3 είναι p_2 , ενώ p_3 ορίζεται η πιθανότητα μετάβασης από την κατάσταση X_3 στη X_1 .



Σχήμα 3.3: Ένα παράδειγμα αλυσίδας Μαρκόβ [5].

Τα μοντέλα Μαρκόβ χρησιμοποιούνται ευρέως σε αλγόριθμους πρόβλεψης, λόγω της αποτελεσματικότητας, της απλότητας και της χαμηλής υπολογιστικής πολυπλοκότητας που τα χαρακτηρίζει [44]. Ένα χαρακτηριστικό μοντέλο πρόβλεψης αλυσίδας Μαρκόβ λαμβάνει υπόψη την τελευταία ακολουθία k τοποθεσιών που

επισκέφθηκε τελευταία ένας χρήστης για να προβλέψει την επόμενη τοποθεσία που θα βρεθεί στο μέλλον. Το μήκος k αυτής της ακολουθίας τοποθεσιών αντιπροσωπεύει την τάξη της αλυσίδας Μαρκόβ και αναφερόμαστε σε αυτό το μοντέλο ως τάξης- k μοντέλο Μαρκόβ (k -order Markov model).

Οι μέθοδοι πρόβλεψης που βασίζονται σε αλυσίδες Μαρκόβ μπορούν να ταξινομηθούν στις παρακάτω τέσσερις κατηγορίες [45]:

1. Μέθοδοι Μαρκόβ τάξης k , που χρησιμοποιούν μόνο το ιστορικό τοποθεσιών για να ανακαλύψουν μεμονωμένα μοτίβα κίνησης [46] [47].
2. Τάξης- k μοντέλα Μαρκόβ που λαμβάνουν επίσης υπόψη εξωτερικές πληροφορίες εκτός από ιστορικές τροχιές κίνησης [48] [49]
3. Υβριδικά μοντέλα Μαρκόβ που συμπληρώνονται από άλλες μεθόδους πρόβλεψης [50] [51]
4. Εξελιγμένοι αλγόριθμοι βασισμένοι σε μοντέλα Μαρκόβ, όπως ο αλγόριθμος LeZi [52] ή αλγόριθμοι που βασίζονται σε κρυφά μοντέλα Μαρκόβ [53] [54].

3.4.1 Τάξης- k μοντέλα Μαρκόβ

Τα κλασικά μοντέλα Μαρκόβ χαμηλής τάξης, συνήθως τάξης 1 ή 2, χρησιμοποιούνται εδώ και τουλάχιστον δεκαπέντε χρόνια για την ανάλυση της ανθρώπινης κινητικότητας. Οι χαμηλής τάξης προβλέψεις των μοντέλων αυτών παρουσίαζαν καλύτερα αποτελέσματα από πιο χρονοβόρους και περίπλοκους αλγόριθμους, όπως των προβλέψεων μερικής ταύτισης (Prediction by Partial Matching, PPM)

Μαρκόβ Predictor Στο [55], ο Μαρκόβ predictor τάξης k (ή $O(k)$) θεμελιώνεται ως εξής: Ο predictor λαμβάνει μια σειρά συμβόλων $(\alpha_1, \alpha_2, \dots, \alpha_n)$ ως συμβολοσειρά ιστορικότητας και επιχειρεί να προβλέψει το επόμενο σύμβολο από τα τρέχοντα συμφραζόμενα. Ως συμφραζόμενα ορίζεται η ακολουθία των k πιο πρόσφατων συμβόλων στο ιστορικό $(\alpha_{n-k_1}, \dots, \alpha_n)$.

Έστω ιστορικό $H = \alpha_1\alpha_2\dots\alpha_n$ και το τμήμα αυτού $H(i, j) = \alpha_1\alpha_{i+1}\dots\alpha_j$ i, j με $1 \leq i \leq j \leq n$. Έστω X μια τυχαία μεταβλητή-σύμβολο και $X(i, j)$ μια συμβολοσειρά $X_iX_{i+1}\dots X_j$ που αναπαριστά μια ακολουθία διακριτών τυχαίων μεταβλητών $X_i, X_{i+1}\dots X_j$ για κάθε

i, j με $1 \leq i \leq j \leq n$. Επίσης, θεωρούμε πλαίσιο $c = H(n - k + 1, n)$ και A το σύνολο όλων των πιθανών συμβόλων.

Εάν το X είναι μια στατική κατανομή τάξης k , τότε για όλα τα $\alpha \in A$ και $i \in \{1, \dots, n - k\}$, η κατανομή του X ικανοποιεί τις κάτωθι εξισώσεις:

$$\begin{aligned} P(X_{n+1} = \alpha | X(1, n) = H) \\ &= P(X_{n+1} = \alpha | X(n - k + 1, n) = c) \\ &= P(X_{i+k+1} = \alpha | X(i + 1, i + k) = c). \end{aligned}$$

Ανά πάσα στιγμή λοιπόν μπορούμε να υπολογίσουμε την πιθανότητα μετάβασης στο α κάνοντας χρήση του τρέχοντος ιστορικού ή και του τρέχοντος περιβάλλοντος (context) c των πιο πρόσφατων k συμβόλων ως:

$$P(X_{n+1} = \alpha | H) \approx P(X_{n+1} = \alpha | c) = \frac{N(c\alpha, H)}{N(c, H)},$$

όπου $N(s', s)$ υποδηλώνει τον αριθμό των φορών που το τμήμα συμβολοσειράς s' εμφανίζεται στη συμβολοσειρά s . Ο Μαρκόβ predictor επιστρέφει το πιο πιθανό επόμενο σύμβολο ως την έξοδο του παρακάτω:

$$X_{n+1} = \operatorname{argmax}_{\alpha \in A} (P(X_{n+1} = \alpha)).$$

Στο [55] προσδιορίζεται επίσης ο $O(k)$ εφεδρικός Μαρκόβ predictor (fallback Markov predictor), ο οποίος ανάγεται σε $O(k - 1)$ Μαρκόβ predictor όταν ο $O(k)$ Μαρκόβ predictor δεν είναι σε θέση να πραγματοποιήσει μια πρόβλεψη. Ο εφεδρικός predictor χρησιμοποιείται όταν το $N(c, H)$ είναι μηδέν, δηλαδή οποτεδήποτε το τρέχον περιβάλλον c δεν έχει εμφανιστεί ποτέ ξανά.

Ορίζεται επίσης, ο predictor "τάξης-0", που επιστρέφει πάντοτε το συχνότερα εμφανιζόμενο σύμβολο στο ιστορικό H . Σε παλιότερή τους εργασία [56], οι Song et al. είχαν διαπιστώσει πως ο $O(2)$ εφεδρικός Μαρκόβ predictor είναι ένας από τους καλύτερους predictors θέσης.

Χρήση predictor Οι Μαρκόβ predictors χρησιμοποιούνται, ανάλογα με το σκοπό της έρευνας για την:

1. Πρόβλεψη διάρκειας μετακίνησης

Οι Μαρκόβ predictors χρησιμοποιούνται για την πρόβλεψη της διάρκειας μιας μετακίνησης. Συνήθως, αυτό συμβαίνει ερευνώντας τα μοτίβα των χρόνων παραμονής σε μια τοποθεσία και όχι της απόλυτης διάρκειας της παραμονής σε αυτήν.

2. Πρόβλεψη χώρου και χρόνου μιας επόμενης κατάστασης

Πολλές εφαρμογές επιδιώκουν την πρόβλεψη της πιθανότητας να βρεθεί κάποιο άτομο σε μια συγκεκριμένη θέση σε μια καθορισμένη χρονική στιγμή. Τα μοντέλα αυτά, αξιοποιούν συχνά Μαρκόβ predictors, αναλύοντας το τρέχον ιστορικό του εκάστοτε χρήστη, προκειμένου να επιτύχουν μια σχετικά ακριβή πρόβλεψη.

Αξιολόγηση αποτελεσμάτων πρόβλεψης Οι προσπάθειες πρόβλεψης μπορεί να καταλήξουν σε τρία ενδεχόμενα:

- Σωστή πρόβλεψη (correct prediction)
- Εσφαλμένη πρόβλεψη (incorrect prediction)
- Αδυναμία πρόβλεψης (prediction miss)

Όταν πραγματοποιηθεί μια πρόβλεψη και αντιστοιχεί στην πραγματική επόμενη τοποθεσία ενός χρήστη, θεωρείται πως έχει γίνει μια σωστή πρόβλεψη (correct prediction). Αντιθέτως, αν προβλεφθεί μια τοποθεσία διαφορετική της πραγματικής επόμενης, παρουσιάζεται μια εσφαλμένη πρόβλεψη (incorrect prediction). Τέλος, εάν ο predictor δεν είναι σε θέση να κάνει μια πρόβλεψη, προκύπτει αδυναμία πρόβλεψης (prediction miss).

Ως **ακρίβεια predictor** (prediction accuracy) ορίζεται ο αριθμός των σωστών προβλέψεων διαιρούμενος με τον αριθμό των συνολικών τοποθεσιών που επισκέφτηκε ένας χρήστης. Ως **απόδοση ενός predictor** θεωρείται η ακρίβεια της πρόβλεψής του. Η **ακρίβεια υπό όρους** ενός predictor (conditional accuracy) είναι ο

αριθμός των σωστών προβλέψεων διαιρούμενος με το συνολικό αριθμό των προβλέψεων που έγιναν. Η υπό όρους ακρίβεια αποκαλύπτει πόσο ακριβής είναι ένας predictor όταν επιτυγχάνει να κάνει μια πρόβλεψη, αφού οι περιπτώσεις αδυναμίας πρόβλεψης παραλείπονται και δεν προσμετρούνται.

3.4.2 Τάξης-k μοντέλα Μαρκόβ με εξωτερικές πληροφορίες

Τα τάξης-k μοντέλα Μαρκόβ με εξωτερικές πληροφορίες αποτελούν ένα υπερσύνολο των τάξης-k μοντέλων Μαρκόβ, στα οποία η πρόβλεψη που πραγματοποιείται ενισχύεται αξιοποιώντας εξωτερικές πληροφορίες. Για να συμβεί αυτό, οι καταστάσεις στα τάξης-k μοντέλα Μαρκόβ εμπλουτίζονται με σχετικές εξωτερικές πληροφορίες. Ωστόσο, αυτό επιβαρύνει την υπολογιστική και χωρική πολυπλοκότητα ανάλυσης των δεδομένων.

Ένα παράδειγμα εφαρμογής των μοντέλων Μαρκόβ τάξης-k με εξωτερικές πληροφορίες για βελτίωση της ακρίβειας πρόβλεψης είναι το [48]. Στη συγκεκριμένη έρευνα, οι Song et al. πρότειναν έναν εφεδρικό Μαρκόβ predictor τάξης-k. Ο predictor αυτός υποβιβάζοταν σε predictor χαμηλότερης τάξης εάν ήταν ανεπιτυχής στην πρόβλεψη της μελλοντικής κατάστασης.

Στις καταστάσεις της έρευνας ενσωματώθηκαν επίσης εξωτερικές πληροφορίες, όπως οι σημειώσεις συμβάντων του Microsoft Outlook και του Ημερολογίου Google. Οι πληροφορίες ημερολογίου παρείχαν επιπλέον πληροφορίες στο μοντέλο Μαρκόβ για τη βελτίωση της ακρίβειας πρόβλεψης. Τα αποτελέσματα της προσομοίωσης που πραγματοποίησαν οι ερευνητές έδειξαν πως ο προτεινόμενος predictor μπορεί να βελτιώσει την πρόβλεψη των παραδοσιακών markov predictors από 3% έως 95%. Η βελτίωση αυτή εξαρτιόταν από το είδος της δραστηριότητας και του μεγέθους μελλοντικής γνώσης. Σε άλλες έρευνες, οι αποστάσεις των διαδρομών, οι χρόνοι ταξιδιού, οι συνήθειες οδήγησης και κοινωνικοί παράγοντες έχουν ενσωματωθεί σε μοντέλα Μαρκόβ τάξης-k με εξωτερικές πληροφορίες για τον υπολογισμό των πιθανών μελλοντικών τοποθεσιών [57, 58].

3.4.3 Κρυφά Μοντέλα Μαρκόβ

Τα **κρυφά μοντέλα Μαρκόβ** (Hidden Markov Models, HMM) είναι μοντέλα Μαρκόβ στα οποία το σύστημα που μοντελοποιείται θεωρείται μια διαδικασία Μαρκόβ, έστω η X , με μη παρατηρήσιμες (κρυφές) καταστάσεις. Τα κρυφά μοντέλα Μαρκόβ υποθέτουν πως υπάρχει μια άλλη διαδικασία, έστω Y , η συμπεριφορά της οποίας εξαρτάται από την X . Ο στόχος λοιπόν μιας κρυφής διαδικασίας Μαρκόβ είναι να αποκτηθεί πληροφορία για τη διαδικασία X , παρατηρώντας τη διαδικασία Y .

Τα κρυφά μοντέλα Μαρκόβ υποθέτουν πως για κάθε χρονική στιγμή n_0 , η εξαρτημένη κατανομή πιθανότητας της n_0 πρέπει να μην εξαρτάται από την ιστορικότητα της X , για κάθε $\{x_n\}_{n < n_0}$.

Ορισμός Έστω X_n και Y_n δυο διακριτού χρόνου στοχαστικές διαδικασίες, όπου $n \geq 1$. Το Ζεύγος (X_n, Y_n) είναι ένα κρυφό μοντέλο Μαρκόβ αν:

- Η X_n είναι μια διαδικασία Μαρκόβ της οποίας η συμπεριφορά δεν είναι άμεσα παρατηρήσιμη (είναι δηλαδή κρυφή)
- $P(Y_n \in A | X_1 = x_1, x_2, \dots, X_n = x_n) = P(Y_n \in A | X_n = x_n)$,
για κάθε $n \geq 1$, x_1, x_2, \dots, x_n και ένα αυθαίρετο σύνολο A

Οι καταστάσεις της διαδικασίας X_n ονομάζονται κρυφές καταστάσεις και η $P(Y_n \in A | X_n = x_n)$ αποκαλείται πιθανότητα μετάδοσης.

Τα κρυφά μοντέλα Μαρκόβ εφαρμόζονται για την προσομοίωση μιας ακολουθίας επισκέψεων ενός ατόμου. Η διαφορά τους με τα κλασικά μοντέλα Μαρκόβ είναι πως οι αλγόριθμοι που αναπτύσσονται μπορούν να συνυπολογίσουν ορισμένα μη παρατηρήσιμα χαρακτηριστικά, δίχως ωστόσο να αγνοούν το ιστορικό τοποθεσίας του ατόμου. Για το λόγο αυτό, τα τελευταία χρόνια, εμφανίζονται στη βιβλιογραφία όλο και περισσότερες τεχνικές που βασίζονται σε κρυφά μοντέλα Μαρκόβ.

3.4.4 Αξιοποίηση των μοντέλων Μαρκόβ στην ανάλυση ανθρώπινης κινητικότητας

Η κίνηση του ανθρώπου στο χώρο μπορεί να αναπαρασταθεί σχετικά αξιόπιστα από μια Μαρκοβιανή στοχαστική διαδικασία. Είναι λοιπόν λογικό να έχουν αναπτυχθεί

αρκετά μοντέλα που την αξιοποιούν για τον υπολογισμό της ανθρώπινης κίνησης και την πρόβλεψη των μελλοντικών της καταστάσεων.

Σε μια έρευνα [45], οι Quiao et al. εκτίμησαν την τάξη του predictor της αλυσίδας Μαρκόβ προσαρμόζοντας το μήκος της στα μοτίβα κινητικότητας του εκάστοτε ατόμου. Έτσι, αντί να χρησιμοποιήσουν έναν στατικό predictor, εξέτασαν την κατανομή χρόνου των μοτίβων κινητικότητας στον υπολογισμό της πιθανότητας μετάβασης στην επόμενη τοποθεσία. Επιπλέον, χρησιμοποίησαν τα αποτελέσματα της πρόβλεψης σε χρήστες με παρόμοιες διαδρομές. Η εφαρμογή των αποτελεσμάτων σε άτομα με παρόμοιες διαδρομές φάνηκε ιδιαίτερα χρήσιμη σε περιπτώσεις καταστάσεων που δεν είχαν καταγραφεί ποτέ ξανά. Η έρευνά τους πραγματοποιήθηκε αξιοποιώντας ένα σύνολο δεδομένων τηλεφωνίας τεχνολογίας Long Term Evolution (LTE) 21 ημερών από περίπου 3500 άτομα. Το υβριδικό μοντέλο που αναπτύχθηκε, διαπίστωσαν πως μπορεί να επιτύχει πρόβλεψη κινητικότητας με ακρίβεια άνω του 56%.

Σε μια άλλη έρευνα, οι Shi et al. [59] συμπεριέλαβαν πλούσια σημασιολογικά δεδομένα, καθώς διαπίστωσαν πως οι υπάρχουσες έρευνες εστίαζαν κυρίως στη μοντελοποίηση χωρικών και χρονικών προτύπων της ανθρώπινης κινητικότητας χωρίς να λαμβάνουν υπόψιν τους άλλα εξωτερικά δεδομένα. Τα δεδομένα που χρησιμοποίησαν αντλήθηκαν από διάφορα κοινωνικά δίκτυα.

Με την προσθήκη της σημασιολογίας κατάφεραν να αποτυπώσουν το κίνητρο των ανθρώπων πίσω από την μετακίνηση. Πρότειναν λοιπόν ένα νέο σημασιολογικό γνωσιακό μοντέλο κινητικότητας που αποτυπώνει τα κίνητρα της ανθρώπινης κινητικότητας, χρησιμοποιώντας μεγάλης κλίμακας και πλούσια σημασιολογικά χωροχρονικά δεδομένα.

Δημιούργησαν μια μέθοδο πολλαπλής ενσωμάτωσης για να συνδυάσουν ετερογενή δεδομένα, όπως το αναγνωριστικό χρήστη, την τοποθεσία, την ώρα και τη δραστηριότητα στον ίδιο χώρο ενσωμάτωσης. Ο χώρος ενσωμάτωσης ήταν βασισμένος στη συχνότητα επανεμφάνισης και δημιουργήθηκε με έναν μη εποπτευόμενο τρόπο, φροντίζοντας ωστόσο να διατηρηθούν τα αρχικά σημασιολογικά δεδομένα της διαδρομής που ακολούθησαν οι χρήστες. Επίσης, το μοντέλο αυτό διέθετε έναν εύκολο τρόπο σύγκρισης δύο καταστάσεων, καθώς αξιοποιούσε μια τυπική μονάδα μέτρησης απόστασης (cosine similarity). Αφού

ενσωματώθηκαν όλες οι διαδρομές στο προταθέν μοντέλο, εφαρμόστηκε ένα κρυφό μοντέλο Μαρκόβ για τον εντοπισμό κρυφών καταστάσεων στις μετακινήσεις στο χώρο ενσωμάτωσης. Οι μετακινήσεις των χρηστών εξετάστηκαν όχι μόνο ως προς το χρόνο και το χώρο, αλλά και ως προς τα κίνητρα του χρήστη.

Η προτεινόμενη μέθοδος αξιολογήθηκε σε δύο μεγάλης κλίμακας σύνολα δεδομένων, όπου επικυρώθηκε η ικανότητά της να παράξει μοντέλα κινητικότητας υψηλής ποιότητας και ακρίβειας. Πραγματοποιήθηκαν επιπροσθέτως εκτενή πειράματα, τα αποτελέσματα των οποίων έδειξαν πως το προταθέν μοντέλο ξεπερνά αρκετά σύγχρονα μοντέλα κινητικότητας, προσδίδοντας μεγαλύτερη ακρίβεια πρόβλεψης και σημαντικά μεγαλύτερη απόδοση.

3.5 Νευρωνικά μοντέλα

Ένα νευρωνικό δίκτυο είναι ένα δίκτυο ή κύκλωμα νευρώνων που αποτελείται από τεχνητούς νευρώνες ή κόμβους. Στις μέρες μας, ένα τέτοιο νευρωνικό δίκτυο αποκαλείται τεχνητό νευρωνικό δίκτυο. Τα νευρωνικά δίκτυα εφαρμόζονται συχνά σε ευφυή συστήματα (Artificial Intelligence) και στη μηχανική μάθηση. Οι συνδέσεις ενός νευρωνικού δικτύου μοντελοποιούνται ως βάρη. Ένα θετικό βάρος αντιπροσωπεύει μια διεγερτική σύνδεση, ενώ οι αρνητικές τιμές αναπαριστούν ανασταλτικές συνδέσεις.

Τα νευρωνικά δίκτυα επίσης χρησιμοποιούνται σε προγνωστικά μοντέλα, σε συστήματα αυτόματου ελέγχου και σε εφαρμογές όπου μπορούν να εκπαιδευτούν (training) μέσω ενός συνόλου δεδομένων. Η αυτομάθηση (self learning) που προκύπτει από την εκπαίδευση μέσω ενός συνόλου δεδομένων μπορεί να πραγματοποιηθεί με τη μορφή νευρωνικών δικτύων. Με τη σειρά τους, τα νευρωνικά δίκτυα αξιοποιούνται στην εξαγωγή συμπερασμάτων από πολύπλοκα και φαινομενικά άσχετα σύνολα πληροφοριών.

3.5.1 Νευρωνικά μοντέλα με ανάδραση

Ένα νευρωνικό δίκτυο με ανάδραση (Recurrent Neural Network, **RNN**) είναι μια κατηγορία νευρωνικού δικτύου που αποσκοπεί στην επεξεργασία διαδοχικών δεδομένων [6]. Τα νευρωνικά δίκτυα με ανάδραση είναι νευρωνικά δίκτυα

εξειδικευμένα στην επεξεργασία μιας σειράς τιμών $x^{(1)}, x^{(2)}, \dots, x^{(\tau)}$ και μπορούν να κλιμακωθούν σε πολύ μεγαλύτερες ακολουθίες απ' ό,τι θα ήταν πρακτικό για δίκτυα μη εξειδικευμένα σε ακολουθιακές εισόδους. Τα νευρωνικά δίκτυα με ανάδραση μπορούν επίσης να αξιοποιήσουν την εσωτερική τους κατάσταση (μνήμη) για να επεξεργαστούν ακολουθίες εισόδου μεταβλητού μήκους. Έχουν ως στόχο να ανακαλύψουν μια συνάρτηση που εκφράζει τον τρόπο μετάβασης από μια κατάσταση στην επόμενη. Το γεγονός αυτό τα καθιστά ικανά να επεξεργαστούν δεδομένα που παρουσιάζουν κάποια χρονική αλληλεξάρτηση.

Η μακροχρόνια βραχυπρόθεσμη μνήμη (Long short-term memory, LSTM) και η περιφραγμένη μονάδα με ανάδραση (Gated recurrent unit, GRU) χρησιμοποιούνται ευρέως στα νευρωνικά μοντέλα με ανάδραση. Η μακροχρόνια βραχυπρόθεσμη μνήμη LSTM αποτελείται από μια κατάσταση μονής κυψέλης (onecell) και από τρεις ελεγχόμενες πύλες για να διατηρήσει και ενημερώσει την κατάσταση της κυψέλης. Με βάση την κατάσταση εισόδου και της τελευταίας κυψέλης, η μακροχρόνια βραχυπρόθεσμη μνήμη LSTM ενημερώνει πρώτα την κατάσταση του κελιού με τα μέρη που πρέπει να διατηρηθούν και τα μέρη που απορρίπτονται. Στη συνέχεια, η μακροχρόνια βραχυπρόθεσμη μνήμη δημιουργεί την έξοδο από την κατάσταση του κελιού με μεταβλητό βάρος. Η περιφραγμένη μονάδα με ανάδραση (GRU) είναι μια δημοφιλής παραλλαγή του LSTM που αντικαθιστά την πύλη ξεκλειδώματος και την πύλη εισόδου με μία μόνο πύλη ενημέρωσης. Οι εκδοχές διαμόρφωσης της ενημέρωσης της περιφραγμένης μονάδας με ανάδραση έχουν ως κάτωθι [60],

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f),$$

$$r_t = \sigma(W_{rx}x_t + W_{rh}h_{t-1} + b_r),$$

$$c_t = \tanh(W_{cx}x_t + r_t * (W_{ch}h_{t-1} + b_c)),$$

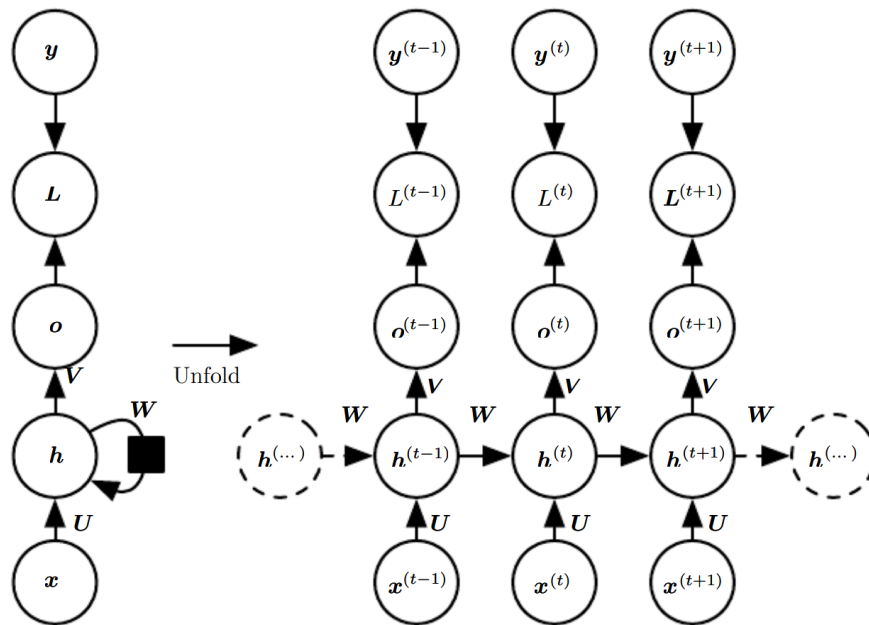
$$h_t = (1 - f_t) * c_t + f_t * h_{t-1},$$

όπου x_t είναι η είσοδος για κάποιο χρόνο t , h_{t-1} η τελευταία έξοδος της περιφραγμένης μονάδας με ανάδραση, f_t το βάρος ενημέρωσης, r_t οι πύλες επανεκκίνησης, c_t ο υποψήφιος και h_t το αποτέλεσμα εξόδου. Η περιφραγμένη

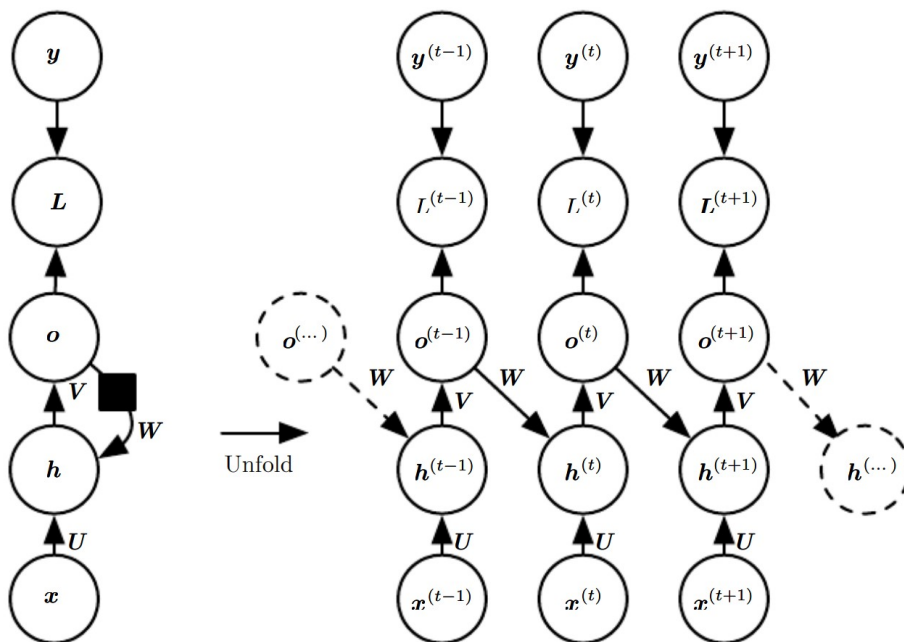
μονάδα με ανάδραση επιτυγχάνει την ίδια απόδοση σε πολλαπλές διεργασίες, ενώ παρουσιάζει σχετικά μικρή υπολογιστική πολυπλοκότητα.

