

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**  
**Σχολή Χρηματοοικονομικής και Στατιστικής**



**Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης**  
**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ**

**ΑΝΑΚΑΛΥΨΗ ΘΑΛΑΣΣΙΩΝ**  
**ΔΙΚΤΥΩΝ ΑΠΟ ΑΙΣ ΔΕΔΟΜΕΝΑ**

**Αλκιβιάδης Κ. Μπέλος**

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

Πειραιάς  
Ιούλιος 2016

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ  
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

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

Τα μέλη της Επιτροπής ήταν:

- Πελέκης Ν. (Επιβλέπων)
- Κούτρας Μ.
- Θεοδωρίδης Γ.

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

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ  
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

**UNIVERSITY OF PIRAEUS**  
**School of Finance and Statistics**



**Department of Statistics and Insurance Science**

**POSTGRADUATE PROGRAM IN  
APPLIED STATISTICS**

**Discovery of “Sea-Networks” from AIS  
data**

By

**Alkiviadis K. Belos**

MSc Dissertation

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

Piraeus, Greece

July 2016

Στους γονείς μου,  
Κώστα και Μύρνα.

## **ΕΥΧΑΡΙΣΤΙΕΣ**

Στο σημείο αυτό, θα ήθελα να ευχαριστήσω ιδιαίτερω, τον επιβλέποντα καθηγητή μου, κύριο Νίκο Πελέκη, για την καθοδήγησή του, τις πολύτιμες συμβουλές και την πολύτιμη βοήθειά του, χωρίς τις οποίες η εκπόνηση της παρούσας εργασίας δεν θα ήταν δυνατή.

## ΠΕΡΙΛΗΨΗ

Το “πρόβλημα” που καλούμαι να αντιμετωπίσω είναι “Η ΑΝΑΚΑΛΥΨΗ ΘΑΛΑΣΣΙΩΝ ΔΙΚΤΥΩΝ ΑΠΟ ΑΙΣΔΕΔΟΜΕΝΑ”. Τα ΑΙΣ, δηλαδή τα αυτόματα συστήματα εντοπισμού ενός καραβιού, θα μας βοηθήσουν να συλλέγουμε πληροφορίες ανά πάσα στιγμή για την τοποθεσία του πλοίου καθώς και για τις παραμέτρους που μπορεί να επηρεάζουν την διαδρομή του. Θεωρώ ότι είναι ένα θέμα που αξίζει να ασχοληθεί κάποιος και να αφιερώσει τον χρόνο του. Το θέμα που θα αναπτύξουμε και θα αναλύσουμε στις επόμενες σελίδες είναι ουσιαστικά μία εφαρμογή με πολύ μεγάλη χρησιμότητα, κυρίως στους ανθρώπους που ασχολούνται με την ναυσιπλοοία στην χώρα μας. Το βασικότερο κίνητρο είναι να μπορέσουμε να αναλύσουμε σε βάθος μεγάλες βάσεις δεδομένων και να μπορέσουμε να καταλάβουμε πώς “συμπεριφέρονται” οι μεταβλητές με την μεγαλύτερη επιρροή πάνω στο πρόβλημα που προσπαθούμε να επιλύσουμε. Η προοπτική που μπορεί να έχει η εφαρμογή αυτή καθώς και η γνώση της συμπεριφοράς των μεγάλων βάσεων δεδομένων είναι ένα τεράστιο κίνητρο για κάποιον να ασχοληθεί με αυτή την εργασία και την εκπόνησή της.

Στο χώρο έχουν γίνει διάφορες εργασίες με στόχο τη μελέτη και την εύρεση αλγορίθμων για την επίλυση των μεγάλων βάσεων δεδομένων σε πολύ μικρούς χρόνους για την καλύτερη και πιο γρήγορη χρησιμοποίησή τους. Οι εργασίες αυτές προσπαθούν να δώσουν λύσεις στα διάφορα προβλήματα που δημιουργούν οι μεγάλες βάσεις δεδομένων. Ένα από τα κυριότερα προβλήματα είναι αυτό των επαναλαμβανόμενων στοιχειοσυνόλων που παρατηρείται στις βάσεις δεδομένων. Οι εργασίες που μελετήθηκαν, τις οποίες θα αναφέρω παρακάτω, σχετίζονται κυρίως με την εύρεση θεωρητικών ή πρακτικών μοντέλων που κύριο στόχο έχουν την σμίκρυνση των τεραστίων βάσεων δεδομένων.

## ABSTRACT

The problem that I am called to solve is the "Discovery of "New Sea-Networks" from AIS Data". The AIS, the Automatic Identification System, of a ship will help us collect information about the location and the parameters that might affect a ship's route at any time. I believe that it is a subject worth to deal with and devote time. The subject that we will develop and analyze through this paper, is practically an application with high utility, especially for people involved with navigation in our country. The main motivation is to be able to analyze in depth large databases and to understand how the variables with greatest influence "act" on the problem that we are called to solve. The prospect that this application might have, as well as the knowledge of the action of large databases, is a huge incentive for someone to involve with this paper and its preparation.

In this field various papers have been made targeting the study and finding algorithms for the solution of dealing with large databases in short time for better and faster use. These papers try to provide solutions in the various problems that large databases create. One of the most basic problems is the one with the repeated item sets that is observed in databases. The papers that have been studied and will be mentioned later are related mostly with the location of theoretical or practical models, which primary aim to shrink the huge databases.

Έχω διαβάσει και κατανοήσει τους κανόνες του ΠΜΣ που περιέχονται στον Οδηγό Συγγραφής ΔΕ και ιδιαίτερα όσα συνιστούν λογοκλοπή. Δηλώνω ότι η παρούσα διπλωματική εργασία αποτελεί προϊόν αποκλειστικά δικής μου προσπάθειας, υπό την καθοδήγηση του επιβλέποντος καθηγητή, ενώ για όλες τις πηγές που χρησιμοποιήθηκαν περιλαμβάνονται οι α- ντίστοιχες αναφορές.



## Table of Contents

<b>ΕΥΧΑΡΙΣΤΙΕΣ</b> .....	5
<b>ΠΕΡΙΛΗΨΗ</b> .....	6
<b>ABSTRACT</b> .....	7
<b>ΠΕΡΙΕΧΟΜΕΝΑ ΠΙΝΑΚΩΝ</b> .....	11
<b>ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ</b> .....	13
<b>ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ</b> .....	16
<b>2.1 ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΦΟΡΑ</b> .....	16
<b>2.1.1 SPATIO-TEMPORAL TRAJECTORY ANALYSIS OF MOBILE OBJECTS FOLLOWING THE SAME ITINERARY</b> .....	16
<b>2.1.2 TRAFFIC ROUTE EXTRACTION AND ANOMALY DETECTION FROM AIS DATA</b> .....	17
<b>2.1.2.1 DBSCAN</b> .....	17
<b>Ορισμός 1: Η γειτονιά-Eps ενός σημείου</b> .....	18
<b>Ορισμός 2: Directly density-reachable</b> .....	18
<b>Ορισμός 3: density-reachable</b> .....	18
<b>Ορισμός 4: density-connected</b> .....	19
<b>Ορισμός 5: Συστάδα</b> .....	19
<b>Ορισμός 6: Θόρυβος</b> .....	20
<b>2.1.3 SEMI-AUTOMATIC SEA LANE EXTRACTION COMBINING PARTICLE FILTERING (PF) AND GEOGRAPHIC INFORMATION SYSTEM (GIS)</b> .....	20
<b>2.1.4.1 SIR</b> .....	21
<b>Βήμα 1</b> .....	21
<b>Βήμα 2</b> .....	21
<b>Βήμα 3</b> .....	21
<b>Βήμα 4</b> .....	21
<b>Βήμα 5</b> .....	22
<b>Βήμα 6</b> .....	22
<b>Βήμα 7</b> .....	22
<b>2.1.4 FROM GPS TRACES TO A ROUTABLE ROAD MAP</b> .....	22
<b>2.1.5 MINING GPS TRACES FOR MAP REFINEMENT</b> .....	23
<b>2.1.6 ON VEHICLE TRACKING DATA-BASED ROAD NETWORK GENERATION</b> .....	24
<b>2.1.7 SEGMENTATION-BASED ROAD NETWORK CONSTRUCTION</b> .....	25
<b>2.2 ΑΝΑΛΥΣΗ ΑΛΓΟΡΙΘΜΩΝ</b> .....	26
<b>2.2.1 ΠΡΩΤΟΣ ΑΛΓΟΡΙΘΜΟΣ</b> .....	26
<b>Στάδιο 1</b> :.....	27
<b>Στάδιο 2</b> :.....	27

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ  
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

Στάδιο 3:.....	27
<b>2.2.2 ΔΕΥΤΕΡΟΣ ΑΛΓΟΡΙΘΜΟΣ .....</b>	<b>28</b>
1) Εκτίμηση πυκνότητας.....	28
2) Αρχική γενιά χάρτη .....	28
3) Ταιριαστά ίχνη στον χάρτη:.....	28
4) Βελτίωση της τοπολογίας: .....	28
5) Βελτίωση της γεωμετρίας:.....	28
<b>2.2.3 ΤΡΙΤΟΣ ΑΛΓΟΡΙΘΜΟΣ .....</b>	<b>29</b>
Βήμα 1: .....	29
Βήμα 2: .....	29
Τύπος 1: .....	29
Τύπος 2: .....	29
<b>2.2.4 ΤΕΤΑΡΤΟΣ ΑΛΓΟΡΙΘΜΟΣ .....</b>	<b>30</b>
<b>2.2.5 ΠΕΜΠΤΟΣ ΑΛΓΟΡΙΘΜΟΣ .....</b>	<b>30</b>
<b>2.2.6 ΕΚΤΟΣ ΑΛΓΟΡΙΘΜΟΣ .....</b>	<b>31</b>
<b>2.2.7 ΕΒΔΟΜΟΣ ΑΛΓΟΡΙΘΜΟΣ .....</b>	<b>32</b>
<b>2.2.8 ΟΓΔΟΣ ΑΛΓΟΡΙΘΜΟΣ .....</b>	<b>32</b>
<b>2.3 ΚΑΤΗΓΟΡΙΕΣ ΑΛΓΟΡΙΘΜΩΝ.....</b>	<b>33</b>
2.3.1. POINT CLUSTERING.....	33
2.3.2 INCREMENTAL TRACK INSERTION .....	34
2.3.3 INTERSECTION LINKING .....	34
<b>ΚΕΦΑΛΑΙΟ 3 .....</b>	<b>36</b>
<b>3.ΜΕΤΡΑ ΠΟΙΟΤΗΤΑΣ ΓΙΑ ΣΥΓΚΡΙΣΗ ΧΑΡΤΩΝ .....</b>	<b>36</b>
<b>3.1 ΜΕΤΡΑ ΠΟΙΟΤΗΤΑΣ .....</b>	<b>36</b>
3.1.1 DIRECTED HAUSDORFF DISTANCE.....	36
3.1.2 PATH BASED DISTANCE.....	37
3.1.3 SHORTEST PATH BASED DISTANCE .....	37
3.1.4 GRAPH-SAMPLING BASED DISTANCE.....	38
<b>3.2 ΣΥΓΚΡΙΣΗ ΑΠΟΣΤΑΣΗΣ ΜΕΤΡΩΝ .....</b>	<b>39</b>
<b>ΚΕΦΑΛΑΙΟ ΤΕΤΑΡΤΟ.....</b>	<b>40</b>
<b>4.1 ΑΞΙΟΛΟΓΗΣΗ ΤΩΝ ΑΛΓΟΡΙΘΜΩΝ.....</b>	<b>40</b>
<b>4.2 ΣΥΜΠΕΡΑΣΜΑ .....</b>	<b>42</b>
<b>ΚΕΦΑΛΑΙΟ ΠΕΜΠΤΟ .....</b>	<b>44</b>
<b>5. ΠΕΙΡΑΜΑΤΙΚΗ ΜΕΛΕΤΗ .....</b>	<b>44</b>
<b>5.1 ΠΡΟΣΕΓΓΙΣΗ.....</b>	<b>44</b>
<b>5.2 DATASET.....</b>	<b>45</b>
<b>5.3 ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΛΓΟΡΙΘΜΩΝ .....</b>	<b>48</b>
<b>5.4 ΣΥΜΠΕΡΑΣΜΑΤΑ .....</b>	<b>52</b>
<b>ΠΕΡΙΟΡΙΣΜΟΙ.....</b>	<b>53</b>
<b>ΠΑΡΑΡΤΗΜΑ .....</b>	<b>54</b>
<b>1. ΑΛΓΟΡΙΘΜΟΣ ΜΕΤΑΤΡΟΠΗΣ ΔΕΔΟΜΕΝΩΝ 1 .....</b>	<b>54</b>
<b>2. ΑΛΓΟΡΙΘΜΟΣ ΜΕΤΑΤΡΟΠΗΣ ΔΕΔΟΜΕΝΩΝ 2 .....</b>	<b>57</b>
<b>ΑΛΓΟΡΙΘΜΟΣ ΜΕΤΑΤΡΟΠΗΣ ΔΕΔΟΜΕΝΩΝ-3 .....</b>	<b>59</b>

<b>ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ.....</b>	<b>63</b>
<b>ΞΕΝΕΣ ΑΝΑΦΟΡΕΣ.....</b>	<b>63</b>
<b>ΑΝΑΦΟΡΕΣ ΣΤΟ ΔΙΑΔΙΚΤΥΟ .....</b>	<b>65</b>