Τα νευρωνικά δίκτυα με ανάδραση επαναπροωθούν τα αποτελέσματά τους στην είσοδό τους, πράγμα που αποτελεί την ειδοποίησή τους διαφορά από τα απλά προωθητικά δίκτυα. Τα πιο συνηθισμένα είδη σχημάτων RNN είναι τα ακόλουθα:

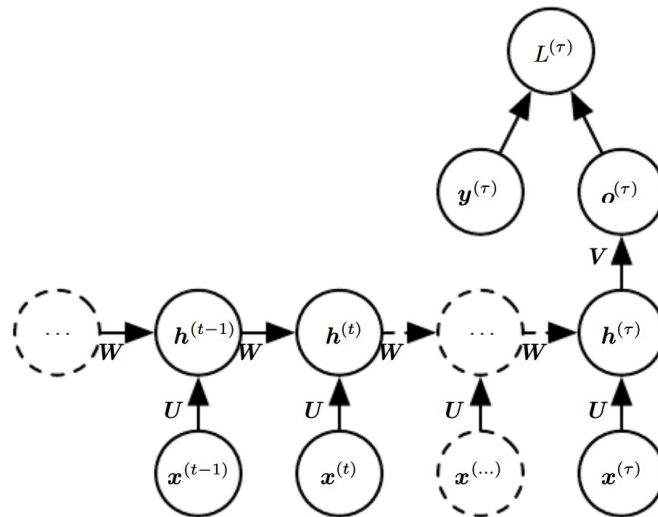
1. RNN που παράγουν έξοδο σε κάθε βήμα και έχουν ανατροφοδοτούμενες συνδέσεις μεταξύ νευρώνων κρυφού επιπέδου (βλ. σχήμα 3.4). Στο σχήμα, το δίκτυο διαθέτει δεδομένα εκπαίδευσης, ετικέτες $y(t)$ και εξόδους $o(t)$ για κάθε t . Το δεξί τμήμα του σχήματος αναπαριστά το χρονο-αναδιπλωμένο υπολογιστικό γράφημα, στο οποίο κάθε κόμβος συνδέεται με μια συγκεκριμένη χρονική στιγμή.
2. RNN που παράγουν έξοδο σε κάθε βήμα και έχουν ανατροφοδοτούμενες συνδέσεις μόνο από την έξοδο σε κάθε βήμα προς τους νευρώνες κρυφού επιπέδου στο επόμενο βήμα (βλ. σχήμα 3.5). Στο σχήμα, σε κάθε βήμα t , η είσοδος είναι x_t , οι ενεργοποιήσεις του κρυφού επιπέδου είναι $h^{(t)}$, οι εξόδοι $o^{(t)}$, οι στόχοι $y^{(t)}$ και η απώλεια $L^{(t)}$. Το RNN του σχήματος είναι εκπαιδευμένο να τοποθετεί μια συγκεκριμένη τιμή εξόδου στο o , και το o είναι η μόνη πληροφορία που επιτρέπεται να στείλει στις μελλοντικές καταστάσεις. Το δεξί τμήμα του σχήματος αναπαριστά τους νευρώνες του επιπέδου h , οι οποίοι ανατροφοδοτούνται με την διέλευση του χρόνου.
3. RNN με ανατροφοδοτούμενες συνδέσεις μεταξύ κρυφών τμημάτων που διαβάζουν μια ολόκληρη ακολουθία και παράγουν μια μόνο έξοδο (βλ. σχήμα 3.6).



Σχήμα 3.4: RNN με συνδέσεις ανατροφοδότησης των κρυφών μονάδων [6].



Σχήμα 3.5: RNN με ανάδραση από την έξοδο (output layer) προς τους νευρώνες κάποιου κρυφού επιπέδου (hidden units). [6].



Σχήμα 3.6: Χρονικά αναδιπλωμένο νευρωνικό δίκτυο με ανατροφοδότηση των κρυφών μονάδων στον εαυτό τους και με μια έξοδο στο τέλος της ακολουθίας. Το δίκτυο αυτό μπορεί να χρησιμοποιηθεί για να συνοψιστεί και να χαρακτηριστεί μια ακολουθία δεδομένων [6].

3.5.2 Αξιοποίηση των νευρωνικών μοντέλων στην ανάλυση της ανθρώπινης κινητικότητας

Η πρόβλεψη της ανθρώπινης κινητικότητας έχει μεγάλη σημασία για ένα ευρύ φάσμα εφαρμογών. Ωστόσο, η πρόβλεψη της δεν είναι απλοϊκή, λόγω της ύπαρξης τριών προκλήσεων που προκύπτουν κατά τον υπολογισμό της:

1. Συχνά παρουσιάζονται σύνθετες μεταβατικές καταστάσεις που απαρτίζονται από κάποια χρονοεξαρτώμενη και υψηλής τάξης περιοδικότητα.
2. Η πολλαπλών επιπέδων περιοδικότητα της ανθρώπινης κινητικότητας
3. Η ετερογένεια και η αραιότητα των συλλεγόμενων δεδομένων.

Τα παραδοσιακά εργαλεία ανάλυσης ανθρώπινης κινητικότητας δεν ήταν αρκετά για την αντιμετώπιση των προκλήσεων που προέκυπταν, αλλά και του όγκου των δεδομένων και των πολύπλοκων κανόνων που τη χαρακτηρίζουν. Ωστόσο, η αυξημένη υπολογιστική δύναμη των υπολογιστών σε συνδυασμό με την ανάπτυξη ευφυών αλγορίθμων μηχανικής μάθησης έδωσε νέα εργαλεία στους ερευνητές. Τα

τελευταία χρόνια, στη βιβλιογραφία εμφανίζονται συχνά μέθοδοι πρόβλεψης της ανθρώπινης κινητικότητας βασισμένες σε νευρωνικά δίκτυα.

Σε μια έρευνα, οι Feng et al. [60] εφάρμοσαν ένα attentional νευρωνικό δίκτυο με ανάδραση και πρότειναν το DeepMove. Το μοντέλο αυτό ήταν ικανό να εξάγει συμπεράσματα κινητικότητας δεχόμενο ως είσοδο δεδομένα μεγάλων αποστάσεων, τα οποία επίσης ήταν αραιά καταγεγραμμένα στο χρόνο. Σχεδίασαν λοιπόν ένα νευρωνικό δίκτυο πολλαπλών επιπέδων με ανάδραση για να εντοπίσουν τις περίπλοκες κινήσεις των ανθρώπων και να ανακαλύψουν τους παράγοντες που διέπουν την ανθρώπινη κινητικότητα. Έτσι, πρότειναν ένα attentional μοντέλο που λαμβάνει υπόψιν του το ιστορικό κινήσεων, που διαθέτετε δύο μηχανισμούς προκειμένου να πραγματοποιηθεί η καταγραφή της περιοδικότητας πολλαπλών επιπέδων.

Οι ερευνητές δοκίμασαν το μοντέλο τους στην πράξη σε τρία πραγματικά αντιπροσωπευτικά σύνολα δεδομένων κινητικότητας. Τα αποτελέσματα της αξιολόγησης έδειξαν πως το μοντέλο τους παρουσίασε καλύτερα αποτελέσματα συγκριτικά με άλλα σύγχρονα μοντέλα. Μάλιστα, τα αποτελέσματά της έρευνας παρουσίαζαν συμπεράσματα κατά τουλάχιστον 10% βελτιωμένα, σε σχέση με τα συμπεράσματα άλλων διαθέσιμων μοντέλων.

Σε μια άλλη έρευνα, οι Zion et al [61] εκμεταλλεύτηκαν την ιδιότητα της επαναληψιμότητας των καθημερινών διαδρομών στις κινήσεις των ανθρώπων. Τις καθημερινές διαδρομές που αναγνώρισαν τις αποκάλεσαν "μοτίβα κινητικότητας", και τις αξιοποίησαν για να εκπαιδεύσουν νευρωνικά μοντέλα εντοπισμού και πρόβλεψης του τρόπου ζωής (lifestyle) ενός ατόμου.

Για την πειραματική εφαρμογή του μοντέλου τους, αξιοποίησαν δεδομένα τηλεφωνίας ενός συνόλου ανθρώπων και αντιπροσώπευσαν την καθημερινή τους πορεία χρησιμοποιώντας σημασιολογικές ετικέτες (όπως σπίτι, εργασία, γυμναστήριο), που έδωσαν στα κύρια σημεία ενδιαφέροντος (Points of Interest, POI) των ατόμων. Τα σημεία ενδιαφέροντος καθορίστηκαν συλλογικά, πραγματοποιώντας συνεντεύξεις σε όλα τα άτομα που συμμετείχαν στην έρευνα.

Χρησιμοποιώντας ένα σύστημα νευρωνικών δικτύων, κατόρθωσαν

- Να αναγνωρίσουν τις καθημερινές διαδρομές στα πλαίσια του κοινωνικού τρόπου ζωής των ανθρώπων

- Να εντοπίσουν μοτίβα κινητικότητας στις τροχιές που αναγνώρισαν
- Να εκπαιδεύσουν ένα μοντέλο για την πρόβλεψη του τρόπου ζωής (lifestyle) των ανθρώπων, που βασίζεται στην αναπαράσταση ενός μοτίβου κινητικότητας.

Κάνοντας χρήση του νευρωνικού δικτύου και του αναπτυχθέντος αλγορίθμου, οι ερευνητές ήταν σε θέση να προβλέψουν μια ετικέτα τρόπου ζωής για μια τροχιά που δεν είχαν παρατηρήσει προηγουμένως. Το προταθέν σύστημα εμφάνισε ακρίβεια και επεκτασιμότητα στην ταυτοποίηση και την πρόβλεψη του τρόπου ζωής, τόσο για μια νέα τροχιά, όσο και για έναν νέο χρήστη. Παράλληλα, τα αποτελέσματα που προέκυψαν κρίθηκαν από τους ερευνητές καλύτερα από εκείνα των αλγορίθμων τελευταίας τεχνολογίας.

3.6 Συγκριτική παρουσίαση μεθοδολογιών, σύνοψη

Στην ενότητα αυτή, σημειώνεται μια συνοπτική ταξινόμηση των μοντέλων ανθρώπινης κινητικότητας που εμφανίζονται συχνά στη βιβλιογραφία. Η μέθοδος ταξινόμησης των μοντέλων σε λίστα βασίζεται στη μελέτη και την κατανόηση των ομοιοτήτων και των διαφορών μεταξύ των πολυάριθμων μοντέλων και των πεδίων στα οποία εμφανίζουν κοινά στοιχεία ή διαφέρουν μεταξύ τους.

Μια διάκριση των τεχνικών ανάλυσης ανθρώπινης κινητικότητας, είναι ο σκοπός για τον οποίο αναπτύχθηκαν. Τα υπάρχοντα μοντέλα σε μικρό ή μεγαλύτερο βαθμό, διαφέρουν. Για παράδειγμα, επιστήμονες της ανάλυσης δεδομένων αναπτύσσουν μοντέλα εντελώς διαφορετικά από τα μοντέλα των συγκοινωνιολόγων. Επίσης, ορισμένα μοντέλα εξετάζονται αναλύοντας δεδομένα έξυπνων συσκευών, όπως για παράδειγμα τα γεωγραφικά δεδομένα ή οι καταγραφές λεπτομερειών κλήσεων. Παράλληλα, άλλα εξετάζονται τοποθετώντας αντίστοιχες συσκευές σε οχήματα που κινούνται στον αστικό ιστό, όπως λόγου χάρη στα ταξί και τα λεωφορεία. Συνήθως όμως, οι κύριες κατηγορίες μοντέλων διαθέτουν παραδείγματα εφαρμογής με τις περισσότερες πηγές δεδομένων.

Διάκριση μοντέλων Η διάκριση των μοντέλων η οποία επιλέχθηκε στην παρούσα διπλωματική εργασία, αφορά το κύριο πεδίο στο οποίο εφαρμόζονται. Συνοψίζοντας

και ακολουθώντας τη δομή της εργασίας, τα μοντέλα που αναλύθηκαν ήταν τα τυχαία κίνησης (Μπράουν, Λέβι), τα βαρυτικά, τα rule-based, τα Μαρκόβ και τα Νευρωνικά.

Τα μοντέλα ανθρώπινης κίνησης, ως προς τον σκοπό τους, μπορούν να διακριθούν σε δύο κατηγορίες:

- Μοντέλα προσομοίωσης ανθρώπινης κίνησης (βλ. πίνακα 3.2)

Τα μοντέλα αυτά επιδιώκουν την αναπαράσταση της ανθρώπινης κίνησης εφαρμόζοντας μαθηματικά μοντέλα. Στόχος τους λοιπόν είναι να ανακαλύψουν τους κανόνες πίσω από την κίνηση των ανθρώπων και των πληθυσμών, να προβλέψουν τις μελλοντικές τους κινήσεις καθώς και να τις αναπαραστήσουν σε πίνακες προέλευσης προορισμού. Τα μοντέλα που ανήκουν στην κατηγορία αυτή είναι τα βασισμένα σε κίνηση Μπράουν, Levy Flights, τα Μαρκόβ και τα βαρυτικά.

Μοντέλα προσομοίωσης ανθρώπινης κίνησης		Πηγές δεδομένων στη βιβλιογραφία						Παραδείγμ. εφαρμογής
Όνομα Μοντέλου	Πεδίο εφαρμογής	Έρευνες	CDR	GPS	3G/4G	POI	Κοιν. Δίκτυα	
Κίνηση Μπράουν, Levy Flights	Ατομική κίνηση	Ναι	Ναι	Ναι				[22, 40, 62]
Μαρκόβ	Ατομική κίνηση		Ναι	Ναι	Ναι	Ναι	Ναι	[63, 64]
Βαρυτικό	Κινήσεις πληθυσμών		Ναι	Ναι	Ναι	Ναι	Ναι	[65-68]

Πίνακας 3.2: Συγκριτικός πίνακας μοντέλων προσομοίωσης ανθρώπινης κίνησης.

- Μοντέλα ανάλυσης ανθρώπινης κίνησης (βλ. πίνακα 3.3)

Τα μοντέλα αυτά αποσκοπούν στην εξαγωγή βασικών μονάδων μέτρησης ανθρώπινης κινητικότητας και την συλλογή πληροφοριών από σύνολα δεδομένων. Τα μοντέλα αυτά χαρακτηρίζονται από ποικιλομορφία ως προς τα αξιοποιούμενα σύνολα δεδομένων. Τα μοντέλα που ανήκουν στην κατηγορία αυτή είναι τα Rule-based, τα Μαρκόβ και τα Νευρωνικά.

Μοντέλα ανάλυσης ανθρώπινης κίνησης		Πηγές δεδομένων στη βιβλιογραφία						Παραδείγμ. εφαρμογής
Όνομα μοντέλου	Πεδίο εφαρμογής	CDR	GPS	3G/4G	POI	Κοιν. Δίκτυα		
Rule-based	Αναγνώριση σημείων, μοτίβων και περιοχών	Ναι	Ναι		Ναι	Ναι	[16, 21, 69]	
Μαρκόβ	Αναγνώριση σημείων	Ναι	Ναι	Ναι	Ναι	Ναι	[38, 45, 59, 70]	
Νευρωνικά	Ανάλυση κίνησης, αναγνώριση σημείων	Ναι	Ναι		Ναι	Ναι	[60, 71]	

Πίνακας 3.3: Συγκριτικός πίνακας μοντέλων ανάλυσης ανθρώπινης κίνησης.

Βάσει των πινάκων 3.3 και 3.2, βλέπουμε πως τα εκάστοτε μοντέλα χαρακτηρίζονται από κάποιο κύριο πεδίο εφαρμογής. Είναι λοιπόν εμφανές πως τα

μοντέλα ανθρώπινης κινητικότητας μπορούν να διακριθούν ως προς το πεδίο εφαρμογής στις εξής τέσσερις κατηγορίες:

1. Μοντέλα προσομοίωσης ατομικής κίνησης
2. Μοντέλα προσομοίωσης κινήσεων πληθυσμών
3. Μοντέλα αναγνώρισης σημείων
4. Μοντέλα αναγνώρισης περιοχών

Μοντέλα προσομοίωσης ατομικής κίνησης Τα μοντέλα αυτά αποσκοπούν στην ανάλυση της κίνησης των ανθρώπων ως ξεχωριστές οντότητες και όχι ως μέλη μεγαλύτερων ομάδων πληθυσμού. Τα μοντέλα προσομοίωσης ατομικής κίνησης, κατεξοχήν τα Μαρκόβ και παλαιότερα τα βασισμένα σε κινήσεις Μπράουν και Λέβι flights, επιδιώκουν την εξαγωγή συμπερασμάτων κίνησης των ανθρώπων και την πρόβλεψη των μελλοντικών της θέσεων.

Προκειμένου να εξαχθούν συμπεράσματα, αξιοποιούνται στη βιβλιογραφία οι περισσότερες πηγές δεδομένων που έχουν καταγραφεί στην παρούσα εργασία. Για παράδειγμα, στην έρευνά τους οι Quiao et al. [45] εφάρμοσαν ένα μοντέλο Μαρκόβ τάξης-k προκειμένου να αναλύσουν την ατομική κίνηση, αξιοποιώντας δεδομένα **LTE**. Σε μια άλλη έρευνα, οι Lv et al. εφάρμοσαν ένα κρυφό μοντέλο Μαρκόβ που ανέλυε **δεδομένα τηλεφωνίας** για την εξαγωγή συμπερασμάτων ατομικής κίνησης και πρόβλεψης μελλοντικών της καταστάσεων. Ένα ακόμη παράδειγμα χρησιμοποιηθέντων δεδομένων είναι η έρευνα των Rhee et al. [22], στην οποία εφαρμόστηκε ένα μοντέλο περιπάτων Λέβι σε δεδομένα **CDR** προκειμένου να προσομοιωθεί με μεγαλύτερη ακρίβεια η ανθρώπινη κίνηση.

Μοντέλα προσομοίωσης κινήσεων πληθυσμών Οι τεχνικές προσομοίωσης κινήσεων πληθυσμού αξιοποιούν χωροχρονικά δεδομένα που κατεξοχήν αποτελούνται από πίνακες προέλευσης-προορισμού (origin-destination matrices). Οι πίνακες προέλευσης-προορισμού συνήθως διαθέτουν πληροοφορίες ως προς τον όγκο των ατόμων που μετακινήθηκαν, την προέλευση, τον προορισμό τους καθώς και τη χρονική στιγμή η οποία αντιπροσωπεύει τη μετακίνηση αυτή. Αντί λοιπόν να

υπολογίζεται η μετακίνηση μεταξύ δυο ή περισσότερων διαδοχικών θέσεων, το πρόβλημα αντιμετωπίζεται αναλύοντας τις μετακινήσεις μεταξύ περιοχών.

Στη βιβλιογραφία, τα βαρυτικά μοντέλα εφαρμόζονται συχνά για την προσομοίωση κινήσεων ομάδων πληθυσμού. Αντίστοιχα με την προηγούμενη κατηγορία, και στην περίπτωση αυτή διαπιστώνεται αξιοποίηση των περισσότερων πηγών δεδομένων που έχουν καταγραφεί στην παρούσα εργασία. Σε μια πρόσφατη έρευνα, οι Ciavarella et al. [65] εφάρμοσαν ένα βαρυτικό μοντέλο αξιοποιώντας δεδομένα λεπτομερειών καταγραφής κλήσεων (CDRs) προκειμένου να εξάγουν συμπεράσματα σε ειδικά διαμορφωμένα χωρικά δεδομένα κίνησης. Παραμετροποιώντας και μεταβάλλοντας το μέγεθος των περιοχών, κατάφεραν να διερευνήσουν την επιρροή του μεγέθους μιας περιοχής στην αξιοπιστία της ανάλυσης.

Μοντέλα αναγνώρισης σημείων Τα μοντέλα αναγνώρισης σημείων αποσκοπούν στην επιτυχή εξαγωγή συμπερασμάτων ως προς το είδος των σημείων που σταθεμούν ή διέρχονται οι χρήστες. Τα μοντέλα αυτά, συνήθως επιστρατεύουν αλγόριθμους βασισμένους σε κανόνες, μοντέλα Μαρκόβ και Νευρωνικά δίκτυα.

Η έρευνα των Andrienko et al. [35] αποτελεί ένα παράδειγμα εφαρμογής rule-based αλγορίθμων για το χαρακτηρισμό περιοχών και δραστηριοτήτων των χρηστών κατά τη διάρκεια της ημέρας. Ο ερευνητής, εξέτασαν δεδομένα κινητής τηλεφωνίας και κοινωνικών δικτύων προκειμένου να ανιχνεύσουν σημεία τα οποία οι χρήστες επισκέπτονταν συχνά ή περιοδικά. Κατάφεραν μάλιστα, να προσαρμόσουν την λύση τους ώστε να διαφυλάξουν τα ψηφιακά ίχνη και πνευματικά δικαιώματα των χρηστών.

Μοντέλα αναγνώρισης περιοχών Οι τεχνικές αναγνώρισης περιοχών αποσκοπούν στην εξαγωγή συμπερασμάτων ως προς το είδος μιας περιοχής και του σκοπού που αυτή επιτελεί στη λειτουργία μιας πόλης ή μιας περιοχής. Κατά κύριο λόγο, τα rule-based μοντέλα είναι αυτά που χρησιμοποιούνται για το σκοπό αυτό.

Ένα χαρακτηριστικό παράδειγμα μοντέλου αναγνώρισης περιοχών είναι η έρευνα των Yuan et al. [16], όπου οι ερευνητές παρουσίασαν έναν αλγόριθμο αναγνώρισης περιοχών διαφορετικών λειτουργιών σε μια πόλη (Framework that Discovers Regions

of different Functions in a City, DRoF). Στην έρευνά τους, μια πόλη διατομείται σε περιοχές, οι οποίες οριοθετούνται από το δίκτυο μεγάλων δρόμων και λεωφόρων. Εφαρμόζουν έναν rule-based αλγόριθμο προκειμένου να συμπεράνουν εμμέσως τα ιδιαίτερα στοιχεία μιας περιοχής. Προκειμένου να το επιτευχθεί η εξαγωγή συμπερασμάτων, προσμετρώνται στοιχεία όπως το είδος των σημείων ενδιαφέροντος που περιλαμβάνονται στην περιοχή και ο όγκος και ο χρόνος που εισέρχονται ή εξέρχονται οι άνθρωποι σε αυτή. Καταλήγοντας, η περιοχή κατατάσσεται βάσει του χαρακτηριστικού που την αντιπροσωπεύει περισσότερο, δηλαδή αυτού του οποίου συγκεντρώνει την μεγαλύτερη βαθμολογία σε σχέση με τα υπόλοιπα χαρακτηριστικά. Για την πειραματική επιβεβαίωση του μοντέλου, αξιοποιήθηκαν δεδομένα GPS και POI του Πεκίνου, μεταξύ 2010 και 2011.

4. Πειραματική Εφαρμογή

Μεθοδολογιών Σε Εμπειρικά Δεδομένα

Στην ενότητα αυτή θα εξεταστούν δυο μοντέλα ανθρώπινης κινητικότητας. Το πρώτο μοντέλο, αφορά δεδομένα ατομικής κινητικότητας, ενώ το δεύτερο αφορά δεδομένα κινήσεων πληθυσμών. Προτού όμως αναφερθεί η πειραματική υλοποίηση των μοντέλων κινητικότητας, θα γίνει μια σύντομη καταγραφή της γλώσσας προγραμματισμού και των βιβλιοθηκών της που αξιοποιήθηκαν για τους σκοπούς της διπλωματικής εργασίας.

Το σύνολο του κώδικα που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής εργασίας μπορεί να βρεθεί στην ακόλουθη ηλεκτρονική διεύθυνση: https://github.com/kapelios/human_mobility.

4.1 Γλώσσα προγραμματισμού και βιβλιοθήκες

Για την ανάλυση της ανθρώπινης κινητικότητας αξιοποιήθηκε η γλώσσα προγραμματισμού Python [72], και συγκεκριμένα η version 3.7. Η Python είναι μια γλώσσα προγραμματισμού υψηλού επιπέδου γενικού σκοπού, και χρησιμοποιείται ευρέως στην επιστήμη της ανάλυσης δεδομένων.