## **ΠΕΡΙΕΧΟΜΕΝΑ ΠΙΝΑΚΩΝ**

TABLE 1: ΚΑΤΗΓΟΡΙΕΣ ΑΛΓΟΡΙΘΜΩΝ .....	35
TABLE 2: AIS ΔΕΔΟΜΕΝΑ .....	44
TABLE 3: AIS ΔΕΔΟΜΕΝΑ 2 .....	45
TABLE 4: GPS ΔΕΔΟΜΕΝΑ.....	47
TABLE 5 ΑΗΜΕD-1 .....	48
TABLE 6 ΑΗΜΕD-2 .....	48
TABLE 7 ΑΗΜΕD-3 .....	49
TABLE 8 CAO-1 .....	50
TABLE 9 CAO-2 .....	50
TABLE 10 CAO-3.....	51
TABLE 11 EDELKAMP-1 .....	51
TABLE 12 EDELKAMP-3 .....	52

## ΕΙΣΑΓΩΓΗ

Η εφαρμογή αυτή θα βοηθήσει να κατανοήσουμε τι είναι τα θαλάσσια δίκτυα, σε τι χρησιμεύουν και σε τι μπορούμε να βρούμε λύσεις ώστε να δημιουργήσουμε νέα θαλάσσια μονοπάτια. Ουσιαστικά, το θέμα μας σχετίζεται με την ανάλυση μεγάλων βάσεων δεδομένων. Θα προσπαθήσουμε να προσεγγίσουμε τις μεταβλητές που επηρεάζουν τις αλλαγές των κατευθύνσεων και κατ'επέκταση την αλλαγή της διαδρομής των πλοίων. Μέσω ενός αλγορίθμου θα προσπαθήσουμε να βρούμε ποια ή ποιες από τις μεταβλητές έχουν την μεγαλύτερη και ουσιαστικότερη συμμετοχή στην διαδικασία ανακάλυψης νέων μονοπατιών. Η βάση δεδομένων που θα αναπτυχθεί θα προέρχεται από ηλεκτρονικά συστήματα αυτόματης καταγραφής δεδομένων και θα περιέχει αναφορές θέσεων των πλοίων.

Η επιτυχία του αποτελέσματος θα κριθεί από το αν μέσω του αλγορίθμου που θα χρησιμοποιηθεί, θα ανακαλυφθούν θαλάσσια δίκτυα.

Συμπερασματικά, σκοπός της εργασίας είναι η ανάλυση μεγάλης βάσης δεδομένων. Η βάση αυτή περιέχει μεγάλο αριθμό μεταβλητών και ο στόχος είναι να μπορέσουμε να καταλάβουμε πώς συμπεριφέρονται οι μεταβλητές αυτές και πώς επηρεάζουν τα αποτελέσματα στην ανάζητηση των θαλάσσιων διαδρομών που θα επιχειρήσουμε να ανακαλύψουμε.

## ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ

Η εργασία αυτή θα ασχοληθεί με την ανακάλυψη των θαλάσσιων δικτύων με την χρήση συστημάτων αυτόματου εντοπισμού(AIS).

Τα θαλάσσια δίκτυα είναι μία προσέγγιση με την οποία θα προσπαθήσουμε να κατανοήσουμε "πάνω" σε τι μονοπάτια κινούνται τα πλοία. Όπως και στη στερία υπάρχουν τα οδικά δίκτυα στα οποία κινούνται όλα τα κινητά αντικείμενα, έτσι θα προσπαθήσουμε να δημιουργήσουμε τα αντίστοιχα θαλάσσια μονοπάτια. Η δυσκολία στο εν λόγω εγχείρημα έγκειται στη μη σταθέρη κίνηση και εκπομπή σημάτων των πλοίων σε σχέση με τα αυτοκίνητα. Αυτό δυσκολεύει την ανάλυση και κατ'επέκταση τη χρησιμοποίηση των AIS δεδομένων σε αλγόριθμους που σχετίζονται με τα οδικά δίκτυα.

Τα AISδεδομένα χρησιμοποιούνται για την αναγνώριση και τον εντοπισμό πλοίων. Τα πλοία χρησιμοποιούν συνήθως για την αποστολή των δεδομένων τους ραντάρ ή συχνότητες VHF, οι οποίες μπορεί να περιέχουν id-θέσης, πορεία διαδρομής,συνθήκες πλευσης καθώς και την ταχύτητα με την οποία κινείται το πλοίο.Άλλες πληροφορίες που μπορούν να αντληθούν είναι το όνομα του πλοίου ή το αναγνωριστικό idπου έχει το κάθε πλοίο, αν αυτό έχει εγκατασταθεί στο αντίστοιχο πρόγραμμα εντοπισμού πλοίων. Τα σήματα που εκπέμπει κάθε πλοίο μπορούν να εμφανίζονται στα πλοία που κινούνται γύρω από αυτό μέσω μίας οθόνης που λειτουργεί όπως ακριβώς και το ραντάρ. Τα AISδεδομένα περιλαμβάνουν πολλές υποκατηγορίες που καθορίζουν τον εντοπισμό των πλοίων. Κάθε πλοίο θα πρέπει να περιέχει κάποια προϊόντα τα οποία πληρούν τις προδιαγραφές των συστημάτων AIS. Για τα συστήματα αυτά υπάρχουν δύο κύριοι τύποι προϊόντων που πρέπει να φέρει πάνω του ένα πλοίο για να μπορεί να διασφαλιστεί η ομαλή λειτουργία του συστήματος των AIS δεδομένων. Αυτά αναφέρονται κυρίως σε εμπορικά πλοία.

Κάθε πλοίο θα πρέπει να διαθέτει έναν πομποδέκτη με τον οποίο θα λαμβάνει καθώς και θα μεταδίδει τις πληροφορίες της θέσης του. Αυτό γίνεται για να μπορεί να παραμένει ενημερωμένο το σύστημα των AIS δεδομένων. Η εκπομπή αυτών των σημάτων γίνεται κάθε λίγα δευτερόλεπτα.

Ο δεύτερος τύπος απαιτεί την τοποθέτηση gpsστα πλοία καθώς και κάποιων δεικτών LED. Και εδώ χρησιμοποιείται ωστόσο ο πομποδέκτης. Σε αυτή την κατηγορία, η εκπομπή δεδομένων γίνεται ανά 30 δευτερόλεπτα, αλλά χωρίς αυτό να είναι δεσμευτικό. Αν αλλάξει για παράδειγμα η ταχύτητα του πλοίου, αλλάζει και ο χρόνος μετάδοσης της εκπομπής των AISδεδομένων.

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ  
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

Η σημαντική διαφορά των δεδομένων αυτών σε σχέση με τα gpsείναι ότι οι πομποδέκτες που δέχονται τα δεδομένα δε βασίζονται στις προδιαγραφές που έχουν θέσει τα AISδεδομένα και ο λόγος είναι ότι αυτά δεν μεταδίδουν τα σήματα. Έτσι, κάθε πλοίο θα πρέπει να έχει ένα σύστημα εντοπισμού που θα πληροί όλες τις προδιαγραφές των μεταδόσεων των AISδεδομένων.

Μία σημαντική λεπτομέρεια που κάνει ακόμη πιο δύσκολη την εύρεση θαλάσσιων δικτύων έχει να κάνει με το κόστος. Πολλά πλοία που χρησιμοποιούν τη μετάδοση των AISδεδομένων χρησιμοποιούν single-channels, με αποτέλεσμα να χάνονται πληροφορίες. Μόνο όσα πλοία χρησιμοποιούν dual-channels, θα λαμβάνουν όλες τις πληροφορίες που μεταδίδονται από τα AISδεδομένα.