Η βιβλιοθήκη pandas Η βιβλιοθήκη της python "**pandas**" [73], είναι μια βιβλιοθήκη της Python εξειδικευμένη στη διαχείριση και ανάλυση δεδομένων. Διαθέτει δομές δεδομένων και λειτουργίες για τη διαχείριση αριθμητικών πινάκων και χρονοσειρών. Τα βασικά χαρακτηριστικά που την απαρτίζουν είναι τα ακόλουθα:

- Υποστήριξη των αντικειμένων DataFrame, για τη διαχείριση δεδομένων. Τα αντικείμενα αυτά διαθέτουν ενσωματωμένα ευρετήρια.
- Εργαλεία ανάγνωσης και εξαγωγής δεδομένων διαφόρων τύπων συνόλων δεδομένων και αρχείων.
- Διαχείριση δεδομένων που λείπουν, αυτόματη λύση συμπλήρωσης κενών.

- Συναρτήσεις αναδιαμόρφωσης και αναδίπλωσης συνόλων δεδομένων.
- Εύκολη και απλή εξαγωγή υποσυνόλων συνόλων δεδομένων.
- Προσθαφαίρεση στηλών σε όλες τις δεδομένων.
- Υποστήριξη εντολών τύπου SQL, όπως ομαδοποίηση ανά τύπο, ταξινόμηση κ.α.
- Η συγχώνευση και η ένωση συνόλων δεδομένων επιτυγχάνεται με απλό τρόπο.
- Φιλτράρισμα δεδομένων βάσει κριτηρίων.

Η βιβλιοθήκη Geopandas Το GeoPandas [74] είναι ένα έργο ανοιχτού κώδικα που διευκολύνει την εργασία με χωροχρονικά δεδομένα στην Python. Το GeoPandas επεκτείνει τους τύπους δεδομένων της βιβλιοθήκης pandas, παρέχοντας έτσι περισσότερες συναρτήσεις και δυνατότητες επεξεργασίας των χωρικών δεδομένων. Η υποστήριξη των γεωμετρικών λειτουργιών οφείλεται στο shapely, το οποίο αποτελεί ένα πακέτο της Python για διαχείριση και ανάλυση επίπεδων γεωμετρικών αντικείμενων. Επίσης, το Geopandas αξιοποιεί τη βιβλιοθήκη matplotlib για την εξαγωγή σχεδίων.

Η πρακτική αξία του Geopandas είναι πως επιτρέπει την εύκολη επεξεργασία χωροχρονικών δεδομένων στην Python, που υπό άλλες συνθήκες θα απαιτούσαν ειδικά εργαλεία ανάλυσης χωροχρονικών δεδομένων (όπως λόγω χάρη το PostGIS).

Η βιβλιοθήκη folium Η βιβλιοθήκη folium [75] διευκολύνει την οπτικοποίηση χωροχρονικών δεδομένων που έχουν υποστεί επεξεργασία στην Python, όπως λόγω χάρη τα **pandas datasets**. Παρέχει, λοιπόν, εργαλεία δημιουργίας διαδραστικών χαρτών. Επιτρέπει επίσης, την επισήμανση δεδομένων σε έναν χάρτη καθώς και την επισύναψη διανυσμάτων, raster ή HTML δεικτών.

Η βιβλιοθήκη διαθέτει αρκετές διεπαφές και ενσωματώσεις με μεγάλες πλατφόρμες χαρτών. Μερικές εξ αυτών είναι το OpenStreetMap, το Mapbox και το Stamen.

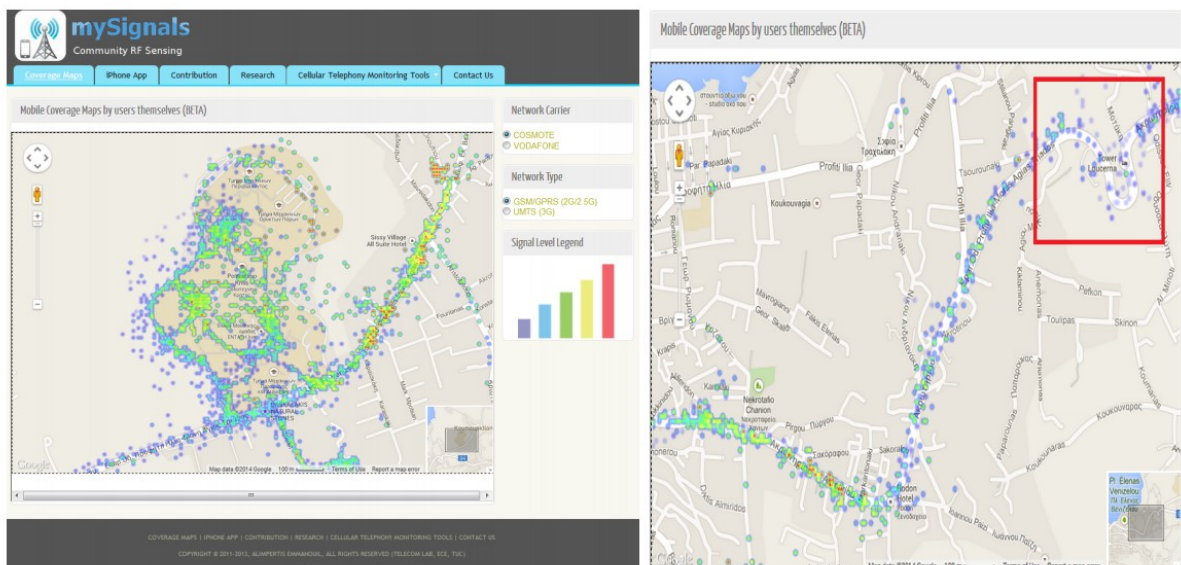
Η βιβλιοθήκη scikit-mobility Η βιβλιοθήκη scikit-mobility [76] είναι μια βιβλιοθήκη για την ανάλυση της ανθρώπινης κινητικότητας. Είναι γραμμένη και υποστηρίζει τη γλώσσα προγραμματισμού Python. Η βιβλιοθήκη αυτή παρέχει τα κάτωθι εργαλεία ανάλυσης της ανθρώπινης κίνησης:

- Αναπαράσταση των ατομικών διαδρομών και των ροών κινητικότητας σε κατάλληλες δομές δεδομένων. Οι δομές αυτές ορίζονται ως TrajDataFrame και FlowDataFrame αντίστοιχα.
- Διαχείριση και επεξεργασία δεδομένων κινητικότητας διαφόρων μορφών (αρχεία λεπτομερειών καταγραφής κλήσεων, δεδομένα GPS, δεδομένα προερχόμενα από τα κοινωνικά δίκτυα, δεδομένα ερευνών κ.α.)
- Εξαγωγή μονάδων μέτρησης ανθρώπινης κινητικότητας και μοτίβων από τα δεδομένα, τόσο σε ατομικό όσο και σε μαζικό επίπεδο
- Δημιουργία τεχνητών (synthetic) ατομικών διαδρομών αξιοποιώντας διάφορα μαθηματικά μοντέλα
- Δημιουργία τεχνητών (synthetic) ροών κινητικότητας, αξιοποιώντας διάφορα μοντέλα προσομοίωσης κινήσεων πληθυσμών.
- Αξιολόγηση κινδύνου παραβίασης της ιδιωτικότητας (privacy risk) που δύναται να προκληθεί από κακόβουλη χρήση του συνόλου δεδομένων κινητικότητας.

4.2 Μοντέλο ατομικής κινητικότητας - Ανάλυση δεδομένων καταγραφών λεπτομερειών κλήσεων (CDR) για μετακινήσεις στην περιοχή των Χανίων

Για την ανάλυση της ατομικής κινητικότητας αναζητήθηκε κατάλληλο σύνολο δεδομένων. Χρησιμοποιήθηκε το σύνολο δεδομένων της έρευνας των Alibertis et al. [7]. Στην έρευνα αυτή, αναπτύχθηκε εφαρμογή, την οποία αποκάλεσαν MySignals. Ουσιαστικά, πρόκειται για μια μέθοδο ανίχνευσης ραδιοσυχνοτήτων, που χρησιμοποιεί ως αισθητήρες τα κινητά τηλέφωνα των χρηστών που συμμετείχαν στην έρευνα. Η εφαρμογή δημιουργεί χάρτες κάλυψης δικτύου, βάσει μετρήσεων μιας εφαρμογής για κινητά τηλέφωνα μάρκας iPhone που αναπτύχθηκε για το σκοπό αυτό. Πιο αναλυτικά, κατασκευάζει ένα γεωγραφικό σύστημα πληροφοριών (GIS), αποτελούμενο από την εφαρμογή για iPhone, μια κεντρική βάση δεδομένων και μια σελίδα που παρουσιάζει την κάλυψη δικτύου στο χάρτη. Όπως φαίνεται και στην

εικόνα 4.1, στην εφαρμογή οι χρήστες της μπορούν να εφαρμόσουν φίλτρα. Έτσι, ο χάρτης μπορεί να προσαρμοστεί βάσει του τηλεπικοινωνιακού παρόχου και της τεχνολογίας δικτύου. Επίσης, στο δεξί τμήμα της εικόνας και συγκεκριμένα στην επισημασμένη περιοχή, φαίνεται η χαμηλή ποιότητα σήματος. Οι ερευνητές αιτιολογούν την χαμηλή κάλυψη δικτύου στη γεωμορφολογία της περιοχής αυτής, μιας και βρίσκεται μεταξύ δυο λόφων. Στο χάρτη, τα σημεία χαμηλής κάλυψης δικτύου χρωματίζονται με ωχρό μοβ χρώμα.



Σχήμα 4.1: Παρουσίαση της ιστοσελίδας MySignals [7].

Το σύνολο δεδομένων που συνέλεξαν οι ερευνητές αποτελείται από περίπου ένα εκατομμύριο καταγραφές ισχύος λαμβανόμενου σήματος (Received Signal Strength, RSS). Στην έρευνα, πραγματοποιήθηκε επίσης μια χωροχρονική ανάλυση στα πραγματικά δεδομένα και ανακαλύφθηκε η διακύμανση RSS με την πάροδο του χρόνου. Έτσι, ανακαλύφθηκαν περιοχές των Χανίων με χαμηλή κάλυψη δικτύου.

Βλέπουμε λοιπόν, πως το σύνολο δεδομένων των ερευνητών δεν είχε χρησιμοποιηθεί για το σκοπό της ανάλυσης της ανθρώπινης κίνησης. Συνεπώς, η ανάλυσή που πραγματοποιείται στη συνέχεια δίνει μια νέα διάσταση στο σύνολο των δεδομένων αυτών, και παρουσιάζει εμπράκτως την επαναχρησιμοποίηση δεδομένων για διαφορετικούς σκοπούς από αυτούς για τους οποίους συλλέχθηκαν.

4.2.1 Σύνολο δεδομένων (dataset)

Το σύνολο δεδομένων που αξιοποιήθηκε για την ανάλυση ατομικής κίνησης, ήταν τα δεδομένα που καταγράφηκαν από την έρευνα MySignals. Η MySignals αποτελείται από περίπου 445 χιλιάδες ξεχωριστές χωροχρονικές καταγραφές σε μορφή **CSV**. Η σημαντική αριθμητική διαφορά των χωροχρονικών καταγραφών με τον αριθμό των καταγραφών ισχύος λαμβανόμενου σήματος (RSS), προκύπτει από το γεγονός πως οι τελευταίες σημειώνονταν πολλαπλές φορές σε σύντομο χρονικό διάστημα ανά καταγραφή. Τα δεδομένα συλλέχθηκαν από ένα δίκτυο δέκα χρηστών κινητών τηλεφώνων μάρκας iPhone. Η συλλογή των δεδομένων έγινε μέσω της εφαρμογής MySignals, που αναπτύχθηκε για το λογισμικό iOS. Η έρευνα διήρκεσε περίπου 8 μήνες. Η εφαρμογή MySignals συγκρατούσε ανά τακτά χρονικά διαστήματα τη λαμβανόμενη ισχύ σήματος (RSSI), σε dBm, την τηλεπικοινωνιακή κυψέλη, καθώς και τη θέση των χρηστών μέσω του αισθητήρα GPS του κινητού τους τηλεφώνου. Επίσης, συλλέχθηκαν και άλλες πληροφορίες περιβάλλοντος, όπως η χρονική σήμανση, το αναγνωριστικό της συσκευής και άλλα. Τα δεδομένα χρησιμοποιήθηκαν κυρίως για τον εντοπισμό σταθμού βάσης, αλλά και για την ανάλυση κάλυψης δικτύου. Οι μετρήσεις έγιναν για έναν κεντρικό Σταθμό Βάσης (central Base Station, BS) ενός μεγάλου παρόχου κινητής τηλεφωνία των Χανίων.

Η αποθήκευση των πληροφοριών και των μετρήσεων γινόταν αρχικά τοπικά σε μία βάση δεδομένων sqlite. Ανά τακτά χρονικά διαστήματα, τα δεδομένα μεταφορτώνονταν σε έναν δικτυακό διακομιστή προκειμένου να αποθηκευτούν μόνιμα και να αποδεσμευτεί χώρος των κινητών τηλεφώνων. Οι ερευνητές σημείωσαν πως δεν συγκράτησαν προσωπικά δεδομένα των χρηστών, ωστόσο, σε επόμενη ενότητα διαπιστώνουμε πως κάτι τέτοιο δεν είναι αρκετό, καθώς ένας κακόβουλος χρήστης θα μπορούσε να ανακαλύψει ευαίσθητες πληροφορίες των χρηστών, όπως τη θέση οικίας, τις επισκεπτόμενες τοποθεσίες και άλλα.

Οι καταγεγραμμένες μετρήσεις περιλαμβάνουν τα ακόλουθα πεδία: χρονική σήμανση, αναγνωριστικό συσκευής, ισχύς λαμβανόμενου σήματος, γεωγραφικό πλάτος και μήκος, ακρίβεια εντοπισμού θέσης, αναγνωριστικό κυψέλης και άλλα. Οι μετρήσεις ξεκίνησαν την 1ή Σεπτεμβρίου του 2012 και διήρκεσαν ως και την 15η Απριλίου 2013. Όλα τα πεδία του συνόλου δεδομένων καθώς και μια σύντομη περιγραφή τους καταγράφεται στον πίνακα 4.1.

AA	Πεδίο	Περιγραφή
1	timestamp	Η ακριβής ημερομηνία και στιγμή της μέτρησης.
2	iPhoneUID	Μια μοναδική συσκευή iPhone, αντιστοιχισμένη με ένα ψευδο-αναγνωριστικό. Δεν χρησιμοποιούνται ευαίσθητα προσωπικά στοιχεία.
3	dt1	Η χρονική στιγμή που καταγράφηκε το rssi1 προκύπτει ως timestamp + dt1. Κάθε μέτρηση αποτελείται από 5 διαδοχικές μετρήσεις RSS.
4	dt2	Η χρονική στιγμή που καταγράφηκε το rssi2
5	dt3	Η χρονική στιγμή που καταγράφηκε το rssi3
6	dt4	Αντίστοιχα με τα 3,4,5
7	dt5	Αντίστοιχα με τα 3,4,5
8	rssi	Η τιμή RSSI που προκύπτει από το API του iOS. Προκύπτει ως μέσος όρος 5 διαδοχικών τιμών. Το χρονικό διάστημα μεταξύ των τιμών δε γνωστοποιείται από το iOS.
9	rssi1	Το RSSI, σε dBm, εξάγεται από το baseband του iPhone σε timestamp + dt1.
10	rssi2	Το RSSI, σε dBm, εξάγεται από το baseband του iPhone σε timestamp + dt1 + dt2
11	rssi3	Αντίστοιχα με τα 9,10
12	rssi4	Αντίστοιχα με τα 9,10
13	rssi5	Αντίστοιχα με τα 9,10
14	latitude	Το γεωγραφικό πλάτος που καταγράφεται από το aGPS του iPhone, κατά το timestamp
15	longitude	Το γεωγραφικό μήκος που καταγράφεται από το aGPS του iPhone, κατά το timestamp
16	finalLatitude	Το γεωγραφικό πλάτος που καταγράφεται από το aGPS του iPhone, κατά το timestamp + dt1 + dt2 + dt3 + dt4 + dt5
17	finalLongitude	Το γεωγραφικό μήκος που καταγράφεται από το aGPS του iPhone, κατά το timestamp + dt1 + dt2 + dt3 + dt4 + dt5
18	horizontalAccuracy	Οριζόντια ακρίβεια που παρέχεται από το aGPS του iPhone. Πειραματικά, διαπιστώθηκε πως ο υπολογισμός του iOS είναι απαισιόδοξος.
19	isMoving	0 εάν δεν υπάρχει διαφορά μεταξύ των 14,15 και 16,17. αλλιώς 1.
20	txPower	Ισχύς Μετάδοσης
21	cellID	Το αναγνωριστικό του κελιού
22	LAC	Κωδικός περιοχής τοποθεσίας. (Location Area Code)
23	MNC	Κωδικός δικτύου κινητής τηλεφωνίας, δηλαδή ο πάροχος δικτύου στον οποίο είναι εγγεγραμμένος ο χρήστης. Για παράδειγμα, το MNC = 1 αντιστοιχεί στην Cosmote.
24	ARFCN	Απόλυτος αριθμός καναλιού ραδιοσυχνότητας (Absolute radio-frequency channel number.)
25	freq_dlink	Downlink frequency carrier (υπολογίζεται με ARFCN)
26	freq_uplink	Uplink frequency carrier (υπολογίζεται με ARFCN)

Πίνακας 4.1: Τα πεδία του συνόλου δεδομένων και μια σύντομη περιγραφή τους [7].

4.2.2 Προ-επεξεργασία δεδομένων

Επιλογή υποσυνόλου των δεδομένων Για την επιθυμητή ανάλυση, επιλέχθηκαν τα εξής πεδία του πίνακα 4.1: timestamp, iPhoneUID, latitude και longitude, . Συνεπώς, από τις στήλες που φαίνονται στον πίνακα 4.1, αξιοποιήθηκαν στην ανάλυση μονάχα αυτές με αύξοντα αριθμό 1,2,14,15.

Φόρτωση δεδομένων σε pandas dataframe Προκειμένου να αξιοποιηθούν οι δυνατότητες της βιβλιοθήκης scikit-mobility, τα δεδομένα μεταφράστηκαν σε δομή δεδομένων κατάλληλη προς επεξεργασία.

Για το σκοπό αυτό, σε πρώτο στάδιο αξιοποιήθηκε η συνάρτηση της βιβλιοθήκης pandas "read_csv", η οποία μετέφρασε το σύνολο των δεδομένων CSV σε ένα **pandas dataframe**. Τα dataframes είναι πίνακες δυο διαστάσεων και δεν διαθέτουν στατικό μέγεθος. Μπορούν επίσης να αποτελούνται από ετερογενή δεδομένα, δηλαδή δεδομένα διαφορετικών τύπων. Στην προκειμένη περίπτωση, το πεδίο timestamp ήταν πεδίο ημερομηνίας, ενώ τα iPhoneUID, γεωγραφικό πλάτος και μήκος καταχωρήθηκαν ως αλφαριθμητικά.

Αξίζει να σημειωθεί πως τα pandas dataframes περιέχουν αριθμημένους άξονες (σειρές και στήλες). Οι αριθμητικές λειτουργίες που παρέχονται από τη βιβλιοθήκη για την επεξεργασία των δεδομένων, μπορούν να εφαρμοστούν σε καθέναν από τους δυο άξονες. Τα dataframes αποτελούν την κύρια δομή δεδομένων της βιβλιοθήκης pandas.

Μετατροπή τους σε trajectory dataframe Στη συνέχεια, το pandas dataframe μεταφράστηκε στην κλάση trajectory dataframe του scikit-mobility. Η κλάση skmob.core.trajectorydataframe.TrajDataFrame δέχτηκε ως είσοδο το pandas dataframe που κατασκευάστηκε στο προηγούμενο βήμα, ενώ προσδιορίστηκαν και αντιστοιχήθηκαν οι εξής στήλες:

1. Γεωγραφικό μήκος
2. Γεωγραφικό πλάτος
3. Ημερομηνία και ώρα
4. Αναγνωριστικό χρήστη

Επίσης, προσδιορίστηκε η κωδικοποίηση του συστήματος γεωγραφικών συντεταγμένων που χρησιμοποιούν οι συσκευές iPhone, το espg:4326.

Φιλτράρισμα δεδομένων Για το φιλτράρισμα του συνόλου δεδομένων της μορφής dataframe και την αφαίρεση θορύβου, αξιοποιήθηκε η συνάρτηση preprocessing.filtering.filter της βιβλιοθήκης scikit-mobility (βλ. ενότητα 4.1), η οποία είναι βασισμένη στην εργασία των Zheng et al. [77].

Πιο συγκεκριμένα, προκειμένου να μην προσμετρηθούν ακραίες τιμές στους υπολογισμούς, αφαιρέθηκαν οι διαδρομές των οποίων η σχετική τους ταχύτητα ήταν μεγαλύτερη των 200 χιλιομέτρων την ώρα.

Στη συνέχεια, καταμετρήθηκε ο συνολικός αριθμός καταγραφών ανά αναγνωριστικό συσκευής. Τα αποτελέσματα της καταμέτρησης παρουσιάζονται στον πίνακα 4.2. Το αναγνωριστικό συσκευής με κωδικό "510635002cb29804d54bff664cab52b" αφαιρέθηκε από το σύνολο δεδομένων, καθώς οι καταγραφές που το αφορούσαν ήταν ελάχιστες και δεν θα μπορούσαν να εξαχθούν χρήσιμα συμπεράσματα από αυτό.

Αναγνωριστικό συσκευής	Συνολικός αριθμός καταγραφών
1e33db5d2be36268b944359fbdbdad21	51372
22223276ea84bbce3a62073c164391fd	144219
510635002cb29804d54bff664cab52be	2
6882f6cf8c72d6324ba7e6bb42c9c7c2	69361
7023889b4439d2c02977ba152d6f4c6e	838
7cbc37da05801d46e7d80c3b99fd5adb	84475
8425a81da55ec16b7f9f80c139c235a2	537
892d2c3aae6e51f23bf8666c2314b52f	25419
a841f74e620f74ec443b7a25d7569545	68439

Πίνακας 4.2: Μοναδικά αναγνωριστικά συσκευών και συνολικός αριθμός καταγραφών ανά συσκευή.

Παρατηρείται επίσης πως ο δεύτερος κατά σειρά χρήστης του πίνακα 4.2 είναι και αυτός για τον οποίο περιέχονται οι περισσότερες καταγραφές (144 χιλιάδες). Οι

Αναγνωριστικό Συσκευής	Ψευδώνυμο
1e33db5d2be36268b944359fbdbdad21	device1
22223276ea84bbce3a62073c164391fd	device2
6882f6cf8c72d6324ba7e6bb42c9c7c2	device3
7023889b4439d2c02977ba152d6f4c6e	device4
7cbc37da05801d46e7d80c3b99fd5adb	device5
8425a81da55ec16b7f9f80c139c235a2	device6
892d2c3aae6e51f23bf8666c2314b52f	device7
a841f74e620f74ec443b7a25d7569545	device8

Πίνακας 4.3: Πίνακας αντιστοίχισης αναγνωριστικού συσκευή με ψευδώνυμο.

υπόλοιποι πέντε χρήστες διαθέτουν είκοσι με ενενήντα χιλιάδες καταγραφές, ενώ διατηρούνται δυο χρήστες με λίγη πληροφορία για αυτούς (κάτω των χιλίων πεδίων).

Πειραματικά, από τις 445 χιλιάδες καταγραφές αφαιρέθηκαν κατά το φιλτράρισμα περίπου οι 750 αυτών.

Μετονομασία στήλης Για λόγους αναγνωσιμότητας και ευκολίας στην ανάλυση των δεδομένων, τα μοναδικά αναγνωριστικά των χρηστών μετονομάστηκαν. Η αντιστοίχιση των αναγνωριστικών συσκευής και των ψευδωνύμων καταγράφεται στον πίνακα 4.3. Η σύμβαση αυτή διατηρείται στο υπόλοιπο κεφάλαιο, για όσες αναλύσεις πραγματοποιήθηκαν.

4.2.3 Εξαγωγή μονάδων μέτρησης κινητικότητας

Στην ενότητα αυτή σημειώνονται οι διάφορες μονάδες μέτρησης ατομικής κινητικότητας που εξήχθησαν από το φιλτραρισμένο σύνολο δεδομένων. Συνοπτικά, εξήχθησαν οι ακόλουθες μονάδες μέτρησης ατομικής κινητικότητας:

1. Ακτίνα περιστροφής (Radius of Gyration)
2. Κ-οστή ακτίνα περιστροφής (k-Radius of Gyration)
3. Μέγιστη διανυθείσα απόσταση ανά συσκευή

4. Τυχειότητα κίνησης
5. Συνολικός αριθμός ξεχωριστών τοποθεσιών ανά συσκευή
6. Υπολογισμός κινδύνου απώλειας προσωπικών δεδομένων

Ακτίνα περιστροφής Στην ενότητα αυτή υπολογίστηκε η ακτίνα περιστροφής σε χιλιόμετρα των καταγραφών του φιλτραρισμένου dataframe διαδρομών που περιγράφηκε στην προηγούμενη ενότητα. Συγκεκριμένα, η ακτίνα περιστροφής υπολογίστηκε ως εξής [20, 78]:

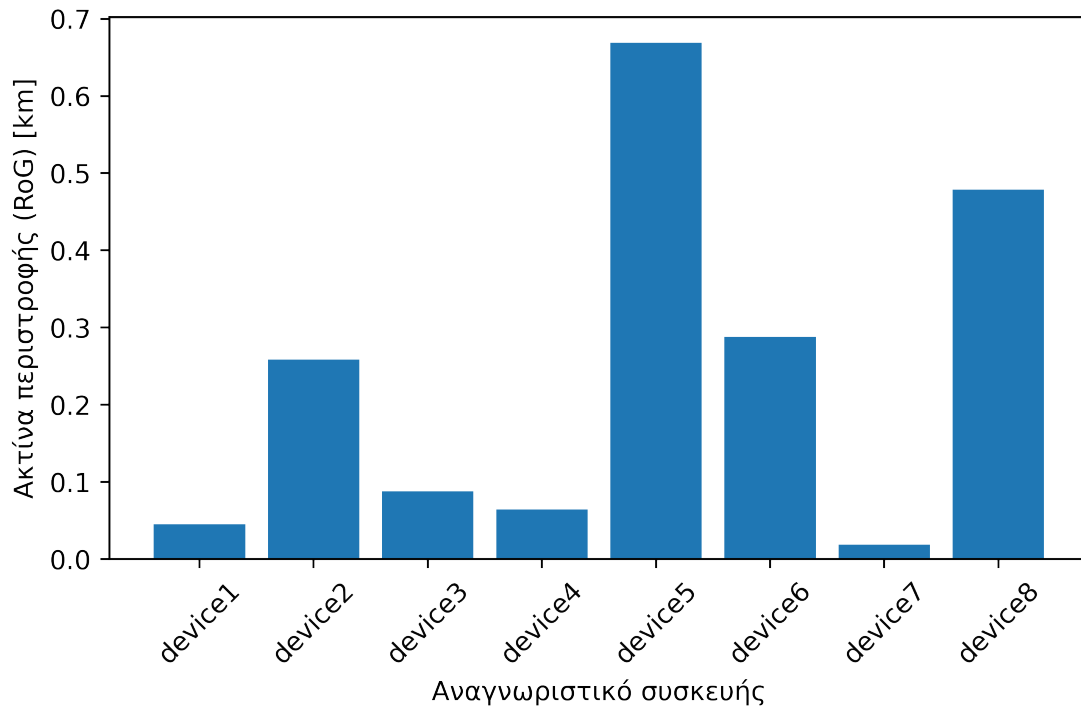
$$r_g(u) = \sqrt{\frac{1}{n_u} \sum_{i=1}^{n_u} \text{dist}(r_i(u) - r_{cm}(u))^2},$$

όπου το $r_i(u)$ αναπαριστά τις n_u θέσεις που καταγράφησαν για τον άνθρωπο με αναγνωριστικό u και το $r_{cm}(u)$ είναι το κέντρο μάζας της διαδρομής του χρήστη u . Όπως έχει ήδη αναφερθεί, η ακτίνα περιστροφής αναπαριστά τη χαρακτηριστική απόσταση που διένυσε ο χρήστης u .