Η ανακάλυψη, λοιπόν, των θαλάσσιων δικτύων θα βοηθήσει στην πλήρη κατανόηση των κινήσεων των πλοίων. Θα μπορέσουμε να αντιληφθούμε καλύτερα γιατί ένα πλοίο αλλάζει ταχύτητα ή ακόμη και πορεία (χωρίς να έχει αλλάξει τον τελικό προορισμό του), ποιοι παράγοντες το επηρεάζουν καθώς και τον χρόνο που κάνει για να φτάσουν στον τελικό προορισμό του.

Μέσω λοιπόν κάποιων αλγορίθμων κατασκευής χαρτών, θα προσπαθήσουμε να αναλύσουμε τα δεδομένα μας. Τα δεδομένα θα πρέπει να τα φέρουμε στη μορφή που χρησιμοποιούν οι αλγόριθμοι αυτοί και να εξάγουμε κάποια συμπεράσματα. Το ιδανικότερο σενάριο θα είναι να κατασκευάσουμε έναν θαλάσσιο χάρτη.

Για να μπορέσω λοιπόν να χρησιμοποιήσω τους συγκεκριμένους αλγορίθμους (θα τους αναλύσουμε εκτενώς στο κεφάλαιο 3) θα πρέπει να μετατρέψω τα δεδομένα στη μορφή που χρησιμοποιεί ο κάθε αλγόριθμος. Θα σχεδιάσω λοιπόν έναν αλγόριθμο μετατροπής δεδομένων στις αντίστοιχες μορφές. Με την εξαγωγή των νέων δεδομένων, θα έχουμε μία πρώτη εικόνα για τις αλλαγές που θα πρέπει να κάνουμε στους αλγορίθμους για να μπορέσουν να τους τρέξουμε. Θα τοποθετήσω τα νέα δεδομένα στους αλγορίθμους και θα προσπαθήσουμε να εξάγουμε ένα γράφημα με τη συμπεριφορά των δεδομένων αυτών. Η δυσκολία της όλης διαδικασίας έχει να κάνει με το ότι οι αλγόριθμοι που θα χρησιμοποιήσουμε είναι σχεδιασμένοι για τη δημιουργία οδικών δικτύων.

Προτού φτάσουμε στο τελικό στάδιο της εργασίας και στην εξαγωγή των αποτελεσμάτων, θα πρέπει να αναλύσουμε τι έχει γίνει στον χώρο της κατασκευής χαρτών μέσω σχετικών εργασιών. Να κατανοήσουμε τα προβλήματα και τις δυσκολίες αυτού του εγχειρήματος και να λάβουμε τη γνώση για τις μεθόδους που ακολουθούν.

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ  
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

Στο επόμενο κεφάλαιο, λοιπόν, θα γίνει μία βιβλιογραφική ανασκόπηση στο τι έχει συμβεί μέχρι τώρα στον συγκεκριμένο τομέα. Με τον τρόπο αυτό, θα μπορέσουμε να κατανοήσουμε λεπτομερώς το πρόβλημα που έχουμε να αντιμετωπίσουμε.

Μέτα θα αναλύσουμε τους σημαντικότερους αλγορίθμους που υπάρχουν στον χώρο για την κατασκευή χαρτών. Θα δούμε πώς λειτουργούν και ποια κριτήρια πρέπει να ικανοποιούν τα δεδομένα μας για να τρέξουν.

Θα συνεχίσουμε με την κατηγοριοποίηση των αλγορίθμων στις τρεις κύριες κατηγορίες που υπάρχουν και θα κλείσουμε με την ποιοτική τους σύγκριση.

Στο τελευταίο κεφάλαιο αυτής της εργασίας, θα ασχοληθούμε με την πειραματική μελέτη και την εξαγωγή των αποτελεσμάτων και των συμπερασμάτων του πειράματός μας.

## ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ

### 2.1 ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΦΟΡΑ

Οι παρακάτω εργασίες είναι η βάση πάνω στην οποία θα στηριχθούμε για την ολοκλήρωση της συγκεκριμένης διπλωματικής εργασίας. Ας δούμε τι έχει μελετηθεί στο χώρο μέσω των εργασιών που θα παραθέσουμε:

#### 2.1.1 SPATIO-TEMPORAL TRAJECTORY ANALYSIS OF MOBILE OBJECTS FOLLOWING THE SAME ITINERARY

Σύμφωνα με τους LaurentETIENNE, ThomasDEVOGELEandAlainBOUJU(2010), τα περισσότερα κινητά αντικείμενα είναι πλέον εξοπλισμένα με αισθητήρες που δίνουν τη δυνατότητα σε όποιον το επιθυμεί να παρακολουθήσει τις κινήσεις τους σε πραγματικό χρόνο. Αυτά τα δεδομένα μπορούν να αποθηκευτούν σε χωροχρονικές βάσεις δεδομένων. Η εργασία αυτή εκτελεί μια εξόρυξη δεδομένων σε σχέση με τις θέσεις των αντικειμένων, προκειμένου να συμπεράνει τη συμπεριφορά τους. Για να διευκολυνθεί ο εντοπισμός των ακραίων τιμών, θα χρησιμοποιηθούν νέες μέθοδοι. Η πρώτη μέθοδος θα είναι ένα γράφημα ζώνης για να καθοριστούν οι διαδρομές. Ακολουθώντας την ίδια διαδρομή που εξάγεται από τη χωροχρονική βάση δεδομένων, ομαδοποιούνται. Η στατιστική ανάλυση που γίνεται οδηγεί σε χωροχρονικά πρότυπα, όπως είναι η κύρια διαδρομή. Ωστόσο, πιθανόν είναι να διακρίνουμε ασυνήθιστες συμπεριφορές, όπως είναι η αλλαγή πλεύσης ή προσανατολισμού, αφού αναφέρεται σε κινήσεις πλοίων σε μια ανοιχτή θαλάσσια περιοχή. Η μέθοδος που χρησιμοποιήθηκε σε αυτή την εργασία έχει τη δυνατότητα να επεκταθεί και σε άλλα είδη κινητών αντικειμένων, αρκεί αυτά να κινούνται σε ανοικτές περιοχές.

Στη συγκεκριμένη εργασία παρουσιάζεται η στατιστική ανάλυση των τροχιών που αναλύεται σε τρεις αλγόριθμους. Πρώτα βλέπουμε τον υπολογισμό της κύριας διαδρομής, ακολουθεί ο υπολογισμός του χωρικού καναλιού και τέλος ο υπολογισμός των χωροχρονικών ζωνών.