Για την πειραματική υλοποίηση αξιοποιήθηκε η συνάρτηση της βιβλιοθήκης `scikit-mobility skmob.measures.individual.radius_of_gyration`.

Τα αποτελέσματα της ανάλυσης φαίνονται στο σχήμα 4.2. Στο σχήμα, γίνεται εμφανές πως η συσκευή με αναγνωριστικό "device5" είναι και αυτός με τη μεγαλύτερη ακτίνα περιστροφής, ενώ κατά σειρά ακολουθούν οι συσκευές 8, 6 και 2. Η ακτίνα περιστροφής είναι η τυπική απόσταση που διανύει ένας χρήστης σε ένα διάστημα παρατήρησης. Οι Gozalez et al. [20] σημειώνουν πως η ακτίνα περιστροφής αυξάνεται, όσο αυξάνεται και το διάστημα παρατήρησης. Παρατηρούμε πως οι συσκευές που παρουσίασαν τη μεγαλύτερη ακτίνα περιστροφής ήταν και αυτές για τις οποίες είχαμε την περισσότερη πληροφορία σε σχέση με τις υπόλοιπες.

Κ-οστή ακτίνα περιστροφής (k-Radius of Gyration) Για τον υπολογισμό της κ-οστής ακτίνας περιστροφής σε χιλιόμετρα του φιλτραρισμένου dataframe διαδρομών που περιγράφηκε στην ενότητα 4.2.2, αξιοποιήθηκε η συνάρτηση `skmob.measures.individual.k_radius_of_gyration` του `scikit-mobility`. Η συνάρτηση αυτή υπολογίζει την κ-οστή ακτίνα περιστροφής βάσει του ακόλουθου τύπου [19]:



Σχήμα 4.2: Ακτίνα περιστροφής ανά αναγνωριστικό συσκευής.

$$r_g^{(k)}(u) = \sqrt{\frac{1}{n_u^k} \sum_{i=1}^k (r_i(u) - r_{cm}^{(k)}(u))^2},$$

όπου το $r_i(u)$ αναπαριστά τις $n_u^{(k)}$ καταγραφείσες τοποθεσίες για το χρήστη u στις k πιο συχνές τοποθεσίες, ενώ το $r_{cm}^{(k)}(u)$ είναι το κέντρο μάζας της διαδρομής του χρήστη u (λαμβάνοντας μόνο υπόψιν τις επισκέψεις στις k πιο συχνές τοποθεσίες). Η k -οστή ακτίνα περιστροφής δείχνει τη χαρακτηριστική απόσταση που διανύθηκε από έναν άνθρωπο βάσει των k συχνότερων τοποθεσιών που εκείνος επισκέφτηκε.

Τα αποτελέσματα της εφαρμογής του αλγόριθμου για $K=2,5$ και 10 φαίνονται στον πίνακα 4.4. Παρατηρούμε πως όσο αυξάνεται το K για κάθε χρήστη, τόσο αυξάνεται η ακτίνα περιστροφής που χαρακτηρίζει την κίνησή του. Αντίστοιχα με την προηγούμενη παράγραφο, φαίνεται πως οι συσκευές για τις οποίες διαθέτουμε περισσότερη πληροφορία εμφανίζουν και μεγαλύτερη k -οστή ακτίνα περιστροφής. Βλέπουμε λοιπόν πως η k -οστή ακτίνα περιστροφής επηρεάζεται από το διάστημα παρατήρησης για το οποίο υπολογίζεται.

Αναγνωριστικό συσκευής	K = 2	K = 5	K = 10
device1	0.006156	0.009839	0.010244
device2	0.326607	0.455353	0.427758
device3	0.008643	0.007174	0.007066
device4	0.025215	0.042611	0.048456
device5	0.066453	0.452976	0.613608
device6	0.038407	0.02899	0.075481
device7	0.010337	0.015883	0.018194
device8	0.0121	0.011382	0.013603

Πίνακας 4.4: K-οστή ακτίνα περιστροφής των μοναδικών συσκευών της έρευνας MySignals σε χιλιόμετρα, για K=2,5 και 10.

Μέγιστη διανυθείσα απόσταση ανά συσκευή Προκειμένου να βρεθεί η μέγιστη διανυθείσα απόσταση ανά συσκευή αξιοποιήθηκε η συνάρτηση `skmob.measures.individual.maximum_distance` της βιβλιοθήκης `scikit-mobility`. Η συνάρτηση αυτή υπολογίζει τη μέγιστη διανυθείσα απόσταση d_{max} ως εξής [29]:

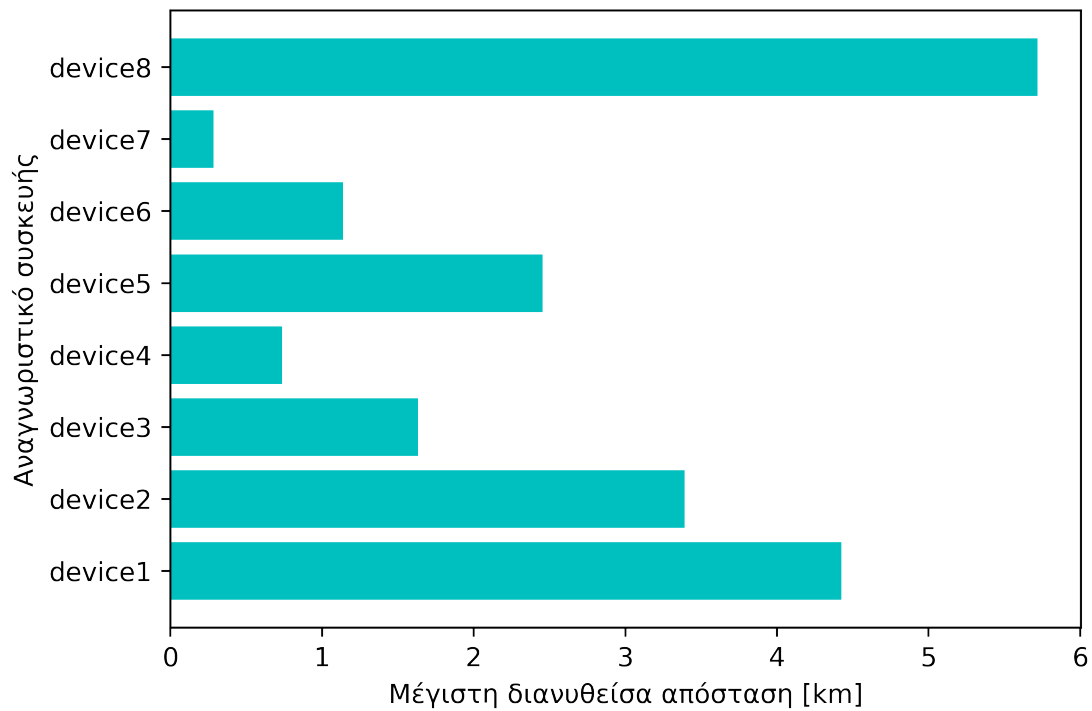
$$d_{max} = \max_{1 \leq i < j < n_u} dist(r_i, r_j),$$

όπου n_u ο αριθμός των σημείων που καταγράφηκαν για το χρήστη u , ενώ r_i και $r_i + 1$ οι γεωγραφικές συντεταγμένες δυο συναπλών σημείων καταγραφής. Τέλος, το `dist` είναι η γεωγραφική απόσταση των δύο σημείων.

Η μέγιστη διανυθείσα απόσταση ανά χρήστη φαίνεται στο σχήμα 4.3. Από το σχήμα, είναι προφανές πως η συσκευή 8 είναι και αυτή που εμφάνιζε τη μεγαλύτερη διανυθείσα απόσταση στο διάστημα παρατήρησης, στα 8 χιλιόμετρα. Στη συνέχεια, ακολουθούν οι συσκευές 1,2 και 5.

Συνολικός αριθμός ξεχωριστών τοποθεσιών ανά συσκευή και τυχαιότητα κίνησης Για τον υπολογισμό του συνολικού αριθμού ξεχωριστών τοποθεσιών ανά συσκευή, αξιοποιήθηκε ο αλγόριθμος των Gonzalez et al. [20].

Τα αποτελέσματα της εκτέλεσης του αλγορίθμου παρουσιάζονται στον πίνακα 4.5. Τις περισσότερες διακριτές τοποθεσίες φαίνεται να τις επισκέφτηκε ο χρήστης 8, για



Σχήμα 4.3: Μέγιστη διανυθείσα απόσταση σε χιλιόμετρα ανά συσκευή.

τον οποίο άλλωστε παρατηρήθηκε προηγουμένως πως διένυσε και την μεγαλύτερη απόσταση. Τις λιγότερες δε διακριτές τοποθεσίες φαίνεται να τις επισκέφτηκε η συσκευή με αναγνωριστικό "device7" (13). Σε συνδυασμό με τη χαμηλή ακτίνα περιστροφής της συσκευής εκείνης, μπορούμε να υποθέσουμε πως η χρήση της συσκευής καταγραφής γινόταν μονάχα σε ένα συγκεκριμένο χώρο.

Αναγνωριστικό συσκευής	Αριθμός διακριτών τοποθεσιών
device1	71
device2	2114
device3	22
device4	124
device5	1457
device6	173
device7	13
device8	5711

Πίνακας 4.5: Συνολικός αριθμός διακριτών τοποθεσιών που αναγνωρίστηκαν ανά συσκευή.

Όσον αφορά την εντροπία των ανθρώπων που συμμετείχαν στην έρευνα MySignals, υπολογίστηκε αξιοποιώντας την εξής συνάρτηση:

$$E_{rand}(u) = \log_2(N_u),$$

όπου N_u ο αριθμός των διακριτών τοποθεσιών που επισκέφτηκε ο χρήστης u . Η τυχαία εντροπία υπολογίζει το βαθμό προβλεψιμότητας των τοποθεσιών που επισκέπτεται ο χρήστης, εάν επισκέπτεται με ίση πιθανότητα την καθεμία. Η εντροπία των συσκευών της έρευνας MySignals παρουσιάζεται στον πίνακα 4.6. Στο σχήμα, διακρίνεται πως οι συσκευή 7 εμφανίζει τη μεγαλύτερη προβλεψιμότητα στην κίνησή της στο χώρο, ενώ η συσκευές 2 και 5 τη μικρότερη προβλεψιμότητα.

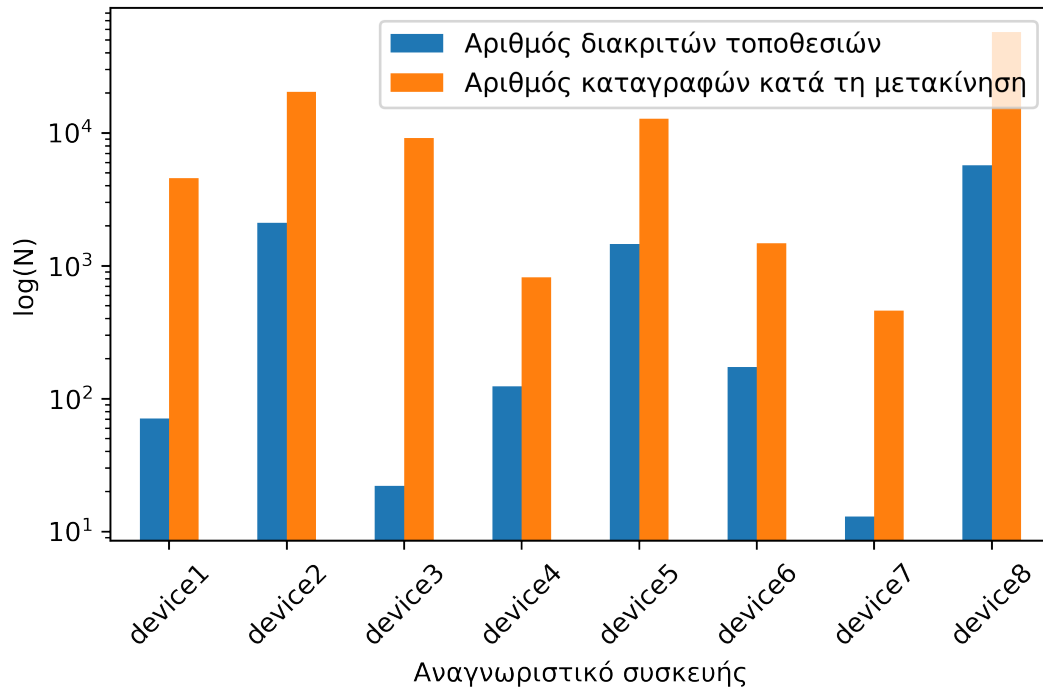
Αναγνωριστικό συσκευής	Τυχαία εντροπία
device1	6.149747
device2	11.04576
device3	4.459432
device4	6.954196
device5	10.50879
device6	7.434628
device7	3.70044
device8	12.47953

Πίνακας 4.6: Εντροπία θέσης ανά συσκευή της έρευνας MySignals.

Συσχέτιση αριθμού διακριτών τοποθεσιών και συνολικών διαδρομών Στην έρευνά τους, οι Axhausen et al. [4] συσχέτισαν τον αριθμό των συνολικών διαδρομών των χρηστών με τον αριθμό των διακριτών σημείων που κατάφεραν να αναγνωρίσουν. Μάλιστα, διαπίστωσαν μια αναλογική σχέση των δυο, καθώς όσο περισσότερο οι άνθρωποι ταξίδευαν, τόσο περισσότερο αυξάνονταν και τα ξεχωριστά σημεία που ήταν σε θέση να αναγνωρίσει ο αλγόριθμός των ερευνητών.

Παρομοίως, και στην έρευνα MySignals παρατηρούμε κάτι αντίστοιχο. Όπως βλέπουμε και στο σχήμα 4.4, όσο μεγαλύτερος ο αριθμός των καταγραφών μετακίνησης ενός χρήστη, τόσο περισσότερο αυξάνεται και ο συνολικός αριθμός διακριτών τοποθεσιών που δύναται να αναγνωρίσει ο αλγόριθμος. Αυτό οφείλεται τόσο στο γεγονός πως αυξάνεται ο όγκος της αναλυόμενης πληροφορίας, όσο και στη διαπίστωση πως σε ένα μεγάλο διάστημα παρατήρησης οι άνθρωποι τείνουν να επισκέπτονται ξεχωριστές τοποθεσίες.

Υπολογισμός κινδύνου απώλειας προσωπικών δεδομένων Αξιοποιώντας τη συνάρτηση `skmob.privacy.attacks.HomeWorkAttack` της βιβλιοθήκης `scikit-mobility`, διαπιστώθηκε πως επιλογή των συντακτών της έρευνας MySignals να αποκρύψουν το πραγματικό αναγνωριστικό συσκευής δεν ήταν αρκετή για να διασφαλίσουν την προστασία των προσωπικών δεδομένων των χρηστών. Για όλες τις συσκευές της έρευνας, κάποιος κακόβουλος χρήστης θα μπορούσε να αποκαλύψει ευαίσθητα



Σχήμα 4.4: Λογαριθμική σύγκριση αριθμού διακριτών τοποθεσιών που αναγνωρίστηκαν και συνολικού αριθμού καταγραφών μετακίνησης χρηστών.

προσωπικά δεδομένα, όπως τη θέση οικίας και εργασίας των χρηστών.

Η συνάρτηση HomeWorkAttack αξιοποιεί τον αλγόριθμο των Pellungrini et al. [36]. Τα αποτελέσματα της εκτέλεσης της παρουσιάζονται στον πίνακα 4.7. Είναι λοιπόν προφανές πως από τη στιγμή που η επικινδυνότητα για κάθε χρήστη είναι στο μέγιστο, τότε τα δεδομένα της έρευνας MySignals εμπεριέχουν πολύ ευαίσθητες πληροφορίες. Σε μελλοντικά σύνολα δεδομένων, οι ερευνητές οφείλουν να εισάγουν τεχνικές τεχνητού θορύβου, προκειμένου να διαφυλάξουν την ιδιωτικότητα των ατόμων.

4.2.4 Αναγνώριση θέσεων οικίας και παρουσίαση αποτελεσμάτων

Στην ενότητα αυτή, τα φιλτραρισμένα δεδομένα αναλύθηκαν κάνοντας χρήση της συνάρτησης ανάλυσης ατομικής κινητικότητας `skmob.individual.home_location` της βιβλιοθήκης `scikit-mobility`.

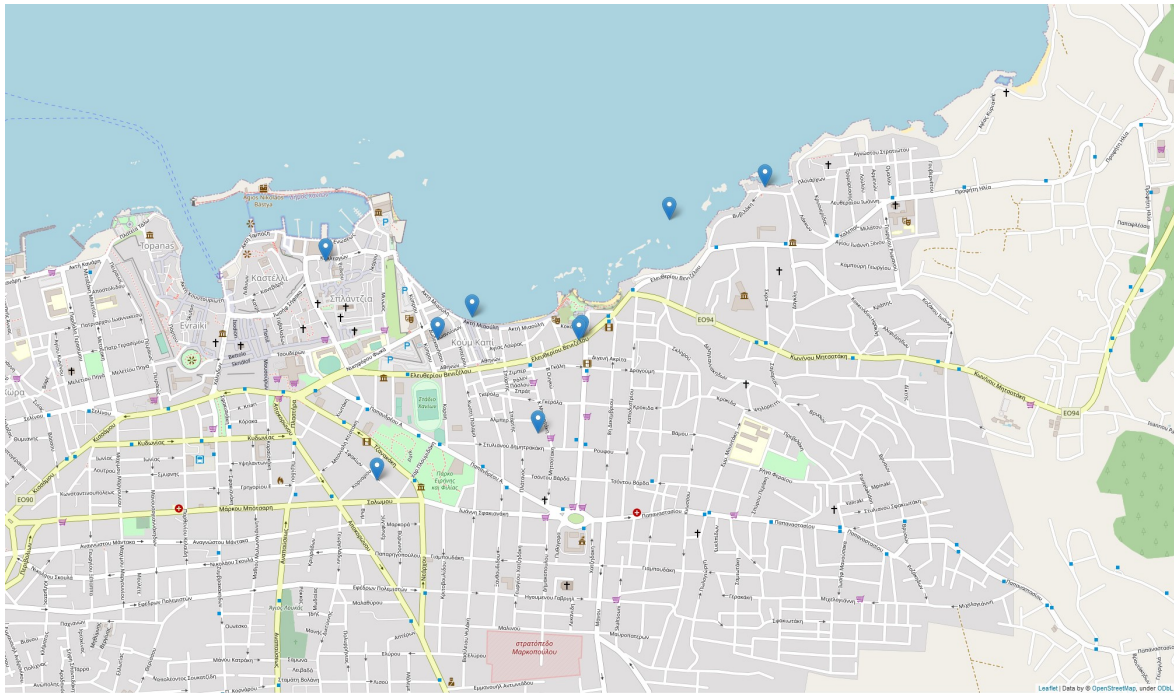
Η συνάρτηση αυτή αξιοποιεί έναν rule-based αλγόριθμο αναγνώρισης της οικίας των ανθρώπων. Ο αλγόριθμος αυτός, ο οποίος είναι βασισμένος στις έρευνες των

Αναγνωριστικό συσκευής	Ρίσκο
device1	1
device2	1
device3	1
device4	1
device5	1
device6	1
device7	1
device8	1

Πίνακας 4.7: Δυνητικά αποτελέσματα εκτέλεσης κακόβουλης επίθεσης για αποκάλυψη προσωπικών δεδομένων. Η κλίμακα ρίσκου είναι κανονικοποιημένη.

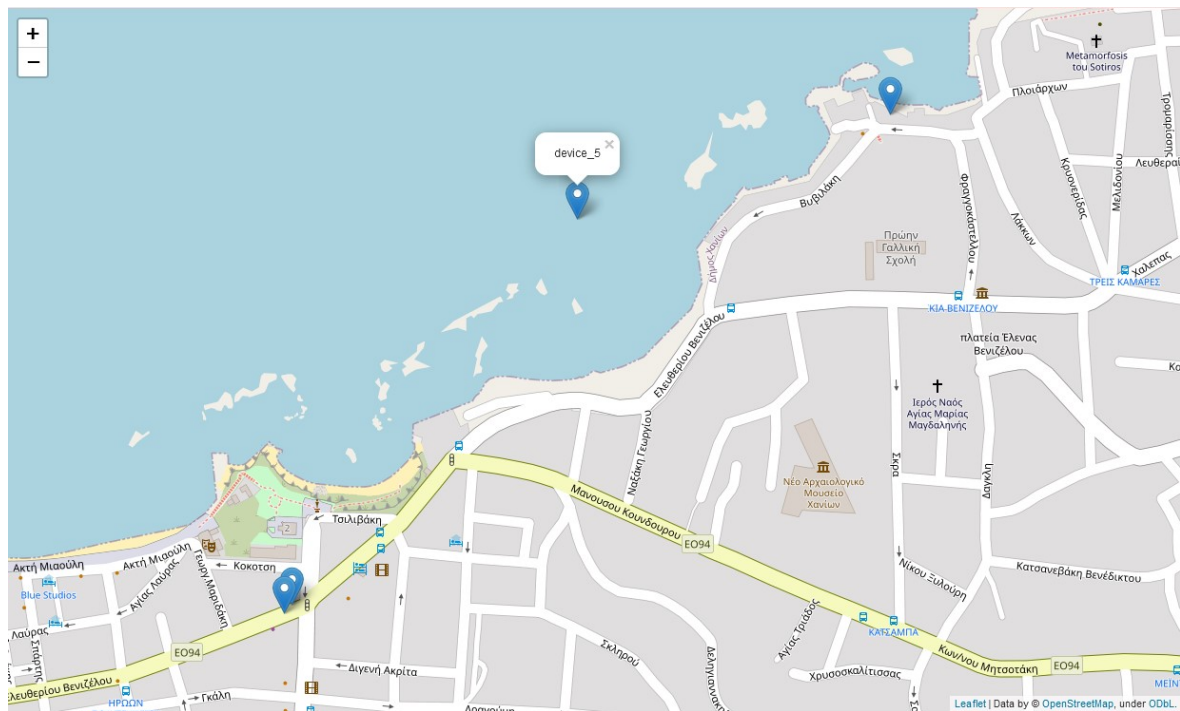
Rhithakkitnukoon et al. και των Csaji et al. [79, 80] παραμετροποιήθηκε κατάλληλα ώστε να είναι σε θέση να αναγνωρίσει τις κατοικίες των ανθρώπων. Συγκεκριμένα, από τις καταγραφές του συνόλου δεδομένων προσμετρήθηκαν μόνο αυτές μεταξύ των ωρών 03:00 και 05:00. Στο χρονικό αυτό πλαίσιο, ως οικία των ανθρώπων ορίστηκε η θέση στην οποία βρίσκονταν συχνότερα.

Τα αποτελέσματα της ανάλυσης οπτικοποιήθηκαν σε διαδραστικό χάρτη OpenStreetMaps αξιοποιώντας τη βιβλιοθήκη folium. Οι θέσεις οικίας των συμμετεχόντων της έρευνας επισημάνθηκαν με κατάλληλους δείκτες. Στην εικόνα 4.5 παρουσιάζονται οι θέσεις που ανιχνεύτηκαν.



Σχήμα 4.5: Οι ανιχνευθείσες θέσεις οικίας των συμμετεχόντων στην έρευνα MySIGNALS.

Αξιολόγηση των αποτελεσμάτων Στο σχήμα 4.6 διαπιστώνουμε πως μια από τις θέσεις οικίας των χρηστών που καταγράφηκε φαίνεται να βρίσκεται εκτός της οικιστικής ζώνης των Χανίων. Το εσφαλμένο αυτό αποτέλεσμα για το χρήστη με μοναδικό αναγνωριστικό "device_5" ενδεχομένως να οφείλεται σε ανακρίβεια των αισθητήρων του κινητού του τηλεφώνου ή σε χαμηλή κάλυψη δικτύου της οικίας του στις ώρες που περιορίστηκε η ανάλυση.

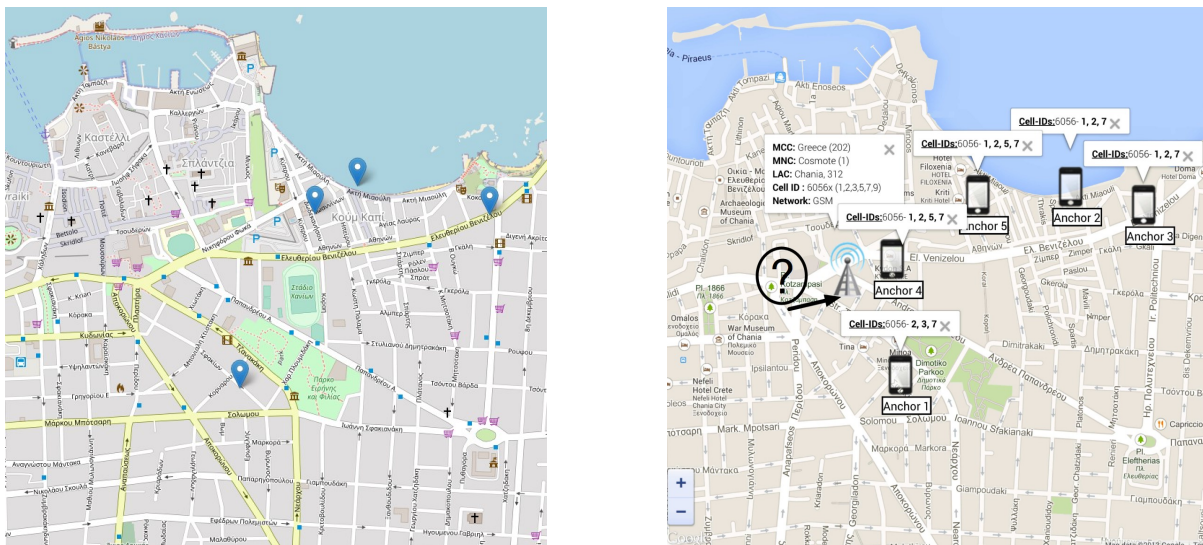


Σχήμα 4.6: Εσφαλμένη αναγνώριση θέσης οικίας ενός χρήστη της έρευνας MySignals.

Σύγκριση αποτελεσμάτων με τις πραγματικές θέσεις οικίας Αρκετό ενδιαφέρον προκύπτει από τη σύγκριση των αποτελεσμάτων της εφαρμογής του αλγορίθμου αναγνώρισης κατοικιών με τις πραγματικές θέσεις των σπιτιών των συμμετεχόντων.

Ως πραγματικές θέσεις των σπιτιών των συμμετεχόντων στην έρευνα, θεωρήθηκαν αυτές που καταγράφονται στην έρευνα των Alimpertis et al. [7]. Όπως φαίνεται και στο σχήμα 4.7, οι τοποθεσίες που αναγνωρίστηκαν από την εφαρμογή του αλγορίθμου εμφανίζουν πλήρη ή μεγάλη ταύτιση με τις πραγματικές. Πράγματι, οι οικίες με δείκτες "Anchor 1, 3 και 5" αναγνωρίστηκαν με μεγάλη ακρίβεια, το σπίτι με δείκτη "Anchor 2" αναγνωρίστηκε μερικά μέτρα εκτός της πραγματικής του θέσης, ενώ η οικία με δείκτη "Anchor 4" δεν κατέστη δυνατό να αναγνωριστεί από τον αλγόριθμο.

Διαπιστώνουμε λοιπόν πως ακόμη και χωρίς να γνωρίζουμε τις πραγματικές θέσεις των κατοικιών των ανθρώπων που συμμετείχαν στην έρευνα, μπορούμε να τις υπολογίσουμε με μεγάλη βεβαιότητα. Αυτό φυσικά προϋποθέτει την ύπαρξη ικανού συνόλου δεδομένων με επαρκείς καταγραφές κατά τη διάρκεια της νύχτας και συγκεκριμένα μεταξύ των ωρών που οι άνθρωποι αναπαύονται.



Σχήμα 4.7: Σύγκριση των κατοικιών χρηστών που αναγνωρίστηκαν αυτόματα από τον αλγόριθμο (αριστερό σχήμα) σε σχέση με τις πραγματικές τοποθεσίες των σπιτιών τους (δεξί σχήμα). Δεξί σχήμα από [7].

4.3 Μοντέλο κινήσεων πληθυσμών - Ανάλυση δεδομένων κινητικότητας μεταξύ των επαρχιών της Ιταλίας πριν και κατά τη διάρκεια της πανδημίας COVID-19

Η πανδημία του COVID-19 έφερε πολλές αλλαγές στις ζωές των ανθρώπων, και οδήγησε πολλές κυβερνήσεις να εφαρμόσουν μέτρα περιορισμού της κινητικότητας. Στην ενότητα αυτή, γίνεται μια προσπάθεια ανάλυσης της επίδρασης στην κινητικότητα μεταξύ περιοχών της Ιταλίας, για το πρώτο εξάμηνο του 2020. Επίσης, γίνεται και μια προσπάθεια εξεύρεσης κατάλληλων παραμέτρων του βαρυτικού μοντέλου, ώστε να προσομοιωθούν πιο αξιόπιστα οι διακοινοτικές μετακινήσεις πληθυσμών σε περιόδους πανδημίας και περιοριστικών μέτρων.

Ιταλία και COVID-19 Η Ιταλία ήταν από τις πρώτες χώρες της Ευρώπης που επλήγησαν σοβαρά από την πανδημία COVID-19. Το πρώτο θετικό κρούσμα του ιού στη χώρα ήταν στις 31 Ιανουαρίου του 2020. Η ασθένεια εξαπλώθηκε γρήγορα σε εθνικό επίπεδο, ενώ τον Απρίλιο του 2020 η Ιταλία ήταν η χώρα με τους

περισσότερους θανάτους από COVID-19 στην Ευρώπη.

4.3.1 Σύνολο δεδομένων

Για την ανάλυση των κινήσεων πληθυσμών μεταξύ των περιοχών της Ιταλίας, αξιοποιήθηκαν τα κάτωθι σύνολα δεδομένων:

1. Δεδομένα κινητικότητας μεταξύ επαρχιών της Ιταλίας από την 18/1/20 έως και την 26/06/20.
2. Δημογραφικά δεδομένα των επαρχιών της Ιταλίας.
3. Γεω-χωρικά δεδομένα των επαρχιών της Ιταλίας.
4. Αριθμός κρουσμάτων COVID-19 ανά ημέρα και επαρχία της Ιταλίας από την 24/02/2020 έως και την 6/12/2020.

Σύνολο δεδομένων κινητικότητας μεταξύ επαρχιών της Ιταλίας Ως πηγή δεδομένων κινητικότητας μεταξύ επαρχιών της Ιταλίας, αξιοποιήθηκε το dataset των Pepe et al. [81]. Οι Pepe et al. συνέλεξαν δεδομένα μεγάλης κλίμακας περίπου 170 χιλιάδων κινητών συσκευών, πριν και κατά τη διάρκεια της επιδημίας σε παν-Ιταλική κλίμακα. Ο στόχος τους ήταν να διευκολύνουν την αξιολόγηση του αντίκτυπου των πολιτειακών παρεμβάσεων, ώστε να περιοριστεί η έκταση της πανδημίας.

Το σύνολο δεδομένων που συνέλεξαν αποτελείται από έναν πίνακα προέλευσης προορισμού από και προς όλες τις επαρχίες της Ιταλίας. Στο σημείο αυτό σημειώνεται πως η Ιταλία διαιρείται συνολικά σε 107 επαρχίες, ενώ η κάθε μια από αυτές διαθέτει μοναδικό κωδικό.

Ο πίνακας προέλευσης-προορισμού περιέχει καταγραφές μετακίνησης από την 18/1/20 έως και την 26/06/20. Συνολικά ο πίνακας διαθέτει 5140 καταχωρήσεις για τις μετακινήσεις μεταξύ των επαρχιών της Ιταλίας και αποτελείται από 163 στήλες, οι οποίες είναι διαμορφωμένες ως εξής:

- Κωδικός επαρχίας προέλευσης
- Κωδικός επαρχίας προορισμού
- Ποσοστό μετακίνησης την 18-01-20

- Ποσοστό μετακίνησης την 19-01-20
- ...
- Ποσοστό μετακίνησης την ημέρα K
- Ποσοστό μετακίνησης την ημέρα K+1
- ...
- Ποσοστό μετακίνησης την 26-06-20

Δημογραφικά δεδομένα των επαρχιών της Ιταλίας Προκειμένου να μετατραπούν οι ποσοστιαίες μετακινήσεις πληθυσμών μεταξύ επαρχιών της Ιταλίας σε αριθμητικές μετακινήσεις, αντλήθηκαν δημογραφικά δεδομένα για τις εκάστοτε επαρχίες. Ως πηγή δημογραφικών στοιχείων αξιοποιήθηκε η Εθνική Στατιστική Υπηρεσία της Ιταλίας [82]. Το σύνολο δεδομένων περιέχει συνολικά 107 καταγραφές, και αποτελείται από τις εξής στήλες:

- Αύξων αριθμός καταγραφής
- Κωδικός επαρχίας
- Συντομογραφία επαρχίας
- Όνομα επαρχίας ολογράφως
- Πληθυσμός επαρχίας

Γεω-χωρικά δεδομένα των επαρχιών της Ιταλίας Για την αναπαράσταση των ροών πληθυσμών της Ιταλίας και τη σωστή τροφοδότηση του βαρυτικού μοντέλου, οι επαρχίες της Ιταλίας περιγράφησαν με κατάλληλα γεω-χωρικά δεδομένα. Τα δεδομένα αυτά αντλήθηκαν από το σύνολο δεδομένων των Celata et al. [83].

Το σύνολο δεδομένων των Celata et al. ήταν σε μορφή geo-JSON, δηλαδή σε μια μορφή κωδικοποίησης δομών γεωγραφικών δεδομένων. Περιλαμβάνει τις εξής πληροφορίες:

- Γεωμετρία επαρχίας

- Όνομα επαρχίας
- Κωδικός επαρχίας
- Συντομογραφία επαρχίας
- Όνομα περιοχής
- Κωδικός περιοχής

Αριθμός κρουσμάτων ανά ημέρα και επαρχία Τέλος, ως πηγή δεδομένων κρουσμάτων και εξέλιξης της πανδημίας η επίδραση της επιδημιολογικής κατάστασης μιας περιοχής στη συμπεριφορά των πολιτών αξιοποιήθηκε το [84]. Το συγκεκριμένο σύνολο δεδομένων περιέχει τον αριθμό των νέων κρουσμάτων ανά ημέρα και ανά επαρχία της Ιταλίας, για την περίοδο από 24/02/2020 έως και 6/12/2020. Συνολικά, το σύνολο διαθέτει 40 χιλιάδες καταγραφές και αποτελείται από τις εξής στήλες:

- Αύξων αριθμός καταγραφής
- Ημερομηνία
- Χώρα
- Κωδικός περιοχής
- Όνομα περιοχής
- Κωδικός επαρχίας
- Όνομα επαρχίας
- Συντομογραφία επαρχίας
- Γεωγραφικό πλάτος κέντρου επαρχίας
- Γεωγραφικό μήκος κέντρου επαρχίας
- Σύνολο νέων θετικών κρουσμάτων

4.3.2 Προ-επεξεργασία δεδομένων

Για την επεξεργασία των δεδομένων, επιλέχθηκε το σύνολο δεδομένων `pandas dataframe` ως το πλέον κατάλληλο. Όπως έχει ήδη αναφερθεί, τα `pandas dataframes` είναι δομές δεδομένων κατάλληλες για επεξεργασία μεγάλων συνόλων δεδομένων. Έτσι, όλες οι πηγές δεδομένων της προηγούμενης υπο-ενότητας μεταφράστηκαν σε `dataframes` ακολουθώντας τη διαδικασία που περιγράφεται παρακάτω.

Μετατροπή του γεω-χωρικού συνόλου δεδομένων σε `pandas dataframe` Καθώς το αρχείο γεω-χωρικών δεδομένων ήταν σε μορφή `geo-JSON`, αρχικά καταχωρήθηκε σε ένα `geopandas dataframe` κάνοντας χρήση της συνάρτησης `geopandas.read_file`.

Από τις συνημμένες πληροφορίες στα γεωγραφικά δεδομένα διατηρήθηκαν για τους σκοπούς της ανάλυσης μονάχα ο μοναδικός κωδικός της επαρχίας και η γεωμετρία που την περιγράφει.

Στη συνέχεια, το `geopandas dataframe` μεταφράστηκε σε `pandas dataframe` αξιοποιώντας τη συνάρτηση `pandas.DataFrame`.

Ενσωμάτωση δημογραφικών στοιχείων επαρχίας στο σύνολο δεδομένων Βάσει του κωδικού επαρχίας, στο σύνολο δεδομένων της προηγούμενης παραγράφου προστέθηκαν τα δημογραφικά στοιχεία των επαρχιών. Τελικά, το `pandas dataframe` που προέκυψε διέθετε τα εξής πεδία:

1. Αύξων αριθμός
2. Κωδικός επαρχίας
3. Γεωμετρία
4. Όνομα επαρχίας
5. Πληθυσμός επαρχίας

Μετατροπή πίνακα προέλευσης-προορισμού σε `dataframe` ροών πληθυσμού Τα `dataframes` ροών πληθυσμού (`flow dataframes`) είναι μια δομή δεδομένων της βιβλιοθήκης `scikit-mobility`, η οποία είναι βασισμένη σε `pandas dataframes`. Τα `flow`

dataframes διαθέτουν συναρτήσεις που διευκολύνουν την επεξεργασία δεδομένων μετακίνησης πληθυσμών.

Ο πίνακας προέλευσης προορισμού των Pepe et al. [81], προκειμένου να καταχωρηθεί σε μια δομή δεδομένων βασισμένη σε pandas dataframes της βιβλιοθήκης scikit-mobility δέχτηκε κατάλληλη επεξεργασία.

Πιο συγκεκριμένα, οι στήλες του αρχικού πίνακα αναλύθηκαν μέσω κώδικα γραμμένου σε γλώσσα Python. Ο κώδικας λάμβανε υπόψιν του τους πληθυσμούς των εκάστοτε επαρχιών και πολλαπλασίαζε την ποσοστιαία μετακίνηση μεταξύ πληθυσμών με το συνολικό πληθυσμό κάθε επαρχίας. Οι πληθυσμοί και οι αντιστοιχισή τους με τις επαρχίες προέκυψε από το pandas dataframe της προηγούμενης παραγράφου. Ο τελικός πίνακας **ροής πληθυσμού** που προέκυψε διέθετε συνολικά 6.5 εκατομμύρια καταγραφές ροής πληθυσμών και περιείχε ως στήλες την επαρχία προέλευσης, την επαρχία προορισμού, την ημερομηνία και τη ροή πληθυσμού μεταξύ των επαρχιών.

Τα δεδομένα που συλλέχθηκαν σε προηγούμενα βήματα καταχωρήθηκαν σε ένα flowdataframe. Η δομή αυτή, πέρα από τον πίνακα ροής πληθυσμού περιλάμβανε και τα γεω-χωρικά δεδομένα της κάθε περιοχής.

Φιλτράρισμα δεδομένων Καθώς πολλές από τις καταγραφές του flowdataframe διέθεταν μηδενικές ροές, γεγονός που δυσχέραινε την επεξεργασία, εφαρμόστηκε κατάλληλη συνάρτηση φιλτραρίσματος της βιβλιοθήκης pandas προκειμένου να απαλειφθούν οι τιμές αυτές. Η δομή δεδομένων που προέκυψε από το στάδιο αυτό περιορίστηκε στις 150 χιλιάδες καταγραφές.

4.3.3 Επίδραση των περιοριστικών μέτρων στην κίνηση μεταξύ επαρχιών

Προκειμένου να περιοριστεί η εξάπλωση της ασθένειας COVID-19, οι Ιταλικές αρχές εφάρμοσαν διάφορα μέτρα. Ένας συνοπτικός πίνακας των μέτρων που εφαρμόστηκαν στην Ιταλία παρουσιάζεται στο 4.8. Όπως φαίνεται και στον πίνακα, οι πρώτοι σημαντικοί περιορισμοί τέθηκαν στις 8 και 9 Μαρτίου του 2020. Στη συνέχεια, ως και τις αρχές Μαΐου τα μέτρα σταδιακά γινόντουσαν πιο αυστηρά.

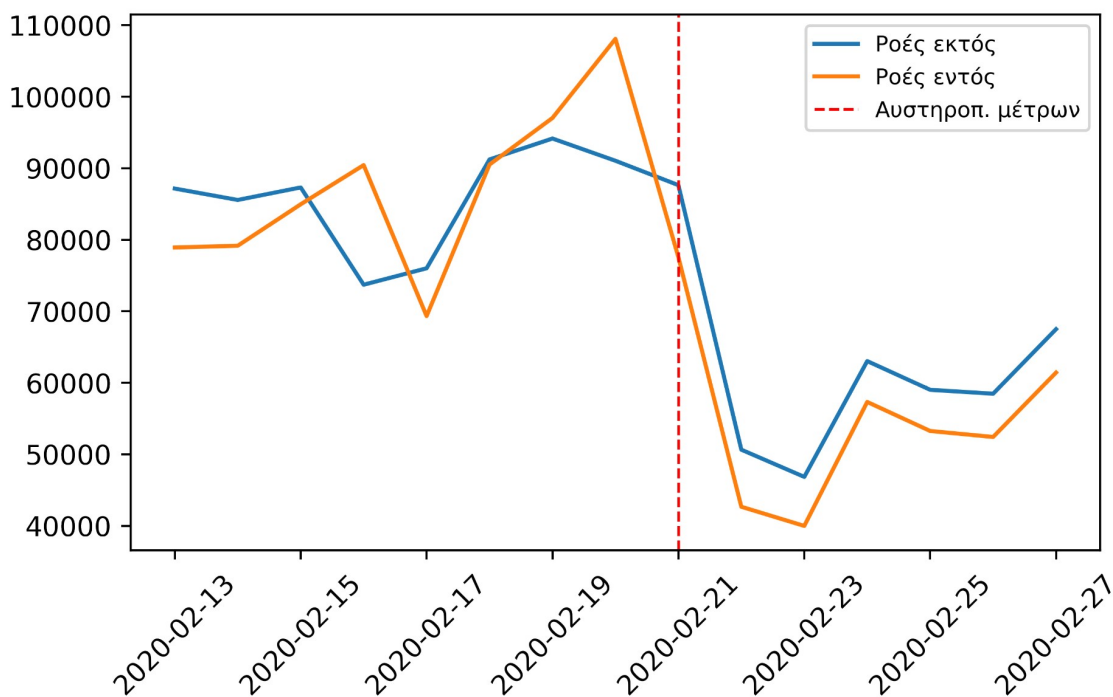
Τον Μάιο, η κυβέρνηση της Ιταλίας ανακοίνωσε σταδιακή άρση των περιορισμών, ενώ τον Νοέμβρη αναγκάστηκε να θέσει εκ νέου μέτρα συγκράτησης της κυκλοφορίας, καθώς ο αριθμός των κρουσμάτων είχε ανέλθει σε μεγάλα επίπεδα.

Ημερομηνία	Τύπος μέτρων	Σημειώσεις	Πηγή
21 Φεβ. 2020	Περιορισμοί	Τοποθετήθηκαν οδοφράγματα για τον περιορισμό μετακίνησης μεταξύ περιοχών.	[85]
		Τα μέτρα αφορούσαν 10 πόλεις στην επαρχία Lodi, 1 πόλη της επαρχίας Παντούα.	
8 Μαρ. 2020	Περιορισμοί	Απαγορεύτηκε η μετακίνηση προς και από τις περιοχές με υψηλή εξάπλωση της COVID-19.	[86]
		Τα μέτρα αφορούσαν πολλές περιοχές της Ιταλίας.	
9 Μαρ. 2020	Περιορισμοί	Απαγορευτικό σε όλη την επικράτεια	[87]
21 Μαρ. 2020	Περιορισμοί	Επέκταση απαγορευτικού και αυστηροποίηση μέτρων	[88]
22 Μαρ. 2020	Περιορισμοί	Απαγόρευση κάθε φυσικής δραστηριότητας στην Λομβαρδία.	[89]
4 Μαΐου 2020	Άρση περιορισμών	Ανακοινώθηκε πλάνο σταδιακής άρσης των περιορισμών	[90]
5 Νοεμβ. 2020	Περιορισμοί	Νυχτερινός περιορισμός κυκλοφορίας	[91]
13 Νοεμβ. 2020	Περιορισμοί	Προσθήκη νέων περιοχών στις κόκκινες ζώνες.	[92]

Πίνακας 4.8: Ημερομηνίες ανακοίνωσης περιοριστικών μέτρων καθώς και η ημερομηνία ανακοίνωσης της άρσης τους.

Η επαρχία Lodi Η επαρχία Lodi της Λομβαρδίας, ήταν η πρώτη περιοχή της Ιταλίας στην οποία επιβλήθηκαν μέτρα απαγόρευσης της κυκλοφορίας εισόδου και εξόδου από αυτήν. Ήταν άλλωστε η επαρχία που εμφάνισε τα πρώτα και περισσότερα κρούσματα στην αρχή της εξάπλωσης της πανδημίας στην Ιταλία.

Με τα μέτρα που εφαρμόστηκαν στις 21 Φεβρουαρίου, τοποθετήθηκαν οδοφράγματα στα σύνορα της επαρχίας προκειμένου να συγκρατηθεί η COVID-19



Σχήμα 4.8: Ροές από και προς την επαρχία Lodi πριν και μετά την ανακοίνωση περιοριστικών μέτρων. Με μπλε χρώμα σημειώνεται το σύνολο των ανθρώπων που έφευγε εκτός της επαρχίας, ενώ με πορτοκαλί το σύνολο των πολιτών που εισέρχονταν σε αυτήν.

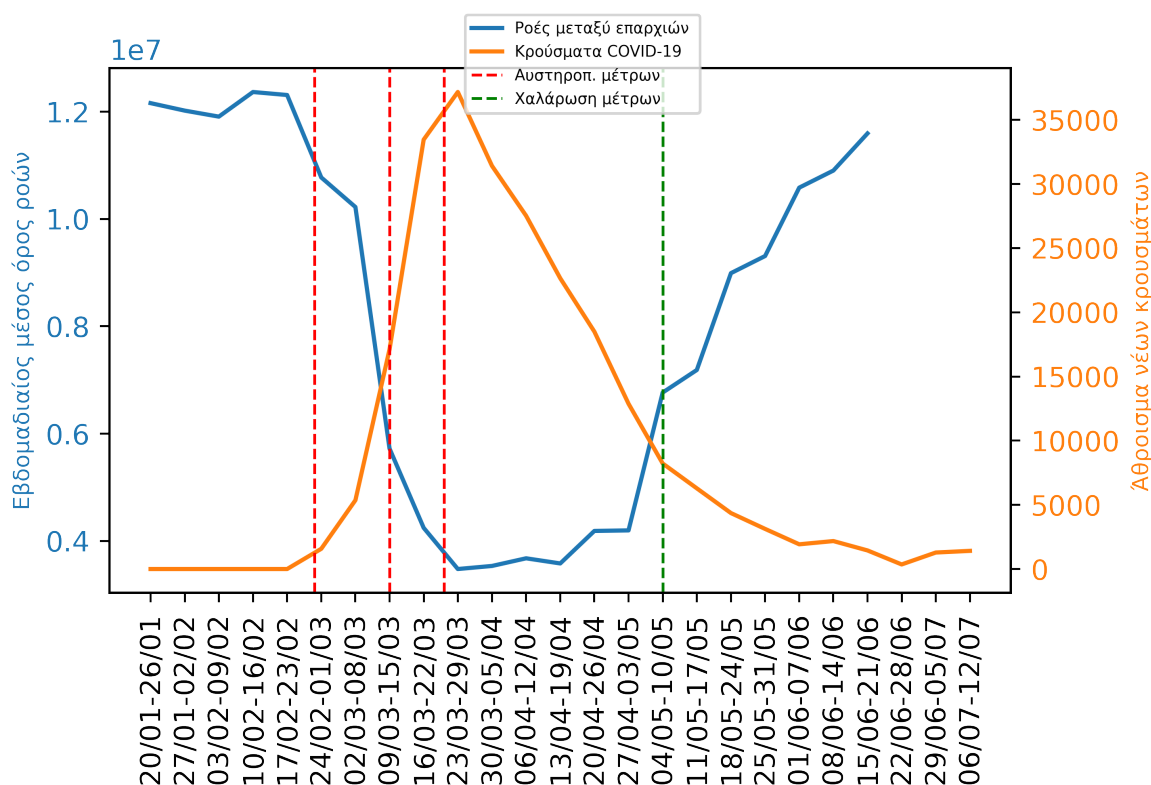
στο εσωτερικό της και να μην εξαπλωθεί στις γύρω πόλεις. Αυτό είχε ως αποτέλεσμα τη σημαντική μείωση ροών μεταξύ της Lodi και των γύρω περιοχών. Τα αποτελέσματα της εφαρμογής των κυκλοφοριακών ρυθμίσεων ήταν θεαματικά.

Όπως φαίνεται και στο σχήμα 4.8, τις ημέρες πριν την ανακοίνωση των περιοριστικών μέτρων παρουσιάστηκε μεγάλη αύξηση του συνόλου των ανθρώπων που εξερχόταν της επαρχίας Lodi. Πολλοί άνθρωποι, στην προσπάθειά τους να προφυλαχθούν, έσπευσαν να απομακρυνθούν ή και να επιστρέψουν στους τόπους καταγωγής τους. Έτσι, παρατηρούμε αυξημένη κινητικότητα αμέσως πριν την επιβολή περιορισμών. Ωστόσο, τις ημέρες μετά την εφαρμογή των μέτρων, ο αριθμός των ροών τόσο εκτός όσο και εντός της επαρχίας μειώθηκε αισθητά.

Επίδραση των περιοριστικών μέτρων στην επικράτεια της Ιταλίας Στην Ιταλία, η εφαρμογή των περιοριστικών μέτρων είχε ως άμεση συνέπεια τη μείωση των ροών μεταξύ των επαρχιών της. Η Ιταλική κυβέρνηση, εφάρμοσε σταδιακά όλο και αυστηρότερα μέτρα περιορισμού. Όπως φαίνεται και στο σχήμα 4.9, οι σταδιακές

αυστηροποιήσεις μέτρων προκαλούσαν μεγάλη μείωση των συνολικών ροών μεταξύ περιοχών. Στο διάστημα μεταξύ 29/03 και 19/04, τον μήνα των πιο μεγάλων απαγορεύσεων, διακρίνεται πως ο μέσος όρος εβδομαδιαίων ροών ήταν κάτω από το 1/3 των ροών πριν επέλθει η πανδημία.

Αξιοσημείωτο επίσης είναι πως η χρονική στιγμή θέσπισης νέων μέτρων συμπίπτει με τις μεγάλες αυξήσεις του συνολικού αριθμού νέων κρουσμάτων. Ως συνέπεια εφαρμογής των μέτρων, διαπιστώνεται πως ο ρυθμός αύξησης των νέων κρουσμάτων μειώνεται. Τα μέτρα της 22ης Μαρτίου, είναι αυτά που φαίνεται να συνδέονται με την πτώση του αριθμού των νέων ασθενών.