Το κύριο πρόβλημα που μελετήθηκε εδώ ήταν οι μεμονομένες ανιχνεύσεις. Η ιδέα της διαδρομής και της τροχιάς είναι να ακολουθούν ένα τόξο δρομολογίου. Έτσι, αποκτούνται οι θέσεις των αντικειμένων και τοποθετούνται σε μία βάση δεδομένων. Στη συνέχεια, ομαδοποιούνται τα αντικείμενα που ακολουθούν τα ίδια τόξα διαδρομών. Κατ'αυτόν τον τρόπο ορίζεται η κύρια διαδρομή. Τα μεταγενέστερα δεδομένα αποθηκεύονται και αυτά με τη σειρά τους στη βάση δεδομένων. Κάθε νέα θέση μπορεί χωρικά και χρονικά να προαχθεί. Αυτές οι διαδικασίες έχουν δοκιμαστεί



σε ένα σημαντικό σύνολο δεδομένων που εφαρμόζονται στο πλαίσιο των θαλάσσιων μεταφορών σε διαφορετικές περιοχές.

### 2.1.2 TRAFFIC ROUTE EXTRACTION AND ANOMALY DETECTION FROM AIS DATA

Στην εργασία τους οι Giuliana Pallotta, Michele Vespe, and Karna Bryan (2013) ανέλυσαν τη χρησιμότητα του Αυτόματου Συστήματος Αναγνώρισης (AIS) για την αποφυγή συγκρούσεων και την on-line παρακολούθηση σκαφών. Εκτός όμως από αυτή την ιδιότητα, τα AIS δεδομένα αποθηκεύονται με στόχο τη δημιουργία μίας βάσης δεδομένων με καταγραφές πλοίων από πολύ παλιά. Παρουσίασαν μια προσέγγιση για να μάθουμε τα μοτίβα της θαλάσσιας κυκλοφορίας. Αυτό επιτυγχάνεται με την επεξεργασία των δεδομένων AIS, ώστε να αφαιρούνται οι περιττές πληροφορίες. Με αυτή τη διαδικασία λαμβάνεται μια βελτιωμένη εικόνα του θαλάσσιου τομέα, ώστε να μπορεί να υποστηριχθεί η γνώση των θαλάσσιων μοτίβων. Επιπλέον, η κατανόηση των ιστορικών δεδομένων δίνει τη δυνατότητα για την ταξινόμηση και την πρόβλεψη των μελλοντικών θέσεων των πλοίων.

Η εργασία αυτή χρησιμοποίησε ένα εργαλείο το οποίο ονομάζεται TREAD και αποδίδει αυτόματη γνώση της θαλάσσιας κυκλοφορίας (λιμάνια, υπεράκτιες πλατφόρμες, σημεία εισόδου-εξόδου και δρομολόγια) μέσω των δεδομένων AIS, χωρίς να έχει υπάρξει προηγούμενη ενημέρωση στην περιοχή που το χρησιμοποίησαν. Οι ανακαλυφθείσες διαδρομές έχουν τρία βασικά χαρακτηριστικά γνωρίσματα: 1) χωρικές, 2) χρονικές και 3) χαρακτηριστικές πληροφορίες, επιτρέποντας την πρόβλεψη των μελλοντικών θέσεων και των πιθανών προορισμών των πλοίων. Η παραγόμενη γνώση από αυτή τη μη επιβλεπόμενη προσέγγιση μπορεί να συνδυαστεί με άλλες πηγές πληροφοριών σε πραγματικό χρόνο για την ενίσχυση της επίγνωσης της κατάστασης σε μια συγκεκριμένη περιοχή ενδιαφέροντος.

#### 2.1.2.1 DBSCAN

Ο αλγόριθμος που χρησιμοποιήθηκε είναι ο DBSCAN. Αυτός ο αλγόριθμος εντοπίζει συστάδες σε μεγάλα σύνολα χωρικών δεδομένων κοιτάζοντας την τοπική πυκνότητα των στοιχείων της βάσης δεδομένων, χρησιμοποιώντας μόνο μία παράμετρο εισόδου. Ο DBSCAN έχει την ικανότητα να καθορίσει ο ίδιος ποιες πληροφορίες θα πρέπει να ταξινομούνται ως θόρυβος ή outliers. Η διαδικασία είναι γρήγορη και κλιμακώνεται πολύ καλά σε σχέση με το μέγεθος της βάσης δεδομένων - σχεδόν γραμμικά.

Ο DBSCAN μπορεί να βρει συστάδες μη συγκεκριμένου σχήματος. Ωστόσο, οι συστάδες, που βρίσκονται σχετικά κοντά η μία στην άλλη, τείνουν να ανήκουν στην ίδια κατηγορία.

Η διαδικασία υπολογισμού του βασίζεται σε έξι κανόνες σύμφωνα με την εργασία των HenrikBäcklund, AndersHedblom , NiklasNeijman [DBSCANADensity-BasedSpatialClusteringofApplicationwithNoise ].

### Ορισμός 1: Η γειτονιά-Eps ενός σημείου

$$NEps(p) = \{q \in D | \text{dist}(p,q) < Eps\}$$

Για ένα σημείο που ανήκει σε μια συστάδα θα πρέπει να υπάρχει τουλάχιστον ένα άλλο σημείο που να βρίσκεται πιο κοντά σε αυτό από την απόσταση EPS.

### Ορισμός 2: Directlydensity-reachable

Υπάρχουν δύο είδη σημείων που ανήκουν σε μια συστάδα: τα συνοριακά σημεία και τα σημεία πυρήνες.

Η γειτονιά-Eps του συνοριακού σημείου τείνει να έχει σημαντικά λιγότερους πόντους από τη γειτονιά Eps-ενός σημείου πυρήνα. Τα συνοριακά σημεία θα εξακολουθούν να είναι ένα μέρος της συστάδας και προκειμένου να συμπεριληφθούν αυτά τα σημεία, θα πρέπει να ανήκουν στην Eps-γειτονιά ενός σημείου πυρήνα.

P: συνοριακό σημείο

q: σημείο πυρήνας

- $P \in NEps(q)$

Σε σχέση με το σημείο q για να μπορεί να χαρακτηριστεί ένα σημείο πυρήνας, πρέπει να έχει έναν ελάχιστο αριθμό σημείων εντός της Eps-γειτονιάς του.

- $|NEps(q)| \geq \text{MinPts}$  (συνθήκη σημείου πυρήνα)

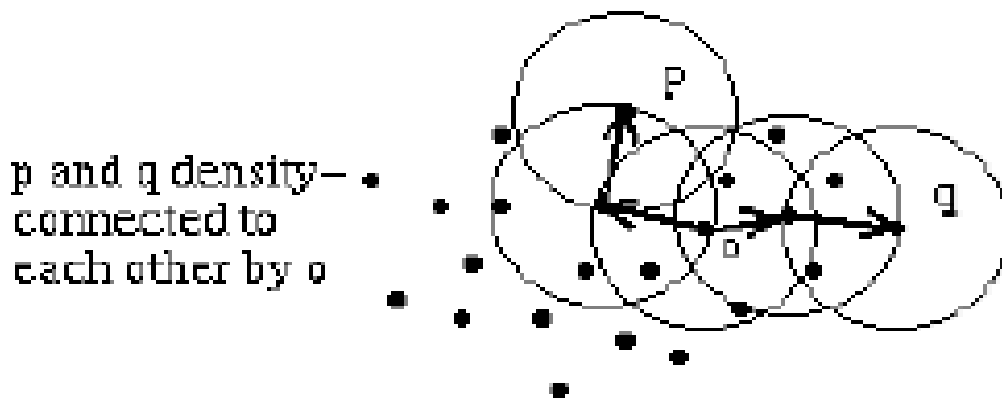
### Ορισμός 3: density-reachable

Ένα σημείο p έχει πρόσβαση πυκνότητας από ένα σημείο q σε σχέση με το EPS και MinPts αν υπάρχει μια αλυσίδα των σημείων  $p_1, \dots, p_n$ , τέτοια ώστε  $p_1 = q$ ,  $p_n = p$  και το  $p_{i+1}$  να έχει απευθείας πρόσβαση πυκνότητας από το  $p_i$ .

#### Ορισμός 4: density-connected

Υπάρχουν περιπτώσεις που δύο συνοριακά σημεία ανήκουν στην ίδια συστάδα, αλλά δε μοιράζονται ένα συγκεκριμένο σημείο πυρήνα. Σε αυτές τις περιπτώσεις, τα σημεία δε θα είναι density-reachable το ένα από το άλλο. Πρέπει, ωστόσο, να υπάρχει ένα σημείο πυρήνας  $q$  από το οποίο είναι και τα δύο density-reachable. Ουσιαστικά δηλαδή, ένα σημείο  $p$  είναι density-connected με ένα σημείο  $q$  σε σχέση με το EPS και MinPts, αν υπάρχει ένα σημείο τέτοιο ώστε και τα δύο,  $p$  και  $q$  να είναι density-reachable από το  $\theta$  σε σχέση με το EPS και MinPts.

Table 1. Density connectivity



#### Ορισμός 5: Συστάδα

Εάν το σημείο  $p$  είναι μέρος μιας συστάδας  $C$  και το σημείο  $q$  είναι η density-reachable από το σημείο  $p$  σε σχέση με μια δεδομένη απόσταση και ένα ελάχιστο αριθμό σημείων μέσα σε αυτή την απόσταση, τότε το  $q$  είναι επίσης ένα μέρος της συστάδας  $C$ .

- $\forall p, q$ : αν το  $p \in C$  και το  $q$  είναι density-reachable από το  $p$  σε σχέση με την  $E_p$  και το  $MinPts$ , τότε  $q \in C$ .
- $\forall p, q \in C$ : το  $p$  είναι density-connected με το  $q$  σε σχέση με την  $E_p$  και το  $MinPts$ .

Δηλαδή, όταν λέμε πως δύο σημεία ανήκουν στην ίδια συστάδα  $C$ , σημαίνει ότι το  $p$  είναι density-connected με το  $q$  σε σχέση με τη δεδομένη απόσταση και τον αριθμό των σημείων εντός της συγκεκριμένης απόστασης.

### Ορισμός 6: Θόρυβος

Ο θόρυβος είναι το σύνολο των σημείων, στη βάση δεδομένων, που δεν ανήκουν σε κάποια από τις συστάδες.

Henrik Bäcklund, Anders Hedblom , Niklas Neijman [DBSCAN A Density-Based Spatial Clustering of Application with Noise ] (1996)

### 2.1.3 SEMI-AUTOMATIC SEA LANE EXTRACTION COMBINING PARTICLE FILTERING (PF) AND GEOGRAPHIC INFORMATION SYSTEM (GIS)

Οι KaterinaTZAVELLA και ο MartinULMKE(2014) ασχολήθηκαν με την εξαγωγή κυκλοφοριακών μοτίβων, τα οποία περιέχουν θόρυβο και ελλιπή δεδομένα που λαμβάνονται μέσω των αισθητήρων. Προσπάθησαν να εισάγουν μια προσέγγιση με την οποία θα είχαν τη δυνατότητα να συνδυάσουν κάποια φίλτρα παρακολούθησης των κινητών αντικειμένων με τις τεχνικές GIS. Τα φίλτρα αυτά είναι τυποποιημένα εργαλεία που έχουν την ικανότητα να εκτιμούν τις θέσεις και τις ταχύτητες των μεμονωμένων αντικειμένων, οι τεχνικές GIS χρησιμοποιούνται για να εξάγουν γεωγραφικά μοτίβα από αυτές τις εκτιμήσεις. Οι τεχνικές GIS επιτρέπουν την ενσωμάτωση ιστορικών δεδομένων γνώσης, όπως ακτογραμμές και βαθυμετρία, για να βελτιωθεί η ποιότητα των εξαγόμενων μοτίβων. Η προσέγγιση που έχει εφαρμοστεί στη θαλάσσια κυκλοφορία, χρησιμοποιεί το Σύστημα Αυτόματης Αναγνώρισης (AIS), το οποίο μπορεί να υπόκειται σε σφάλματα του GPS, πλαστογράφιση, χρονικές καθυστερήσεις και περιόδους χαμένων ανιχνεύσεων. Τα αποτελέσματα είναι συνεχείς διαδρομές πλοίων, ακόμα και όταν ο ρυθμός ενημέρωσης των μεταδόσεων είναι σημαντικά πιο αργός.

Ενώ η υπηρεσία των δεδομένων AIS παρέχει ενημερώσεις των θέσεων του πλοίου στις περιοχές που καλύπτονται από το δίκτυο των πομπών των δεδομένων AIS σχεδόν σε πραγματικό χρόνο, η υπηρεσία SAIS (Satellite AIS) παρέχει ενημερώσεις θέσεων με μικρότερη συχνότητα, κατά διαστήματα που μπορεί να κυμαίνονται από λίγα λεπτά έως αρκετές ώρες. Οι ενημερώσεις SAIS σχετίζονται περισσότερο με πλοία που πλέουν σε ωκεανούς. Σε αυτές τις περιπτώσεις είναι που ο συνδυασμός των φίλτρων παρακολούθησης των αναφερόμενων δεδομένων AIS αποκλίνουν από τους στόχους και η GIS μέθοδος φαίνεται να είναι ένα ισχυρό εργαλείο για την εξαγωγή των θαλάσσιων διαδρόμων. Η μείωση της ποσότητας των σωματιδίων που χρησιμοποιούνται από την προτεινόμενη μεθοδολογία οδηγεί σε ελαχιστοποίηση του κόστους και στην καλύτερη ακρίβεια των αποτελεσμάτων (που εξάγουν θαλάσσιους

διαδρόμους κοντά σε αληθινά σημεία του εδάφους με βάση τις τροχιές που έχουν επιλέξει τα σκάφη). Επιτεύχθηκε, επίσης, μια ολιστική και επιτυχής εκμετάλλευση παρακολούθησης των αποτελεσμάτων και των δεδομένων πληροφοριών, δεδομένου ότι ο αλγόριθμος που περιγράφεται στην εργασία αυτή, είναι ανοιχτός σε πολλές επεκτάσεις (π.χ. αυτοματισμού, πολλαπλή παρακολούθηση στόχου κλπ). Ένα άλλο πλεονέκτημα αυτής της μεθοδολογίας είναι ο υψηλός βαθμός αυτοματοποίησης. Στις παραπάνω εργασίες αναπτύχθηκαν και αναφέρθηκαν κάποια μοντέλα αλγορίθμων που βοηθούν στην εξόρυξη γνώσης μεγάλων βάσεων δεδομένων.