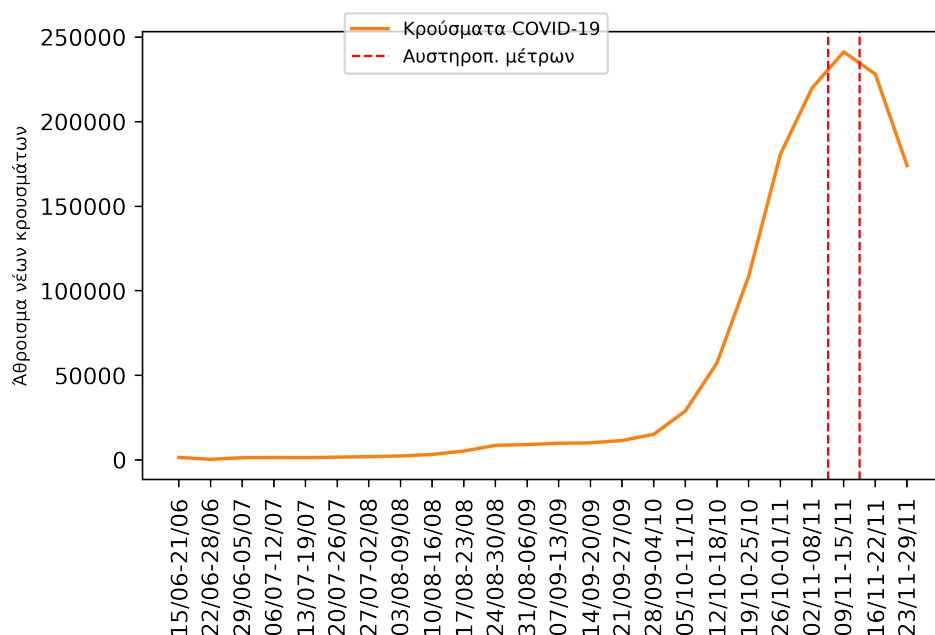


Σχήμα 4.9: Εβδομαδιαίος μέσος όρος ροών μεταξύ των επαρχιών και άθροισμα εβδομαδιαίων νέων κρουσμάτων, για την περίοδο από 20/1/2020 έως και 12/7/2020.

Η σταδιακή άρση των μέτρων, που ξεκίνησε στις 4 Μαΐου, έφερε σταδιακή αύξηση των κινήσεων μεταξύ επαρχιών. Ως και την 21/06, όπου έχουμε καταγραφές κινητικότητας μεταξύ των επαρχιών, φαίνεται πως οι κινήσεις έτειναν να επανέλθουν στα επίπεδα προ κορονοϊού. Αν συνυπολογιστεί και η περίοδος διακοπών, που συνήθως είναι μεταξύ Ιουλίου και Αυγούστου, είναι εύλογο να υποτεθεί πως οι ροές επανήλθαν ή ακόμη και ξεπέρασαν τις ροές που είχαν καταγραφεί πριν την

πανδημία.

Η άρση των μέτρων σε συνδυασμό ίσως με τον τουρισμό ανέτρεψε την κάμψη νέων φορέων COVID-19 που παρατηρήθηκε στο σχήμα 4.9. Στο σχήμα 4.10, φαίνεται πως ο αριθμός των ασθενών ολοένα και αυξανόταν, έχοντας ως συνέπεια την κυβέρνηση της Ιταλίας να αναγκαστεί και πάλι να επιβάλλει περιοριστικά μέτρα τον Νοέμβριο του 2020.



Σχήμα 4.10: Εβδομαδιαίο άθροισμα νέων κρουσμάτων COVID-19 στην Ιταλία, από 15/06 έως και 29/11.

4.3.4 Σύγκριση μετακινήσεων πριν και κατά τη διάρκεια της πανδημίας

Για τη σύγκριση των ροών μετακίνησης μεταξύ των επαρχιών της Ιταλίας πριν και κατά τη διάρκεια της πανδημίας, εξήχθησαν κατάλληλα σύνολα δεδομένων και αξιοποιήθηκε μια μέθοδος σύγκρισης πινάκων προέλευσης-προορισμού.

Πίνακες προέλευσης- προορισμού Οι πίνακες προέλευσης-προορισμού (origin-destination matrices) αντιπροσωπεύουν κινήσεις μεταξύ γεωγραφικών σημείων. Συγκεκριμένα, αναφέρονται σε ροές (flows) προερχόμενες από μια προέλευση (origin) σε έναν προορισμό (destination).

Σύγκριση πινάκων προέλευσης-προορισμού Σε μια πρόσφατη έρευνα, οι Behara et al. [93] παρουσίασαν μια μέθοδο σύγκρισης πινάκων προέλευσης προορισμού που αξιοποιεί την απόσταση Λέβενσταιν (Levenshtein distance). Οι ερευνητές υπογράμμισαν πως η σύγκριση των πινάκων προέλευσης-προορισμού πρέπει να λαμβάνει υπόψιν της τόσο τις διαφορές των συνολικών ροών, αλλά και την κατανομή των ροών αυτών ως προς τα εκάστοτε ζεύγη προελεύσεων-προορισμού.

Η κανονικοποιημένη απόσταση Λέβενσταιν για πίνακες προέλευσης προορισμού ή αλλιώς NLOD (Normalized Levenshtein distance for OD matrices), όπως αποκάλεσαν την μονάδα μέτρησης που κατασκεύασαν παρουσιάζει σημαντικά βελτιωμένη ευαισθησία σε σχέση με άλλες τεχνικές σύγκρισης πινάκων προέλευσης-προορισμού.

Η NLOD, είναι μια συνάρτηση της ομοιότητας δυο πινάκων, που λαμβάνει τιμές από 0 έως 1. Όσο μικρότερη η διαφορά (ή η απόσταση) των δύο πινάκων, τόσο μικρότερη και η διαφορά τους. Δυο ταυτόσημοι πίνακες θα έχουν NLOD 0, ενώ δυο αρκετά διαφορετικοί θα έχουν διαφορά που θα τείνει στο 1.

Στη συνέχεια, για κάθε εβδομάδα από 20/1/2020 έως και 21/6/2020, υπολογίστηκε η απόσταση Λέβενσταιν μεταξύ των επόμενων της. Τα αποτελέσματα παρατίθενται στον πίνακα 4.9. Παρατηρούμε πως όταν επιβάλλονται περιοριστικά μέτρα, η μέση NLOD μεταξύ της εβδομάδας προέλευσης και της εβδομάδας αναφοράς είναι αρκετά μεγάλη. Για παράδειγμα, η απόσταση της εβδομάδας 1 με τις επόμενες της σχεδόν τριπλασιάζεται όταν αρχίζουν να εφαρμόζονται αυστηρά περιοριστικά μέτρα.

Για τη βελτίωση της αναγνωσιμότητας του πίνακα, ομαδοποιήθηκαν και οπτικοποιήθηκαν κάποια αποτελέσματα.

Σύγκριση οχτώ πρώτων εβδομάδων με τις επόμενες τους Όπως φαίνεται και στο σχήμα 4.11, οι εβδομάδες 1-8, δηλαδή οι εβδομάδες από 20/01-26/1 έως 9/3-16/3 παρουσιάζουν παρόμοια μέση NLOD με τις επόμενες τους. Πιθανότατα αυτό οφείλεται στο γεγονός πως όλες οι εβδομάδες προέλευσης αναφέρονται σε περιόδους πριν την εφαρμογή σκληρών περιοριστικών μέτρων εξάπλωσης της ασθένειας COVID-19. Η 21η Μάρτη ήταν η πρώτη μέρα εφαρμογής πολύ αυστηρών μέτρων σε όλη την Ιταλική επικράτεια, δηλαδή τα μέτρα επηρέασαν την αμέσως επόμενη εβδομάδα (9-15/3).

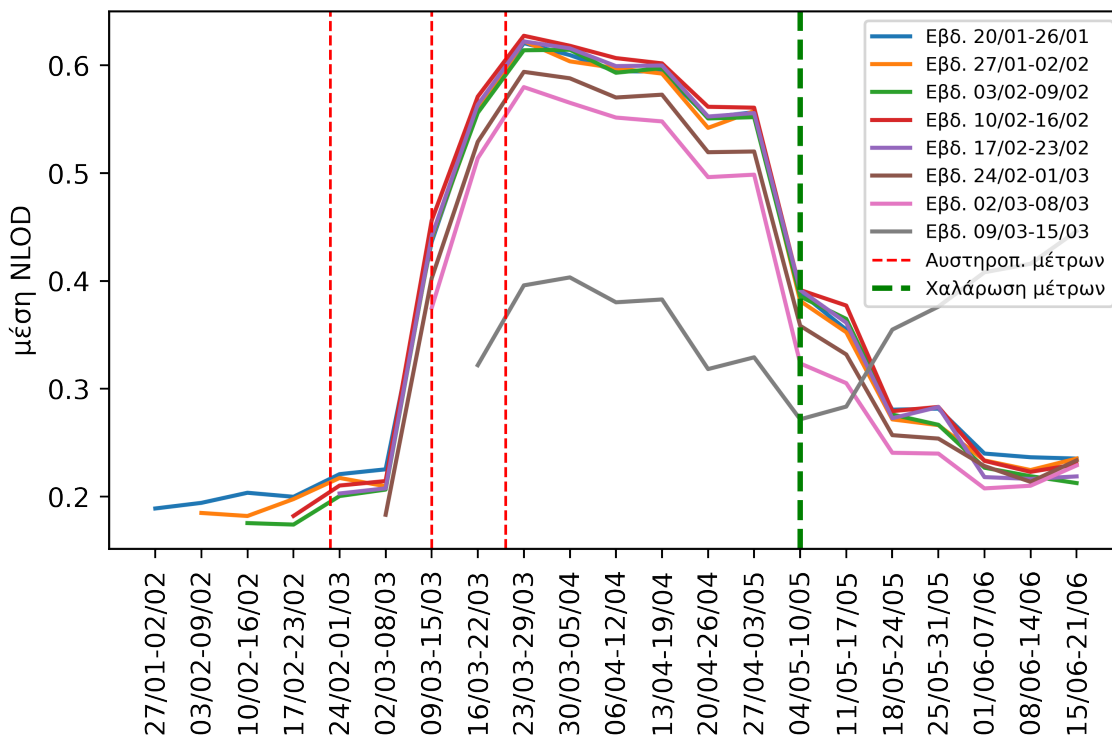
Παρατηρούμε πως όσο εφαρμόζονται περιοριστικά μέτρα και μειώνεται η

	Εβ.2	Εβ.3	Εβ.4	Εβ.5	Εβ.6	Εβ.7	Εβ.8	Εβ.9	Εβ.10	Εβ.11	Εβ.12	Εβ.13	Εβ.14	Εβ.15	Εβ.16	Εβ.17	Εβ.18	Εβ.19	Εβ.20	Εβ.21	Εβ.22
Εβ.1	0.189	0.194	0.204	0.2	0.221	0.225	0.436	0.56	0.621	0.61	0.595	0.594	0.552	0.557	0.389	0.355	0.281	0.282	0.24	0.236	0.235
Εβ.2		0.185	0.182	0.198	0.217	0.21	0.439	0.557	0.623	0.604	0.598	0.592	0.542	0.558	0.381	0.352	0.272	0.266	0.234	0.225	0.235
Εβ.3			0.175	0.174	0.2	0.207	0.441	0.556	0.614	0.615	0.593	0.598	0.551	0.552	0.386	0.365	0.276	0.267	0.227	0.219	0.213
Εβ.4				0.182	0.21	0.214	0.456	0.571	0.628	0.618	0.607	0.602	0.562	0.561	0.392	0.377	0.279	0.283	0.233	0.223	0.23
Εβ.5					0.203	0.208	0.442	0.564	0.622	0.616	0.599	0.6	0.553	0.556	0.392	0.361	0.273	0.283	0.218	0.216	0.219
Εβ.6						0.183	0.403	0.529	0.594	0.588	0.57	0.573	0.52	0.52	0.358	0.332	0.257	0.254	0.228	0.214	0.234
Εβ.7							0.375	0.514	0.58	0.565	0.552	0.548	0.496	0.499	0.324	0.305	0.241	0.24	0.208	0.21	0.229
Εβ.8								0.322	0.396	0.403	0.38	0.383	0.318	0.329	0.272	0.284	0.355	0.376	0.408	0.416	0.446
Εβ.9									0.282	0.296	0.294	0.296	0.267	0.293	0.372	0.387	0.463	0.482	0.536	0.534	0.556
Εβ.10										0.257	0.271	0.281	0.308	0.308	0.442	0.456	0.531	0.546	0.588	0.594	0.606
Εβ.11											0.27	0.247	0.312	0.295	0.427	0.433	0.516	0.534	0.575	0.588	0.603
Εβ.12												0.271	0.284	0.28	0.399	0.426	0.507	0.519	0.566	0.571	0.586
Εβ.13													0.264	0.271	0.396	0.413	0.507	0.516	0.557	0.569	0.59
Εβ.14														0.246	0.344	0.36	0.455	0.468	0.519	0.523	0.553
Εβ.15															0.344	0.37	0.453	0.462	0.51	0.516	0.544
Εβ.16																0.221	0.287	0.288	0.339	0.338	0.381
Εβ.17																	0.246	0.277	0.314	0.332	0.356
Εβ.18																		0.181	0.225	0.237	0.264
Εβ.19																			0.227	0.224	0.27
Εβ.20																				0.188	0.208
Εβ.21																					0.199

Πίνακας 4.9: Συγκριτικός πίνακας της μέσης απόστασης Λεβενστάιν μεταξύ των μελετώμενων εβδομάδων, από 20/1/2020 έως και 21/06/2020.

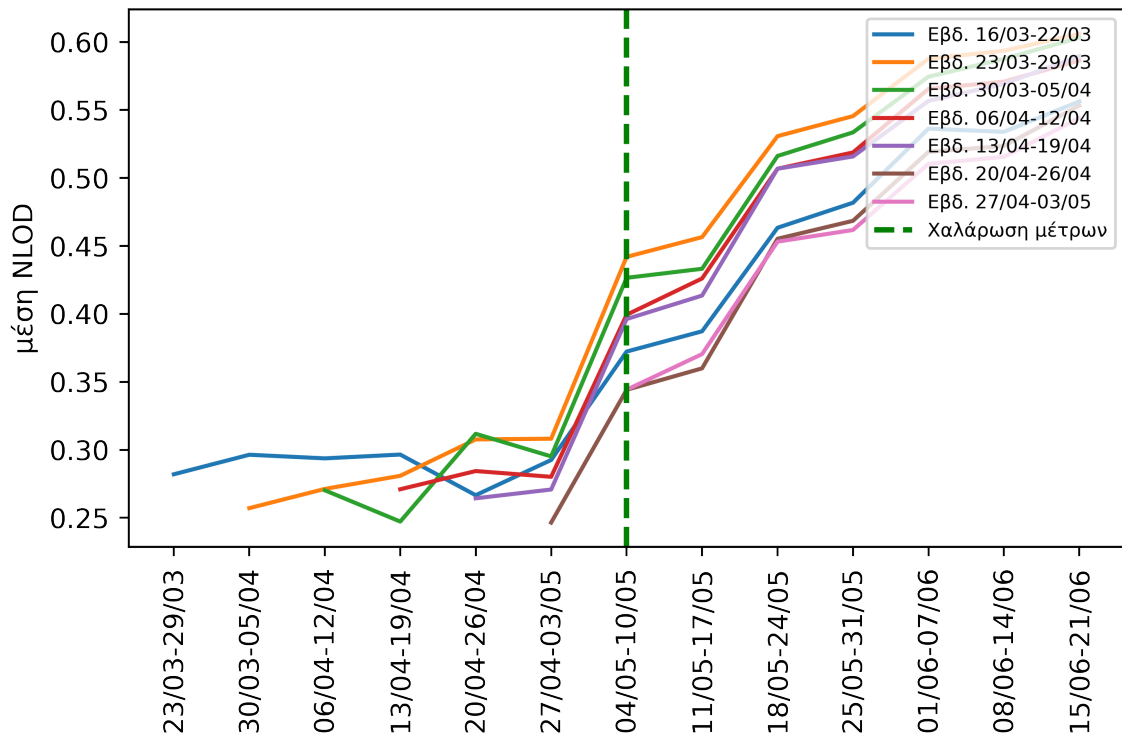
μετακίνηση μεταξύ των επαρχιών της Ιταλίας, τόσο περισσότερο διαφέρουν οι πίνακες προέλευσης προορισμού. Η διαφορά αυτή αποτυπώνεται στο σχήμα 4.11 με την αύξηση της τιμής της μέσης κανονικοποιημένης απόστασης Λέβενσταιν. Από την εβδομάδα που αίρονται τα μέτρα και έπειτα, διαπιστώνεται σημαντική μείωση της απόστασης. Το γεγονός αυτό επιβεβαιώνει πως πράγματι οι ροές μεταξύ των επαρχιών επαναφέρονται σταδιακά στην κανονικότητα.

Η εβδομάδα 9-15/3 έχει χαμηλότερη μέση κανονικοποιημένη απόσταση ως την 4η Μαΐου σε σχέση με τις υπόλοιπες, καθώς εκκινεί σε συνθήκες μερικού περιορισμού, ενώ όταν επανέρχονται οι ροές σε φυσιολογικά επίπεδα εμφανίζει μεγαλύτερη διαφορά για τον ίδιο λόγο.



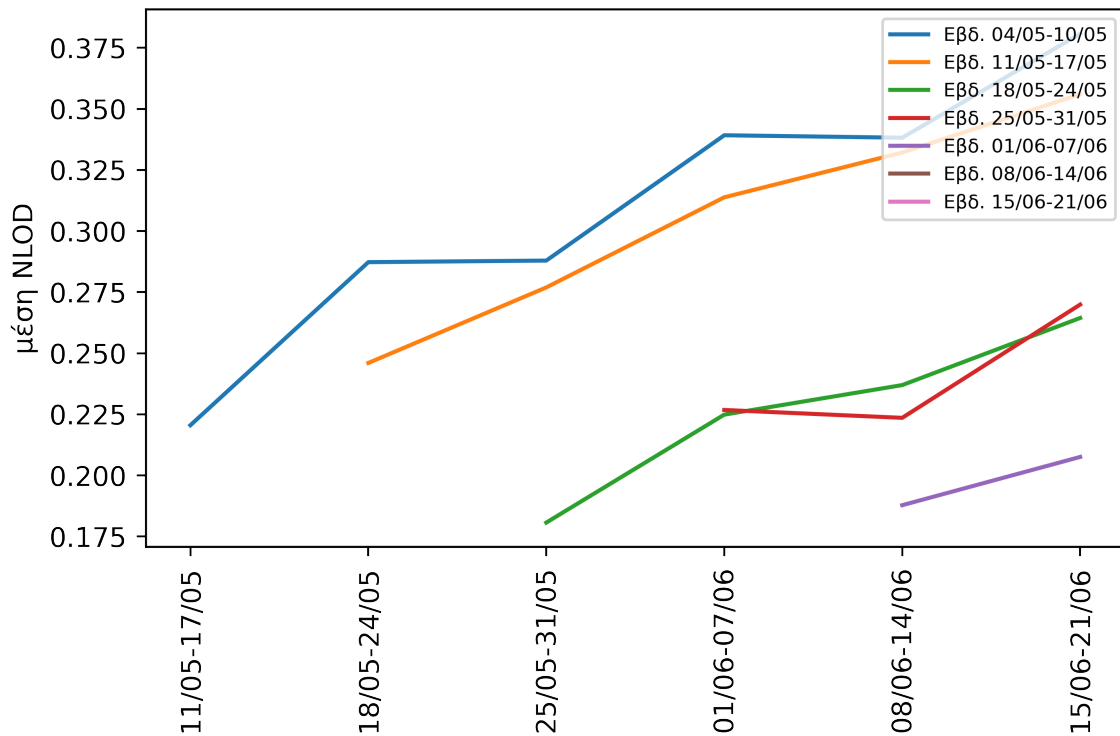
Σχήμα 4.11: Μέση κανονικοποιημένη απόσταση Λέβενσταιν (mean NLOD) μετακινήσεων μεταξύ περιοχών. Στο σχήμα, συγκρίνονται οι οχτώ πρώτες εβδομάδες του διαστήματος που μελετάται με τις εβδομάδες που ακολουθούν.

Σύγκριση 7 επόμενων εβδομάδων Στο σχήμα 4.12, που αφορά όλες τις μέσες NLOD των εβδομάδων πριν την άρση των μέτρων διαπιστώνουμε πως οι εβδομάδες αυτές παρουσιάζουν παρόμοια απόσταση σε σχέση με τις επόμενες τους. Πιο αναλυτικά, μέχρι και πριν την άρση του απαγορευτικού η διακύμανση των ροών είναι μεταξύ 0.25-0.35, ενώ όταν χαλαρώνουν τα μέτρα η μέση απόσταση αυξάνεται και πλησιάζει το 0.6. Αυτό οφείλεται στο γεγονός πως όσο αυξάνονται οι ροές μεταξύ των επαρχιών της Ιταλίας, τόσο περισσότερο θα διαφέρουν οι εβδομάδες κατά τη διάρκεια του lockdown με τις εβδομάδες εκτός των περιορισμών.



Σχήμα 4.12: Μέση κανονικοποιημένη απόσταση Λέβενσταιν (mean NLOD) μετακινήσεων μεταξύ περιοχών. Στο σχήμα συγκρίνονται οι εβδομάδες από 23/3 έως και 21/6/2020 με τις επόμενες τους.

Σύγκριση τελευταίων 7 εβδομάδων Οι τελευταίες 7 εβδομάδες αναπαριστώνται γραφικά στο σχήμα 4.13, δηλαδή οι τελευταίες εβδομάδες του συνόλου δεδομένων βρίσκονται όλες χρονικά σε περίοδο δίχως αυστηρά περιοριστικά μέτρα. Για το λόγο αυτό, παρουσιάζουν μικρές διακυμάνσεις σε σχέση με τις επόμενες. Παρατηρούμε ωστόσο μια αυξητική τάση της μέσης NLOD, η οποία οφείλεται στην αύξηση των ροών μεταξύ των επαρχιών. Η εβδομάδα που βρίσκεται πιο κοντά στην ημερομηνία άρσης των μέτρων είναι και αυτή που εμφανίζει και τη μεγαλύτερη μέση κανονικοποιημένη απόσταση Λέβενσταιν, καθώς εμφανίζει μέση NLOD ίση με 0.544.



Σχήμα 4.13: Μέση κανονικοποιημένη απόσταση Λέβενσταιν (mean NLOD) μετακινήσεων μεταξύ περιοχών. Στο σχήμα συγκρίνονται οι επτά τελευταίες εβδομάδες με τις επόμενες τους.

4.3.5 Δημιουργία συνθετικών ροών

Στην ενότητα αυτή δημιουργούνται τεχνητά σύνολα δεδομένων αξιοποιώντας το βαρυτικό μοντέλο προσομοίωσης των ροών μεταξύ περιοχών και αναζητούνται κατάλληλες παράμετροι για την προσομοίωση συνθηκών πανδημίας και περιορισμού κίνησης.

Περιγραφή μοντέλου που υλοποιήθηκε Για την προσομοίωση ροών μεταξύ περιοχών χρησιμοποιήθηκε η συνάρτηση `skmob.models.gravity.Gravity` της βιβλιοθήκης `scikit-mobility`. Η συνάρτηση αυτή ακολουθεί τη βασική διατύπωση του βαρυτικού μοντέλου. Θεωρεί δηλαδή την πιθανότητα μετακίνησης T_{ij} από μια περιοχή i σε μια περιοχή j ως:

$$T_{ij} \propto \frac{P_i P_j}{r_{ij}}$$

όπου P_i και P_j ο πληθυσμός των περιοχών i και j αντίστοιχα και r_{ij} η απόσταση μεταξύ του i και j . Το βαρυτικό μοντέλο υποθέτει πως ο αριθμός των διαδρομών που εκκινούν από την περιοχή i είναι αναλογικός του πληθυσμού της περιοχής αυτής (P_i), ενώ αντίστοιχα η κατάληξη των διαδρομών αυτών είναι αναλογική του πληθυσμού της περιοχής j . Τέλος, προσμετράται και το κόστος της διαδρομής αυτής ως προς την απόσταση που διανύθηκε [9]. Μια άλλη διατύπωση του βαρυτικού μοντέλου που προσμετρά τις παραμέτρους αυτές είναι η ακόλουθη:

$$T_{ij} = Km_i m_j f(r_{ij}),$$

όπου K σταθερά, οι μάζες m_i και m_j σχετίζονται με τον αριθμό των διαδρομών που εκκινούν της περιοχής i ή έλκονται από την j . Το $f(r_{ij})$, το οποίο ονομάζεται συνάρτηση αναχαίτισης (deterrence function) είναι μια φθίνουσα συνάρτηση της απόστασης.