Ο κύριος αλγόριθμος που χρησιμοποιήθηκε είναι ο SIR. Παρακάτω θα δούμε πώς λειτουργεί ο συγκεκριμένος αλγόριθμος. (KCLi – 2000)

#### 2.1.4.1 SIR

Ας υποθέσουμε ότι  $(y_1, x_1), \dots, (y_n, x_n)$  είναι το αρχικό σύνολο δεδομένων με  $(p+1)$  μεταβλητές και  $n$  περιπτώσεις. Ο αλγόριθμος του SIR αποτελείται από τα ακόλουθα βήματα.

##### Βήμα 1

Ταξινομεί τη βάση δεδομένων βάση του  $Y$ .

##### Βήμα 2

Χωρίζει τα δεδομένα σε  $H$  τμήματα όσο πιο όμοια μπορεί. Τα τμήματα δηλώνονται ως:

$$(Y_{(i)}, X_{(i)}), \text{ με } i = 1, \dots, H.$$

##### Βήμα 3

Για κάθε τμήμα, υπολογίζει τη μέση τιμή του δείγματος  $X$ . Αυτό υπολογίζει τον μέσο όρο των παλινδρομήσεων σε κάθε τμήμα.

$$\bar{X}_{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} n_i x_{(i),j}$$

##### Βήμα 4

Υπολογίζει την συνδιακύμανση του πίνακα των μέσων όρων των τμημάτων  $X$  σταθμίζοντάς τα με βάση το μέγεθος κάθε τμήματος.

$$\widehat{\Sigma}_s = \frac{1}{n} \sum_{i=1}^H n_i (\overline{X}_{(i)} - \bar{X})(\overline{X}_{(i)} - \bar{X})'$$

### Βήμα 5

Βρίσκει την κατεύθυνση του αλγορίθμου μέσω της ανάλυσης των ιδιοτιμών και των ιδιοδιανυσμάτων του  $\widehat{\Sigma}_s^{-1} \widehat{\Sigma}_s$ , όπου  $\widehat{\Sigma}$  είναι η συνδιακύμανση του πίνακα των X. Έχουμε λοιπόν

$$\widehat{\Sigma}_s \widehat{\beta}_i = \lambda_i \widehat{\Sigma} \widehat{\beta}_i, i=1, \dots, p$$

όπου  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ . Το  $\widehat{\beta}_i$  είναι η iη κατεύθυνση του SIR.

### Βήμα 6

Χρησιμοποιεί κάθε κατεύθυνση του SIR για να σχηματίσει έναν γραμμικό συνδυασμό των X.

### Βήμα 7

Σχεδιάζει το Y σε σχέση με το  $\widehat{\beta}_i' X$  για να βρει τη μεταξύ τους σχέση.

KC Li [High dimensional data analysis via the SIR/PHD approach] (2000)

## 2.1.4 FROM GPS TRACES TO A ROUTABLE ROAD MAP

Η εργασία αυτή των LiliCao, JohnKrumm(2009) παρουσιάζει μια μέθοδο για την αυτόματη μετατροπή νέων ιχνών του GPS από οχήματα που κινούνται σε καθημερινή βάση στη δημιουργία ενός οδικού δικτύου. Αρχικά η μέθοδος εξομαλύνει ταίχνη του GPS μέσω μιας νέας τεχνικής συνάθροισης. Η τεχνική αυτή συγκεντρώνει τα ίχνη που ανήκουν στον ίδιο δρόμο. Με τα αποτελέσματα των ομαλών ιχνών, εφαρμόζεται ένας τυχαίος αλγόριθμος ομαδοποίησης για να δημιουργηθεί ένα γράφημα κόμβων και ακμών που εκπροσωπούν το οδικό δίκτυο. Δείχνουν πώς το δίκτυο αυτό μπορεί να χρησιμοποιηθεί για να σχεδιαστούν λογικές διαδρομές οδήγησης. Έχει αποδειχθεί ότι οι αλγόριθμοί τους, χρησιμοποιώντας πραγματικά δεδομένα GPS που συλλέγονται σε δημόσιους δρόμους, αξιολογούν την αποτελεσματικότητα της προσέγγισης αυτής συγκρίνοντας τα αποτελέσματα του σχεδιασμού της διαδρομής που προτείνεται από το γράφημα που δημιουργήθηκε σε έναν σχεδιασμό μιας μαζικής διαδρομής.

Αναπτύχθηκε μια νέα μέθοδος για τη δημιουργία ενός οδικού χάρτη από ίχνη GPS μέσω οχημάτων που κινούνται επί καθημερινής βάσης στο συγκεκριμένο οδικό δίκτυο. Αυτή είναι μια εναλλακτική λύση δημιουργίας ενός χάρτη χωρίς να χρειάζεται να χρησιμοποιήσουμε ειδικούς οδηγούς και οχήματα. Για να γίνει αποδεκτός ο θόρυβος στα δεδομένα GPS, παρουσιάστηκε μια καινοτόμος προσέγγιση, χρησιμοποιώντας προσομοιώσεις των φυσικών έλξεων μεταξύ των ιχνών. Αυτή η διαδικασία ομαδοποιεί τα ίχνη που κινούνται στον ίδιο δρόμο, καθώς και στην ίδια κατεύθυνση του δρόμου αυτού. Παρουσιάζεται ένας απλός αλγόριθμος για τη συγχώνευση των συγκεκριμένων ιχνών σε μια αναπαράσταση του οδικού δικτύου σε σχέση με κόμβους και ακμές, το οποίο θα αναλύσουμε σε επόμενο κεφάλαιο.

### 2.1.5 MINING GPS TRACES FOR MAP REFINEMENT

Παρά την αυξανόμενη δημοτικότητα των συστημάτων οδήγησης, οι υπάρχοντες ψηφιακοί χάρτες εξακολουθούν να είναι ανεπαρκείς για πολλές προηγμένες εφαρμογές. Σε αυτή την εργασία, παρουσιάζεται μία προσέγγιση για να προκύψουν χάρτες υψηλής ακρίβειας από τα ίχνη των οχημάτων που είναι εξοπλισμένα με μεταβλητούς δέκτες GPS (σε σχέση με την προηγούμενη εργασία που χρησιμοποιούσε δέκτες οχημάτων που κινούνταν καθημερινά στους δρόμους). Θεωρείται ότι, λόγω της υπερβολικής χρήσης αυτών των εφαρμογών, τα ειδικά εξοπλισμένα οχήματα θα διατίθενται σχεδόν δωρεάν. Η προσέγγιση αποτελείται από διαδοχικά στάδια επεξεργασίας, τα οποία είναι τα εξής: 1) οι ατομικές τροχιές των οχημάτων χωρίζονται σε οδικά τμήματα και διασταυρώσεις, 2) ένας οδικός άξονας δημιουργείται για κάθε τμήμα, 3) οι θέσεις των γραμμών καθορίζονται από τις ομαδοποιήσεις των κάθετων μετατοπίσεων και 4) οι αλλαγές των ιχνών μεταξύ των τμημάτων που χρησιμοποιούνται στην παραγωγή των μοντέλων διασταύρωσης. Οι συγγραφείς περιγράφουν μια προσέγγιση στην εξόρυξη δεδομένων με έναν συνεχόμενο τρόπο.