Τα **περιορισμένα βαρυτικά μοντέλα** (constrained gravity models) είναι αυτά στα οποία ο αριθμός των ανθρώπων που ξεκινούν από μια περιοχή i είναι περιορισμένος σε μια γνωστή ποσότητα O_i . Στις περιπτώσεις αυτές, το βαρυτικό μοντέλο καλείται να υπολογίσει τον προορισμό. Τα περιορισμένα βαρυτικά μοντέλα διατυπώνονται ως εξής:

$$T_{ij} = K_i O_i m_j f(r_{ij}) = O_i \frac{m_i f(r_{ij})}{\sum_k m_k f(r_{ik})}$$

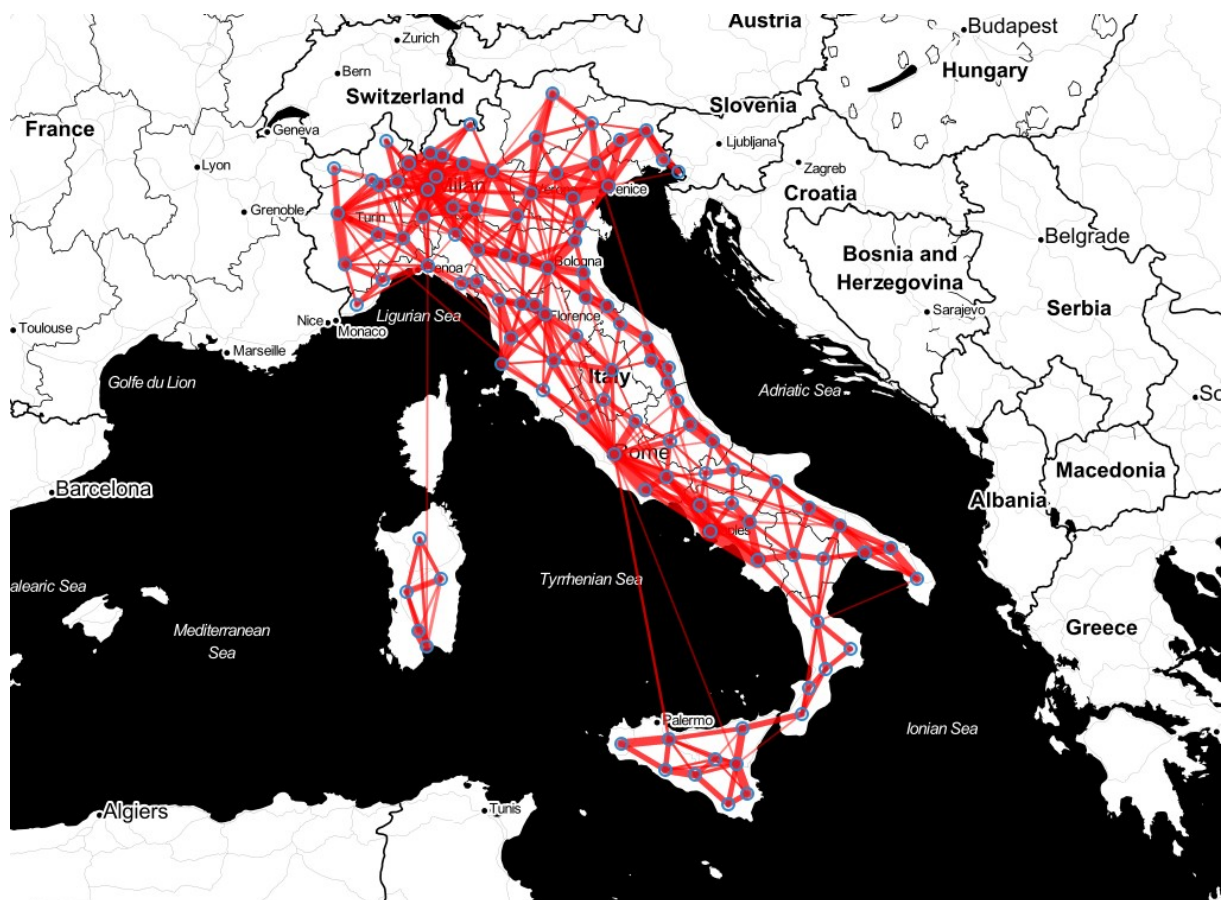
Στη μορφή αυτή, οι σταθερές αναλογικότητας K_i εξαρτώνται από την τοποθεσία εκκίνησης και την απόστασής της ως προς τα άλλα μέρη που προσμετρώνται.

Στην πειραματική εφαρμογή σε εμπειρικά δεδομένα, εφαρμόστηκε το περιορισμένο βαρυτικό μοντέλο, τόσο για την προσομοίωση ροών πριν την έναρξη της πανδημίας, όσο και μετά την έναρξη αυτής.

Προσομοίωση ροών πριν την έναρξη της πανδημίας Για την προσομοίωση πριν την έναρξη της πανδημίας, αξιοποιήθηκαν τα σύνολα δεδομένων που δημιουργήθηκαν στην ενότητα 4.3.2. Συγκεκριμένα, χρησιμοποιήθηκαν τα παρακάτω σύνολα δεδομένων:

1. Το pandas dataframe με τις πληροφορίες γεωμετρίας και πληθυσμού (spatial tessellation)
2. Το flowdataframe που περιείχε τις ροές προέλευσης-προορισμού

Αρχικά, από το σύνολο των ροών του flowdataframe, απομονώθηκαν οι ροές μιας εβδομάδας πριν την έναρξη της πανδημίας. Συγκεκριμένα, απομονώθηκαν οι ροές της περιόδου από 20/1/2020 έως και 26/1/2020. Στη συνέχεια, από τις ροές του flowdataframe υπολογίστηκε για κάθε επαρχία ο αριθμός των ατόμων που εξέρχονταν αυτής (O_i). Η πληροφορία αυτή καταχωρήθηκε στο σύνολο δεδομένων γεωμετρίας και πληθυσμού (tesselation). Μια οπτική απεικόνιση των ροών προέλευσης προορισμού της περιόδου 20-26/1/2020 φαίνεται στο σχήμα 4.14.



Σχήμα 4.14: Ροές άνω των 10.000 ατόμων μεταξύ των επαρχιών της Ιταλίας από 20 έως 26/1/2020.

Έπειτα, εκτελέστηκαν συνολικά 1066 προσομοιώσεις του βαρυτικού μοντέλου για διάφορες παραμέτρους της συνάρτησης αναχαίτισης και της συνάφειας του προορισμού. Δοκιμάστηκαν όλοι οι συνδυασμοί των δυο παραμέτρων για συνάρτηση αναχαίτισης από -4 έως και -2 με προσαύξηση 0.05 (συνολικά 41 τιμές) και της συνάφειας προορισμού από 0 έως 1.5 (συνολικά 26 τιμές) με προσαύξηση 0.05.

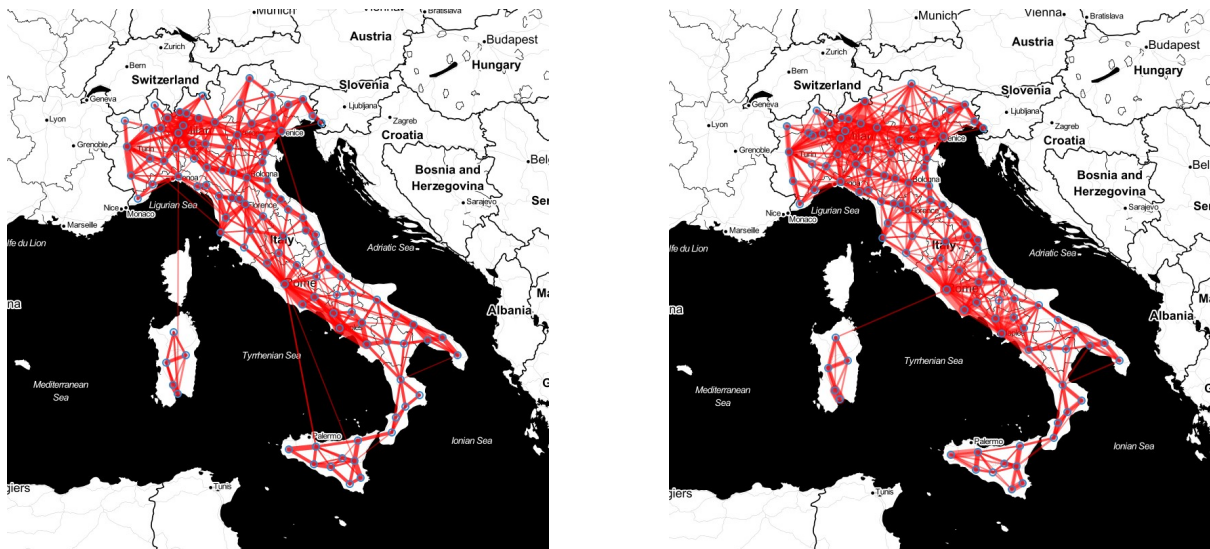
Οι πίνακες προέλευσης-προορισμού που δημιουργήθηκαν συγκρίθηκαν με τους πραγματικούς πίνακες βάσει της απόστασης Λέβενσταιν (βλ. ενότητα 4.3.4). Τα 20 καλύτερα αποτελέσματα της σύγκρισης φαίνονται στον πίνακα 4.10. Σε σχέση με τις βασικές παραμέτρους του βαρυτικού μοντέλου (για το οποίο ορίζεται η τιμή -2 ως όρισμα συνάρτησης αναχαίτισης και 1 για τη συνάφεια προορισμού) καθώς και τους συνδυασμούς παραμέτρων που εξετάστηκαν, διαπιστώνεται πως το βαρυτικό μοντέλο παρήγαγε τα καλύτερα αποτελέσματα για αρκετά μικρότερη τιμή της παραμέτρου αναχαίτισης. Ο μικρότερος συντελεστής αναχαίτισης μεταφράζεται και σε μικρότερη πιθανότητα να συμβεί μια μακρινή μετακίνηση μεταξύ των επαρχιών της Ιταλίας.

Η σημαντική διαφορά της μέσης κανονικοποιημένης απόστασης Λέβενσταιν ενδέχεται να οφείλεται στην ιδιαίτερη μορφολογία της Ιταλίας, καθώς το βαρυτικό μοντέλο συνυπολογίζει την ευθεία απόσταση μεταξύ δυο περιοχών και δε συνυπολογίζει φυσικά εμπόδια (όπως λ.χ. υψηλά βουνά και θάλασσες). Επίσης, πολλές ροές μεταξύ περιοχών της Ιταλίας που φαίνεται να συνέβησαν μέσω αεροπορικών διαδρομών δεν ήταν εύκολο να αποτυπωθούν μέσω του βαρυτικού μοντέλου.

Ορίσματα συνάρτησης αναχαίτισης	Συνάφεια προορισμού	μέση NLOD
-3.35	1.1	0.494
-3.5	1.15	0.494
-3.5	1.15	0.494
-3.2	1.05	0.495
-3.4	1.1	0.495
-3.05	1	0.495
-3.05	1	0.495
-3.25	1.05	0.495
-3.55	1.15	0.495
-3.3	1	0.496
-3.3	1	0.496
-3.45	1.05	0.496
-3.4	1	0.496
-3.4	1	0.496
-3.1	0.95	0.497
-2.9	0.95	0.497
-3.6	1.1	0.497
-3.3	1.1	0.497
-3.45	1.15	0.497
-3.55	1.05	0.497

Πίνακας 4.10: Μέση κανονικοποιημένη απόσταση Λέβενσταιν μεταξύ των πραγματικών και των συνθετικών δεδομένων που παρήχθησαν για την εβδομάδα 20-26/1/2020.

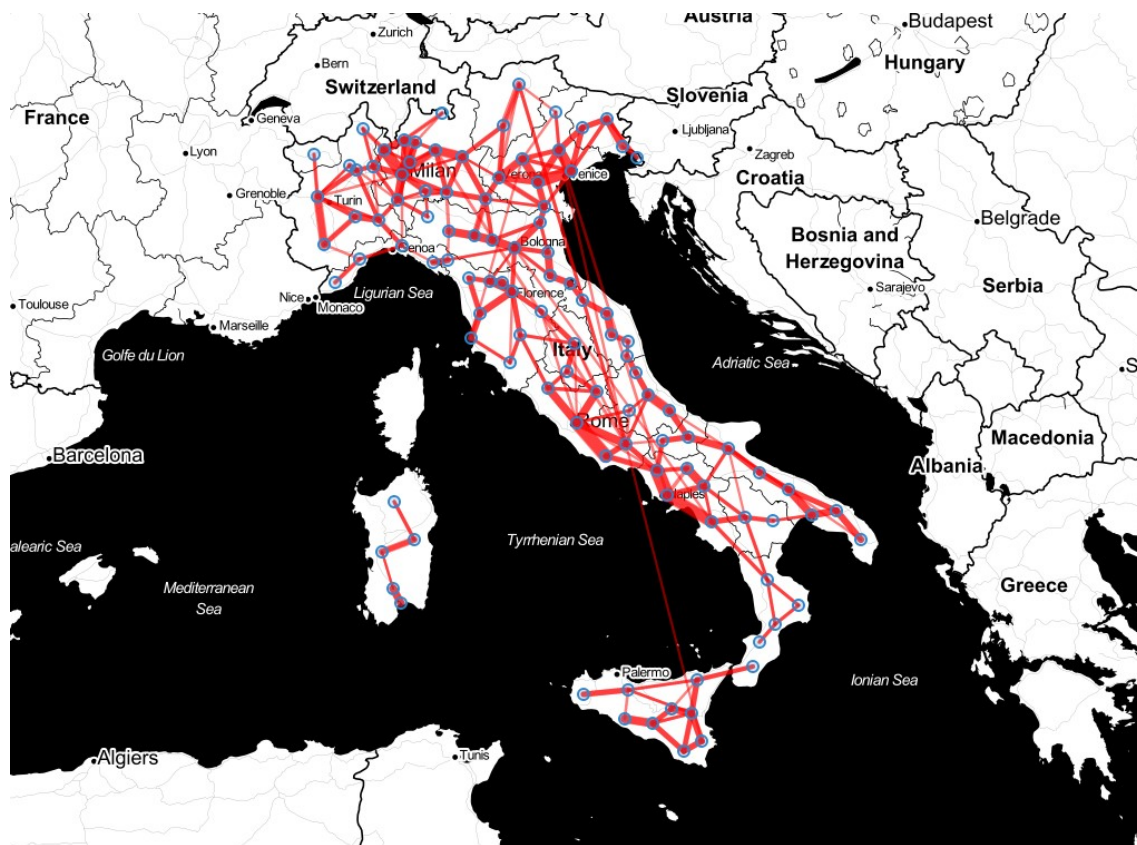
Τέλος, στο σχήμα 4.15 παρουσιάζονται οι ροές άνω των 10.000 ατόμων μεταξύ των επαρχιών της Ιταλίας για το διάστημα 20-26/1/2020 καθώς και οι συνθετικές ροές που προέκυψαν από την εφαρμογή του βαρυτικού μοντέλου. Αρκετές από τις αεροπορικές οδούς της Ιταλίας δεν φαίνεται να δημιουργήθηκαν από το βαρυτικό μοντέλο, επιβεβαιώνοντας τη διαπίστωση πως ο καλύτερος συνδυασμός παραμέτρων που εξετάστηκαν ευνοούσε τις κοντινές μετακινήσεις μεταξύ επαρχιών.



Σχήμα 4.15: Οπτικοποίηση των πραγματικών ρών άνω των 10.000 ατόμων μεταξύ των επαρχιών της Ιταλίας από 20-26/1/2020 (αριστερά) και των συνθετικών ρών που δημιουργήθηκαν από το βαρυτικό μοντέλο (δεξιά).

Προσομοίωση ρών μετά την έναρξη της πανδημίας Για την προσομοίωση των ρών κατά τη διάρκεια της πανδημίας, αξιοποιήθηκαν τα σύνολα δεδομένων που δημιουργήθηκαν στην ενότητα 4.3.2, σε αντιστοιχία με την προηγούμενη παράγραφο.

Από το σύνολο των ρών του flowdataframe, απομονώθηκαν οι ροές μιας εβδομάδας κατά τη διάρκεια της πανδημίας. Συγκεκριμένα, απομονώθηκαν οι ροές της περιόδου από 30/3 έως και 5/4/2020. Στη συνέχεια, από τις ροές του flowdataframe υπολογίστηκε για κάθε επαρχία ο αριθμός των ατόμων που εξέρχονταν αυτής (O_i). Η πληροφορία αυτή καταχωρήθηκε στο σύνολο δεδομένων γεωμετρίας και πληθυσμού (tesselation). Μια οπτική απεικόνιση των ρών προέλευσης προορισμού της περιόδου 30/3-5/4/2020 φαίνεται στο σχήμα 4.16. Σε σχέση με το σχήμα 4.14, διαπιστώνεται πως το πλήθος των ρών άνω των 10.000 ατόμων κατά τον Μάρτιο είναι αρκετά μειωμένο. Αυτό, προφανώς συνέβη λόγω των αυστηρών περιοριστικών μέτρων όσον αφορά τη μετακίνηση μεταξύ των επαρχιών της Ιταλίας.



Σχήμα 4.16: Ροές άνω των 10.000 ατόμων μεταξύ των επαρχιών της Ιταλίας από 30/3 έως 5/4/2020.

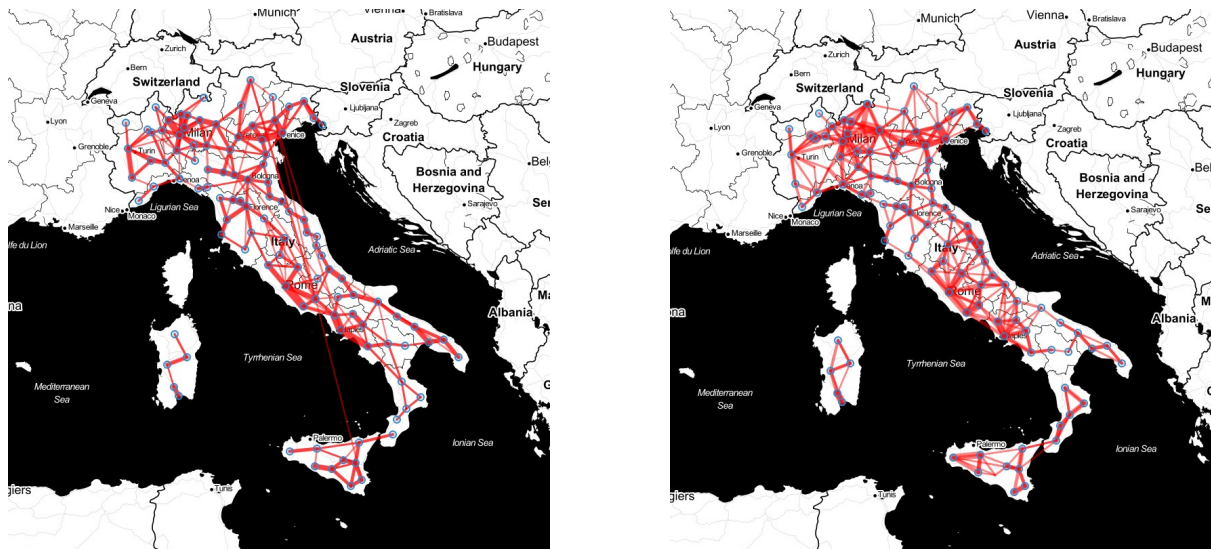
Σε πλήρη αντιστοιχία με την προηγούμενη παράγραφο, εκτελέστηκαν συνολικά 1066 προσομοιώσεις του βαρυτικού μοντέλου για διάφορες παραμέτρους της συνάρτησης αναχαίτισης και της συνάφειας του προορισμού. Δοκιμάστηκαν όλοι οι συνδυασμοί των δυο παραμέτρων για συνάρτηση αναχαίτισης από -4 έως και -2 με προσαύξηση 0.05 (συνολικά 41 τιμές) και της συνάφειας προορισμού από 0 έως 1.5 (συνολικά 26 τιμές) με προσαύξηση 0.05.

Οι πίνακες προέλευσης-προορισμού που δημιουργήθηκαν, συγκρίθηκαν με τις πραγματικές ροές βάσει της απόστασης Λέβενσταιν. Τα 20 καλύτερα αποτελέσματα της σύγκρισης φαίνονται στον πίνακα 4.11. Σε σχέση με τις βασικές παραμέτρους του βαρυτικού μοντέλου, διαπιστώνεται πως καλύτερα αποτελέσματα προκύπτουν από μια αρκετά μεγαλύτερη παράμετρο της συνάρτησης αναχαίτισης. Όσον αφορά τις παραμέτρους που εξετάστηκαν, διαπιστώνεται πως το βαρυτικό μοντέλο παράγει πιο αξιόπιστα αποτελέσματα ευνοώντας τις κοντινές μετακινήσεις σε σχέση με τις πιο μακρινές.

Ορίσματα συνάρτησης αναχαίτισης	Συνάφεια προορισμού	μέση NLOD
-2.9	0.85	0.513
-2.85	0.95	0.513
-3	1	0.513
-2.75	0.8	0.513
-2.7	0.9	0.513
-2.85	0.85	0.514
-3	0.9	0.514
-2.55	0.85	0.514
-2.95	0.9	0.514
-2.7	0.8	0.514
-2.8	0.85	0.514
-2.65	0.8	0.515
-2.8	0.95	0.515
-2.95	1	0.515
-2.65	0.9	0.515
-2.9	0.9	0.515
-2.75	0.85	0.516
-2.4	0.8	0.516
-2.8	0.9	0.516
-2.95	0.95	0.516

Πίνακας 4.11: Μέση κανονικοποιημένη απόσταση Λέβενσταιν μεταξύ των πραγματικών και των συνθετικών δεδομένων που παρήχθησαν για την εβδομάδα αυστηρών περιορισμών 30/4-5/4/2020.

Τέλος, στο σχήμα 4.17 παρουσιάζονται οι ροές άνω των 10.000 ατόμων μεταξύ των επαρχιών της Ιταλίας για το διάστημα 30/4-5/4/2020 καθώς και οι συνθετικές ροές που προέκυψαν από την εφαρμογή του βαρυτικού μοντέλου. Όπως βλέπουμε στα πραγματικά δεδομένα, μερικές επαρχίες της βόρειας Ιταλίας είχαν σημαντικές μετακινήσεις από και προς τις επαρχίες του νότου, ωστόσο το καλύτερο βαρυτικό μοντέλο βάσει των συνδυασμών παραμέτρων που εξετάστηκαν δεν προσμέτρησε



Σχήμα 4.17: Οπτικοποίηση των πραγματικών ροών άνω των 10.000 ατόμων μεταξύ των επαρχιών της Ιταλίας από 30/4-5/4/2020 (αριστερά) και των συνθετικών ροών που δημιουργήθηκαν από το βαρυτικό μοντέλο (δεξιά).

επαρκώς τις αεροπορικές διασυνδέσεις της Ιταλίας. Επίσης, το βαρυτικό μοντέλο δε δημιούργησε σημαντικές ροές πληθυσμών μεταξύ μακρινών επαρχιών, όπως χαρακτηριστικά φαίνεται μεταξύ των επαρχιών της κεντρικής και νότιας Ιταλίας.

4.4 Σύνοψη αποτελεσμάτων πειραματικής εφαρμογής

Στο κεφάλαιο αυτό συνοψίζονται τα αποτελέσματα της πειραματικής εφαρμογής σε εμπειρικά δεδομένα των τεχνικών ανάλυσης ανθρώπινης κινητικότητας.

4.4.1 Σύνοψη αποτελεσμάτων ατομικής κινητικότητας

Όσον αφορά την ανάλυση της ατομικής κινητικότητας, εξηγήθηκαν λεπτομερώς οι βιβλιοθήκες και τα σύνολα δεδομένων που αξιοποιήθηκαν. Στη συνέχεια, περιγράφηκε η προ-επεξεργασία και το φιλτράρισμα των δεδομένων ενώ υπολογίστηκαν βασικές μονάδες μέτρησης της ανθρώπινης κινητικότητας. Μεταξύ αυτών των μονάδων υπολογίστηκε η ακτίνα περιστροφής των χρηστών, η μέγιστη απόσταση που διένυσαν και η τυχαιότητα της κίνησής τους. Επίσης, βρέθηκαν οι μοναδικές τοποθεσίες που επισκέφτηκαν, υπολογίστηκε ο κίνδυνος απώλειας προσωπικών δεδομένων και τέλος υπολογίστηκαν οι κατοικίες των συμμετεχόντων

στην έρευνα.

Συγκρίνοντας τα αποτελέσματα της εφαρμογής του αλγορίθμου επισήμανσης σπιτιών με τις πραγματικές τοποθεσίες αυτών, διαπιστώθηκε μεγάλη συνάφεια και ακρίβεια της ανάλυσης. Σφάλματα που προέκυψαν λόγω εσφαλμένης θέσης, αποτυπώθηκαν οπτικά στον χάρτη και εξηγήθηκαν. Συμπερασματικά, η ανάλυση της ατομικής κινητικότητας, έχοντας επαρκή δεδομένα για τη διεξαγωγή της, δύναται να αποκαλύψει πολλές πληροφορίες για τη ζωή και τις συνήθειες των ανθρώπων. Πρέπει ωστόσο να υπογραμμιστεί πως για τη διαφύλαξη της ιδιωτικότητας των ατόμων των οποίων οι κινήσεις συλλέγονται, πρέπει να λαμβάνεται υπόψιν η ανωνυμοποίηση των δεδομένων και η εισαγωγή τεχνητού θορύβου στα σύνολα δεδομένων.

4.4.2 Σύνοψη αποτελεσμάτων κινήσεων πληθυσμών

Όσον αφορά τη σύνοψη των αποτελεσμάτων του μοντέλου κινήσεων πληθυσμών, για τη συλλογή και την προ-επεξεργασία των δεδομένων ακολουθήθηκε παρόμοια διαδικασία με αυτήν που περιγράφηκε στην προηγούμενη παράγραφο. Αξιοποιήθηκαν σύνολα δεδομένων της εξάπλωσης της ασθένειας COVID-19 στην Ιταλία καθώς και γεω-χωρικά και δημογραφικά δεδομένα των επαρχιών της Ιταλίας. Στη συνέχεια, ανακαλύφθηκε η επίδραση των περιοριστικών μέτρων για τον περιορισμό της εξάπλωσης της COVID-19 στην κίνηση μεταξύ των επαρχιών της Ιταλίας.

Αρχικά, υπολογίστηκαν οι συνολικές ροές εντός και εκτός της επαρχίας Lodi, της πρώτης επαρχίας που εφαρμόστηκαν μέτρα περιορισμού μετακίνησης από και προς αυτήν. Έπειτα, η ανάλυση επεκτάθηκε σε ολόκληρη την επικράτεια της Ιταλίας, όπου διαπιστώθηκε σημαντική μείωση των ροών μεταξύ των επαρχιών της κατά τη διάρκεια των περιοριστικών μέτρων και αύξηση αυτών μετά την άρση των μέτρων. Επίσης, εντοπίστηκε μια συσχέτιση του αριθμού των νέων κρουσμάτων COVID-19 και των ημερομηνιών εφαρμογής περιοριστικών μέτρων. Εφαρμόστηκε ο αλγόριθμος απόστασης πινάκων προέλευσης-προορισμού Λέβενσταιν "NLOD" στα δεδομένα κινητικότητας μεταξύ των επαρχιών για τις εβδομάδες πριν την έναρξη και κατά τη διάρκεια της πανδημίας. Ανακαλύφθηκε μικρή απόσταση Λέβενσταιν μεταξύ των εβδομάδων πριν την εφαρμογή περιοριστικών μέτρων, η οποία αυξανόταν κατά

την εφαρμογή περιορισμών και μειωνόταν κατά την άρση τους.

Τέλος, εφαρμόστηκε το βαρυτικό μοντέλο για την προσομοίωση των ροών μεταξύ των επαρχιών της Ιταλίας. Έγιναν δυο προσομοιώσεις, η μία αφορούσε μια τυχαία εβδομάδα πριν την εφαρμογή μέτρων, ενώ η δεύτερη μια εβδομάδα αυστηρού lockdown. Τα αποτελέσματα των συνθετικών ροών συγκρίθηκαν με τις πραγματικές βάσει της απόστασης Λέβενσταιν. Η απόσταση Λέβενσταιν μεταξύ των πραγματικών ροών και των συνθετικών ροών ήταν παρόμοια, τόσο σε συνθήκες πανδημίας όσο και εκτός αυτών. Για το συνδυασμό παραμέτρων που δοκιμάστηκαν, βρέθηκαν καλύτερα αποτελέσματα του βαρυτικού μοντέλου όταν ευνοούνταν οι κοντινές μετακινήσεις σε σχέση με τις πιο μακρινές. Η σύγκριση των τεχνητών συνόλων δεδομένων με τα πραγματικά, μέσω της κανονικοποιημένης απόστασης Λέβενσταιν κατέδειξε πως το βαρυτικό μοντέλο, ενώ αναπαράγει σε έναν ικανοποιητικό βαθμό τις ροές πληθυσμού μεταξύ επαρχιών, για την περίπτωση της Ιταλίας δεν αρκεί από μόνο του για την προσομοίωση των ροών μεταξύ των επαρχιών της. Αυτό συμβαίνει γιατί το βαρυτικό μοντέλο υπολογίζει την ευθεία απόσταση μεταξύ δυο περιοχών και δεν συνυπολογίζει την ιδιαίτερη μορφολογία μιας περιοχής και τις αεροπορικές διασυνδέσεις. Στην περίπτωση της Ιταλίας, μεγάλες οροσειρές και θάλασσα εμποδίζουν τις μετακινήσεις μεταξύ κοντινών επαρχιών, ενώ υπάρχουν τακτικές αεροπορικές συνδέσεις μεταξύ της βόρειας και της νότιας Ιταλίας. Το βαρυτικό μοντέλο, βάσει των παραμέτρων που εξετάστηκαν, υπολόγισε ροές μεταξύ επαρχιών που χωρίζονταν από οροσειρές και θάλασσα, ενώ δεν αποτύπωσε τις αεροπορικές οδούς της Ιταλίας.

5. Συμπεράσματα και Μελλοντικές

Προεκτάσεις

Μετά την καταγραφή των κυριότερων μεθόδων ανάλυσης ανθρώπινης κινητικότητας, των συνόλων δεδομένων που αυτές αξιοποιούν, την πειραματική εφαρμογή σε εμπειρικά δεδομένα και τη σύγκριση των αποτελεσμάτων, παρουσιάζονται συνολικά τα αποτελέσματα της διπλωματικής εργασίας και καταγράφονται δυνητικές μελλοντικές προεκτάσεις αυτής.

5.1 Συμπεράσματα

Η ανάλυση της ανθρώπινης κίνησης, τόσο σε ατομικό επίπεδο όσο και σε επίπεδο κινήσεων πληθυσμών παρουσιάζει πολλά οφέλη σε πολλά επιστημονικά και οικονομικά επίπεδα. Οι διάφορες τεχνικές ανάλυσής της, ανάλογα με το πεδίο στο οποίο εξειδικεύονται, μπορούν να ανακαλύψουν σημαντικές πληροφορίες σε μεγάλα και φαινομενικά ετερογενή σύνολα δεδομένων.

Η παρούσα διπλωματική εργασία ανέλυσε από θεωρητική σκοπιά τις χαρακτηριστικές τεχνικές και μονάδες μέτρησης της ανθρώπινης κινητικότητας που εμφανίζονται στη βιβλιογραφία. Επίσης, από πρακτική σκοπιά εξετάστηκε ένα σύνολο δεδομένων ατομικής κινητικότητας, και ένα σύνολο ρών πληθυσμού.

Σε **επίπεδο ατομικής κινητικότητας**, η ανάλυση των δεδομένων που προκύπτουν από τις έξυπνες συσκευές που η μεγάλη πλειοψηφία των ανθρώπων έχει στην κατοχή της μπορεί να αποκαλύψει αρκετές πληροφορίες για τη ζωή και τις συνήθειες τους. Πληροφορίες όπως η μέγιστη απόσταση που διανύουν μέσα στην ημέρα, οι καθημερινές διαδρομές που αυτοί πραγματοποιούν είναι άμεσα υπολογίσιμες. Είναι δυνατόν επίσης να εντοπιστούν και να επισημανθούν τοποθεσίες στις οποίες αυτοί συχνά βρίσκονται ή δραστηριότητες που αυτοί πραγματοποιούν και να προσομοιωθούν οι μελλοντικές τους μετακινήσεις. Σε **επίπεδο κινήσεων πληθυσμών**, η ανάλυση της κινητικότητας μπορεί να αποτυπώσει βάσει διάφορων μονάδων μέτρησης τον όγκο και τις τάσεις των ρών μεταξύ των περιοχών μιας

πόλης ή μιας ολόκληρης χώρας, ή ακόμη και να προσομοιώσει τις μελλοντικές ροές αυτών.

Όσον αφορά την εξάπλωση μεταδοτικών ασθενειών, τα εργαλεία που παρέχει η ανάλυση της ανθρώπινης κινητικότητας διευκολύνουν την πρόβλεψη της εξέλιξής της. Επίσης, διευκολύνουν τη θέσπιση στοχευμένων μέτρων περιορισμού της σε επίπεδο πληθυσμού, καθώς και τον εντοπισμό των κοινωνικών επαφών των φορέων της ασθένειας σε ατομικό επίπεδο.

5.2 Μελλοντικές προεκτάσεις

Η ανάλυση της ανθρώπινης κίνησης είναι αρκετά δημοφιλής σε πολλά επιστημονικά πεδία. Τα τελευταία χρόνια, πολλές νέες τεχνικές παρουσιάζονται ενώ οι ήδη υπάρχουσες τεχνικές συνεχώς βελτιώνονται.

Ορισμένες επεκτάσεις που θα μπορούσε να έχει η παρούσα διπλωματική εργασία είναι οι ακόλουθες:

- Εξέταση σύγχρονων τεχνικών ανάλυσης ανθρώπινης κίνησης πάνω σε δεδομένα ατομικής κινητικότητας και δεδομένα κινήσεων πληθυσμών μέσω της μηχανικής μάθησης
- Εφαρμογή κάποιου αλγόριθμου βασισμένου σε κανόνες για επισήμανση δραστηριοτήτων σε δεδομένα ατομικής κινητικότητας, προκειμένου να καταγραφούν περισσότερα στοιχεία για τις συνήθειες των ατόμων. Μια τέτοια ανάλυση, θα μπορούσε να αξιοποιηθεί για την επισήμανση εστιατορίων τα οποία αυτοί επισκέπτονται συχνότερα ή χώρους πέραν της κατοικίας τους στους οποίους περνούν τον περισσότερό τους χρόνο.
- Προσαρμογή του βαρυτικού μοντέλου ώστε να συνυπολογίζει την ιδιαίτερη μορφολογία μιας περιοχής. Συγκεκριμένα, μετατροπή της συνάρτησης αναχαιτίσης κατάλληλα ώστε να υπολογίζεται η οδική απόσταση δυο περιοχών έναντι της ευθείας απόστασης στο χάρτη. Επίσης, θα μπορούσαν να συμπληρωθούν και οι τακτικές αεροπορικές διασυνδέσεις μεταξύ των επαρχιών της Ιταλίας, θέτοντας ως όριο καθημερινών μετακινήσεων το σύνολο των μέγιστων επιβατών που μπορούν να ταξιδέψουν μέσω των αεροπλάνων.

- Αξιοποίηση πλουσιότερων συνόλων δεδομένων, ενδεχομένως με περισσότερα είδη πληροφορίας για τη βελτίωση της ποιότητας των παραγόμενων αποτελεσμάτων. Τέτοιες πληροφορίες θα μπορούσαν να είναι μεταδεδομένα ή δεδομένα τοποθεσίας από κάποιον άλλον αισθητήρα πέραν του GPS, ως εναλλακτική σε περίπτωση που η τοποθεσία του χρήστη δεν είναι δυνατό να εντοπιστεί μέσω των δορυφόρων γεωεντοπισμού.
- Δεδομένα μεγαλύτερης ή μικρότερης κλίμακας προκείμενου να ανακαλυφθεί κάποια συσχέτιση του όγκου των αξιοποιούμενων δεδομένων και της αποτελεσματικότητας των εκάστοτε τεχνικών ανάλυσης ανθρώπινης κίνησης. Η συσχέτιση αυτή, εφόσον βρισκόταν, θα μπορούσε να εφαρμοστεί για την αξιολόγηση των συνόλων δεδομένων ως προς την ικανότητά τους να παράξουν αξιόπιστες μετρήσεις κινητικότητας.

Βιβλιογραφία

- [1] Chaoming Song, Tal Koren, Pu Wang, and Albert-László Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010.
- [2] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37–37, 1996.
- [3] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.
- [4] KW Axhausen, S Schonfelder, J Wolf, M Oliveria, and U Samaga. Eighty weeks of gps traces, approaches to enriching trip information. In *Transportation Research Board Annual Meeting*, pages 1870–06. Citeseer, 2004.
- [5] Wei Huang, Songnian Li, Xintao Liu, and Yifang Ban. Predicting human mobility with activity changes. *International Journal of Geographical Information Science*, 29(9):1569–1587, 2015.
- [6] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [7] Emmanouil Alimpertis, Nikos Fasarakis-Hilliard, and Aggelos Bletsas. Community rf sensing for source localization. *IEEE Wireless Communications Letters*, 3(4):393–396, 2014.
- [8] Barbara Furletti, Paolo Cintia, Chiara Renso, and Laura Spinsanti. Inferring human activities from gps tracks. In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, pages 1–8, 2013.

- [9] Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.
- [10] Luca Pappalardo and Filippo Simini. Data-driven generation of spatio-temporal routines in human mobility. *Data Mining and Knowledge Discovery*, 32(3):787–829, 2018.
- [11] Kate E Jones, Nikkita G Patel, Marc A Levy, Adam Storeygard, Deborah Balk, John L Gittleman, and Peter Daszak. Global trends in emerging infectious diseases. *Nature*, 451(7181):990–993, 2008.
- [12] Sandro Meloni, Nicola Perra, Alex Arenas, Sergio Gómez, Yamir Moreno, and Alessandro Vespignani. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific reports*, 1:62, 2011.
- [13] Vivek Charu, Scott Zeger, Julia Gog, Ottar N Bjørnstad, Stephen Kissler, Lone Simonsen, Bryan T Grenfell, and Cécile Viboud. Human mobility and the spatial transmission of influenza in the united states. *PLoS computational biology*, 13(2):e1005382, 2017.
- [14] Nan Zhang, Hong Huang, Boni Su, Xun Ma, and Yuguo Li. A human behavior integrated hierarchical model of airborne disease transmission in a large city. *Building and Environment*, 127:211–220, 2018.
- [15] Matteo Chinazzi, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science*, 368(6489):395–400, 2020.
- [16] Jing Yuan, Yu Zheng, and Xing Xie. Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 186–194, 2012.

- [17] Jameson L Toole, Michael Ulm, Marta C González, and Dietmar Bauer. Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD international workshop on urban computing*, pages 1–8, 2012.
- [18] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.
- [19] Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. Returners and explorers dichotomy in human mobility. *Nature communications*, 6(1):1–8, 2015.
- [20] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *nature*, 453(7196):779–782, 2008.
- [21] Christian M Schneider, Vitaly Belik, Thomas Couronné, Zbigniew Smoreda, and Marta C González. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84):20130246, 2013.
- [22] Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, Seong Joon Kim, and Song Chong. On the levy-walk nature of human mobility. *IEEE/ACM transactions on networking*, 19(3):630–643, 2011.
- [23] Yves-Alexandre De Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3:1376, 2013.
- [24] John RB Palmer, Thomas J Espenshade, Frederic Bartumeus, Chang Y Chung, Necati Ercan Ozgencil, and Kathleen Li. New approaches to human mobility: Using mobile phones for demographic research. *Demography*, 50(3):1105–1128, 2013.
- [25] Maarten Vanhoof, Fernando Reis, Zbigniew Smoreda, and Thomas Ploetz. Detecting home locations from cdr data: introducing spatial uncertainty to the state-of-the-art. *arXiv preprint arXiv:1808.06398*, 2018.
- [26] Shan Jiang, Marta C Gonzalez, and Joseph Ferreira. Understanding the link between urban activity destinations and human travel pattern. 2011.

- [27] Xiao Liang, Jichang Zhao, Li Dong, and Ke Xu. Unraveling the origin of exponential law in intra-urban human mobility. *Scientific reports*, 3:2983, 2013.
- [28] Shan Jiang, Joseph Ferreira, and Marta C González. Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore. *IEEE Transactions on Big Data*, 3(2):208–219, 2017.
- [29] Nathalie E Williams, Timothy A Thomas, Matthew Dunbar, Nathan Eagle, and Adrian Dobra. Measures of human mobility using mobile phone records enhanced with gis data. *PloS one*, 10(7):e0133630, 2015.
- [30] Martin W Traunmueller, Nicholas Johnson, Awais Malik, and Constantine E Kontokosta. Digital footprints: Using wifi probe and locational data to analyze human mobility trajectories in cities. *Computers, Environment and Urban Systems*, 72:4–12, 2018.
- [31] Feixiong Luo, Guofeng Cao, Kevin Mulligan, and Xiang Li. Explore spatiotemporal and demographic characteristics of human mobility via twitter: A case study of chicago. *Applied Geography*, 70:11–25, 2016.
- [32] Md Shahadat Iqbal, Charisma F Choudhury, Pu Wang, and Marta C González. Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40:63–74, 2014.
- [33] GO MediaCT. The mobile movement: Understanding smartphone users. *Google Inc*, 2011.
- [34] Kai Zhao, Sasu Tarkoma, Siyuan Liu, and Huy Vo. Urban human mobility data mining: An overview. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 1911–1920. IEEE, 2016.
- [35] Natalia Andrienko, Gennady Andrienko, Georg Fuchs, and Piotr Jankowski. Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces. *Information Visualization*, 15(2):117–153, 2016.
- [36] Roberto Pellungrini, Luca Pappalardo, Francesca Pratesi, and Anna Monreale. A data mining approach to assess privacy risk in human mobility data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 9(3):1–27, 2017.

- [37] Ying Zhang. User mobility from the view of cellular data networks. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pages 1348–1356. IEEE, 2014.
- [38] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.
- [39] Albert Einstein. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der physik*, 4, 1905.
- [40] Chaogui Kang, Xiujun Ma, Daoqin Tong, and Yu Liu. Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and its Applications*, 391(4):1702–1717, 2012.
- [41] Juan de Dios Ortúzar and Luis G Willumsen. *Modelling transport*. John wiley & sons, 2011.
- [42] Andrei Andreevich Markov. The theory of algorithms. *Trudy Matematicheskogo Instituta Imeni VA Steklova*, 42:3–375, 1954.
- [43] Andrey Andreyevich Markov. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. *Dynamic probabilistic systems*, 1:552–577, 1971.
- [44] Nicola Bui, Matteo Cesana, S Amir Hosseini, Qi Liao, Ilaria Malanchini, and Joerg Widmer. A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques. *IEEE Communications Surveys & Tutorials*, 19(3):1790–1821, 2017.
- [45] Yuanyuan Qiao, Zhongwei Si, Yanting Zhang, Fehmi Ben Abdesslem, Xinyu Zhang, and Jie Yang. A hybrid markov-based model for human mobility prediction. *Neurocomputing*, 278:99–109, 2018.
- [46] Libo Song, David Kotz, Ravi Jain, and Xiaoning He. Evaluating next-cell predictors with extensive wi-fi mobility data. *IEEE transactions on mobile computing*, 5(12):1633–1649, 2006.

- [47] Daniel Ashbrook and Thad Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous computing*, 7(5):275–286, 2003.
- [48] Michael H Sun and Douglas M Blough. Mobility prediction using future knowledge. In *Proceedings of the 10th ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems*, pages 235–239, 2007.
- [49] Brian D Ziebart, Andrew L Maas, Anind K Dey, and J Andrew Bagnell. Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 322–331, 2008.
- [50] Meng Chen, Xiaohui Yu, and Yang Liu. Mining moving patterns for predicting next location. *Information Systems*, 54:156–168, 2015.
- [51] Dominique Barth, Samir Bellahsene, and Leila Kloul. Combining local and global profiles for mobility prediction in lte femtocells. In *Proceedings of the 15th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*, pages 333–342, 2012.
- [52] Alicia Rodriguez-Carrion, Carlos Garcia-Rubio, and Celeste Campo. Performance evaluation of lz-based location prediction algorithms in cellular networks. *IEEE Communications Letters*, 14(8):707–709, 2010.
- [53] Hongbo Si, Yue Wang, Jian Yuan, and Xiuming Shan. Mobility prediction in cellular network using hidden markov model. In *2010 7th IEEE consumer communications and networking conference*, pages 1–5. IEEE, 2010.
- [54] Juan Antonio Alvarez-Garcia, Juan Antonio Ortega, Luis Gonzalez-Abril, and Francisco Velasco. Trip destination prediction based on past gps log using a hidden markov model. *Expert Systems with Applications*, 37(12):8166–8171, 2010.
- [55] Libo Song, Udayan Deshpande, Ulaş C Kozat, David Kotz, and Ravi Jain. Predictability of wlan mobility and its effects on bandwidth provisioning. 2006.

- [56] Libo Song, David Kotz, Ravi Jain, and Xiaoning He. Evaluating location predictors with extensive wi-fi mobility data. In *IEEE Infocom 2004*, volume 2, pages 1414–1424. IEEE, 2004.
- [57] Vibhav Gogate, Rina Dechter, Bozhena Bidyuk, Craig Rindt, and James Marca. Modeling transportation routines using hybrid dynamic mixed networks. *arXiv preprint arXiv:1207.1384*, 2012.
- [58] Haoyi Xiong, Daqing Zhang, Daqiang Zhang, Vincent Gauthier, Kun Yang, and Monique Becker. Mpaas: Mobility prediction as a service in telecom cloud. *Information Systems Frontiers*, 16(1):59–75, 2014.
- [59] Hongzhi Shi, Yong Li, Hancheng Cao, Xiangxin Zhou, Chao Zhang, and Vassilis Kostakos. Semantics-aware hidden markov model for human mobility. *IEEE Transactions on Knowledge and Data Engineering*, 2019.
- [60] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. Deepmove: Predicting human mobility with attentional recurrent networks. In *Proceedings of the 2018 world wide web conference*, pages 1459–1468, 2018.
- [61] Eyal Ben Zion and Boaz Lerner. Identifying and predicting social lifestyles in people’s trajectories by neural networks. *EPJ Data Science*, 7(1):45, 2018.
- [62] Bin Jiang, Junjun Yin, and Sijian Zhao. Characterizing the human mobility pattern in a large street network. *Physical Review E*, 80(2):021136, 2009.
- [63] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Next place prediction using mobility markov chains. In *Proceedings of the first workshop on measurement, privacy, and mobility*, pages 1–6, 2012.
- [64] Junjie Jiang, Changchun Pan, Haichun Liu, and Genke Yang. Predicting human mobility based on location data modeled by markov chains. In *2016 Fourth International Conference on Ubiquitous Positioning, Indoor Navigation and Location Based Services (UPINLBS)*, pages 145–151. IEEE, 2016.
- [65] Constanze Ciavarella and Neil M Ferguson. Deriving fine-scale models of human mobility from aggregated origin-destination flow data. *PLOS Computational Biology*, 17(2):e1008588, 2021.

- [66] Mariano G Beiró, André Panisson, Michele Tizzoni, and Ciro Cattuto. Predicting human mobility through the assimilation of social media traces into mobility models. *EPJ Data Science*, 5:1–15, 2016.
- [67] Karim Keramat Jahromi, Matteo Zignani, Sabrina Gaito, and Gian Paolo Rossi. Simulating human mobility patterns in urban areas. *Simulation Modelling Practice and Theory*, 62:137–156, 2016.
- [68] Chen Zhao, An Zeng, and Chi Ho Yeung. Characteristics of human mobility patterns revealed by high-frequency cell-phone position data. *EPJ Data Science*, 10(1):5, 2021.
- [69] Mingxiao Li, Song Gao, Feng Lu, and Hengcai Zhang. Reconstruction of human movement trajectories from large-scale low-frequency mobile phone data. *Computers, Environment and Urban Systems*, 77:101346, 2019.
- [70] Qiuqian Lv, Yuanyuan Qiao, Nirwan Ansari, Jun Liu, and Jie Yang. Big data driven hidden markov model based individual mobility prediction at points of interest. *IEEE Transactions on Vehicular Technology*, 66(6):5204–5216, 2016.
- [71] Katarzyna Siła-Nowicka, Jan Vandrol, Taylor Oshan, Jed A Long, Urška Demšar, and A Stewart Fotheringham. Analysis of human mobility patterns from gps trajectories and contextual information. *International Journal of Geographical Information Science*, 30(5):881–906, 2016.
- [72] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.
- [73] Jeff Reback, Wes McKinney, jbrockmendel, Joris Van den Bossche, Tom Augspurger, Phillip Cloud, gfyong, Sinhrks, Adam Klein, Matthew Roeschke, Simon Hawkins, Jeff Tratner, Chang She, William Ayd, Terji Petersen, Marc Garcia, Jeremy Schendel, Andy Hayden, MomIsBestFriend, Vytautas Jancauskas, Pietro Battiston, Skipper Seabold, chris b1, h vetinari, Stephan Hoyer, Wouter Overmeire, alimcmaster1, Kaiqi Dong, Christopher Whelan, and Mortada Mehyar. *pandas-dev/pandas: Pandas 1.0.3*, March 2020.

- [74] Kelsey Jordahl, Joris Van den Bossche, Martin Fleischmann, James McBride, Jacob Wasserman, Jeffrey Gerard, Adrian Garcia Badaracco, Alan D. Snow, Jeff Tratner, Matthew Perry, Carson Farmer, Geir Arne Hjelle, Micah Cochran, Sean Gillies, Lucas Culbertson, Matt Bartos, Giacomo Caria, Nick Eubank, sangarshanan, Sergio Rey, maxalbert, Aleksey Bilogur, Brendan Ward, Christopher Ren, Dani Arribas-Bel, Flavin, Leah Wasser, Levi John Wolf, Martin Journois, and abonte. *geopandas/geopandas: v0.9.0*, February 2021.
- [75] Martin Journois, Filipe, Rob Story, Andrew Bird, Antonio Lima, Pavel Tyslacki, Justin Duke, Anand Patil, shankari, andrew giessel, uier κ ίεωες, Kyle Corbitt, Adam Pah, Aron Ahmadi, Jacob Wasserman, Dennis O'Brien, Kim Younggun, tahmed42, Roy Keyes, Lennart Fricke, and Elliott Sales de Andrade. *folium: Latest release*, November 2015.
- [76] Luca Pappalardo, Filippo Simini, Gianni Barlacchi, and Roberto Pellungrini. *scikit-mobility: a python library for the analysis, generation and risk assessment of mobility data*, 2019.
- [77] Yu Zheng. Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3):1–41, 2015.
- [78] Luca Pappalardo, Salvatore Rinzivillo, Zehui Qu, Dino Pedreschi, and Fosca Giannotti. Understanding the patterns of car travel. *The European Physical Journal Special Topics*, 215(1):61–73, 2013.
- [79] Santi Phithakkitnukoon, Zbigniew Smoreda, and Patrick Olivier. Socio-geography of human mobility: A study using longitudinal mobile phone data. *PloS one*, 7(6):e39253, 2012.
- [80] Balázs Cs Csáji, Arnaud Browet, Vincent A Traag, Jean-Charles Delvenne, Etienne Huens, Paul Van Dooren, Zbigniew Smoreda, and Vincent D Blondel. Exploring the mobility of mobile phone users. *Physica A: statistical mechanics and its applications*, 392(6):1459–1473, 2013.
- [81] Emanuele Pepe, Paolo Bajardi, Laetitia Gauvin, Filippo Privitera, Brennan Lake, Ciro Cattuto, and Michele Tizzoni. Covid-19 outbreak response, a dataset to assess

- mobility changes in italy following national lockdown. *Scientific data*, 7(1):1–7, 2020.
- [82] Istituto Nazionale di Statistica. Results of the Permanent Population Census, 2018.
- [83] Guglielmo Celata and Guillaume. *openpolis/geojson-italy*. 2 2021.
- [84] Carmine Minichini. Novel Coronavirus COVID-19 Italy Dataset, 2021.
- [85] What towns in Italy are on lockdown because of coronavirus? metro, 25 feb. 2020, <https://metro.co.uk/2020/02/25/towns-italy-lockdown-coronavirus-12298246/>.
- [86] Italy: Authorities to lock down Lombardy region, 11 other provinces over COVID-19 outbreak march 8 /update 13.” gardaworld, <https://www.garda.com/crisis24/news-alerts/320551/italy-authorities-to-lock-down-lombardy-region-11-other-provinces-over-covid-19-outbreak-march-8-update-13>. accessed 11 mar. 2021.
- [87] Coronavirus: Italy extends emergency measures nationwide. *BBC News*, March 2020.
- [88] Italy struck by deadliest day as virus prompts industry shutdown, bloomberg.com march 2020.
- [89] Coronavirus: Lombardy region announces stricter measures. *BBC News*, March 2020.
- [90] Coronavirus, Fase 2: dal 4 maggio sì a incontri con familiari. Il 18 riapriranno i negozi, il primo giugno bar, ristoranti, parrucchieri e centri estetici, April 2020.
- [91] Italy: Authorities tighten COVID-19 measures from November 5 /update 48. gardaworld, <https://www.garda.com/crisis24/news-alerts/396716/italy-authorities-tighten-covid-19-measures-from-november-5-update-48>.
- [92] "coronavirus: Italy extends 'red zones' as infections soar". *bbc news*, 13 nov. 2020. www.bbc.com, <https://www.bbc.com/news/world-europe-54937699>. *BBC News*, November 2020.

-
- [93] Krishna NS Behara, Ashish Bhaskar, and Edward Chung. A novel approach for the structural comparison of origin-destination matrices: Levenshtein distance. *Transportation Research Part C: Emerging Technologies*, 111:513–530, 2020.