Η προσέγγιση για την αυτόματη επαγωγή υψηλής ακρίβειας οδικών χαρτών χρησιμοποιεί, όπως προαναφέραμε, τα παγκόσμια δεδομένα εντοπισμού θέσης που προέρχονται από δοκιμαστικά οχήματα.

Η προσέγγιση αυτή χωρίζει τμηματικά τα ίχνη σε οδικά τμήματα και διασταυρώσεις. Αυτή η διαμέριση μπορεί να χρησιμοποιήσει κάποιον μαζικό διαθέσιμο χάρτη είτε να εξαχθεί το συμπέρασμα από έναν αλγόριθμο ομαδοποίησης ιχνών. Η τελευταία μέθοδος βασίζεται σε χάρτες από το μηδέν.

Ο τρόπος με τον οποίο λειτουργεί η συγκεκριμένη μέθοδος είναι η παρακάτω: υπολογίζει έναν οδικό άξονα για κάθε τμήμα χρησιμοποιώντας ένα σταθμισμένο splinesελαχίστων τετραγώνων που αντιπροσωπεύει το γεωμετρικό σχήμα του. Οι θέσεις των μονοπατιών καθορίζονται από έναν αλγόριθμο ομαδοποίησης.

Το σύστημα προσδιορίζει την έκταση των περιοχών διασταύρωσηςανιχνεύοντας τις αποκλίσειςπου παρουσιάζουν οιαλλαγές των ίχνώνμεταξύ των τομέων. Στη συνέχεια, κατασκευάζει ένα μοντέλο διασταύρωσης που περιλαμβάνει μια σειρά από ξεχωριστές διαδρομές για κάθε δυνατό ζεύγος μονοπατιών εισόδου και εξόδου.

Με τη διαδικασία της επαναληπτικής μεθόδου, δίνεται η δυνατότητα να αυξηθεί η ακρίβεια των εξαγόμενων χαρτών.Ένα παρόμοιο σύστημα θα μπορούσε να εφαρμοστεί, ώστε να χαρακτηριστούν καλύτεραοι διασπάσεις καιοι συγχωνεύσεις των μονοπατιών.

Προφανώς, το σύστημααυτό δεν μπορεί να λάβει υπόψιν τα "πραγματικά" όρια των μονοπατιών κυκλοφορίας λόγω των σημάνσεων, αλλά βασίζεται στην παρατηρούμενη συμπεριφορά των οδηγών. Αν κάποια τοποθεσία παρουσιάζει μια σταθερή τάση προς τη μία πλευρά ενός μονοπατιού, ο χάρτης που έχει δημιουργηθεί θα αντικατοπτρίζει το γεγονός αυτό. Επιπλέον, τα μοντέλα μετάβασης-διασταύρωσης βοηθούν να καθοριστείη αλλαγή τωνμονοπατιών, ακόμη και αν δεν υπάρχουν καθόλου σχεδιασμένα μονοπάτια. Η τελική εφαρμογή θα μας βοηθήσει να επιλέξουμε τελικά ποια από αυτές τις δύο πτυχές πρέπει να προτιμηθεί.

Συνοπτικά, οι μέθοδοι που έχουν περιγραφεί παραπάνω έχουν την ικανότητα τόσο να κατασκευάσουν αυτόματα νέους ψηφιακούς οδικούς χάρτες από τα ίχνη τωνGPS, όσο και να αυξήσουν την ακρίβεια στους αρχικούς χάρτες, αρκεί αυτοί να είναι διαθέσιμοι. Μέσω αυτών των μεθόδων υπάρχουν μεγάλες δυνατότητες για την αύξηση της ακρίβειας και της λεπτομέρειας των μελλοντικών χαρτών, μειώνοντας παράλληλα σημαντικά το κόστος της δημιουργίας και της διατήρησής τους. Όσο μεγαλύτερη ενσωμάτωση πληροφοριών έχουμε, τόσο μεγαλύτερη δυνατότητα βελτίωσης των χαρτών θα προκύπτει.

(STEFAN SHROEDL, KIRI WAGSTAFF, SETH ROGERS, PAT LANGLEY, CHRISTOPHER WILSON (2004))

## 2.1.6 ON VEHICLE TRACKING DATA-BASED ROAD NETWORK GENERATION

Στην εργασία που θα αναλύσουμε των SophiaKaragiorgou, DieterPfoser (2012) θα δούμε τη χρησιμότητα των οδικών δικτύων και τον ρόλο που παίζουν στις βάσεις δεδομένων. Παρατηρούμε, λοιπόν, πόσο σημαντικά είναι τα οδικά δίκτυα, όταν χρησιμοποιούνται ως σύνολα δεδομένων για έναν αυξανόμενο αριθμό εφαρμογών.



Παρολαυτά, η δημιουργία και κατά κύριο λόγο η συντήρηση τέτοιων συνόλων δεδομένων έχουν αναπτύξει πολλές και ενδιαφέρουσες ερευνητικές προκλήσεις. Η συγκεκριμένη εργασία προτείνει έναν αλγόριθμο αυτόματης δημιουργίας οδικού δικτύου που λαμβάνει τα δεδομένα εντοπισμού του οχήματος με τη μορφή τροχιών (μέσω GPS) ως είσοδο και παράγει ένα γράφημα οδικού δικτύου. Ωστόσο, η μέθοδος αυτή αντιμετωπίζει τις προκλήσεις των εξελισσόμενων συνόλων δεδομένων χαρτών, εστιάζοντας ειδικότερα σε τρεις βασικές προκλήσεις: 1) την παραγωγή των αυτόματων γνωρισμάτων του χάρτη (μέσω των βαρών), 2) τα αυτόματα γνωρίσματα του οδικού δικτύου, και 3) την πρόβλεψη της αξιολόγησης της ποιότητας.

Ουσιαστικά, ο αλγόριθμος (θα τον αναλύσουμε στον επόμενο κεφάλαιο) εκμεταλλεύεται τις αλλαγές που δημιουργούνται στη μορφή του μοτίβου χρησιμοποιώντας τις στροφές ως μέσο για τον εντοπισμό των κόμβων της διασταύρωσης. Η διασταύρωση των κόμβων αυτών χρησιμοποιείται αποτελεσματικά για να συσσωρευτούν οι τροχιές των οχημάτων και οι δεσμοί μεταξύ των διασταύρωσεων που προέρχονται από τις ενώσεις τους. Ο αλγόριθμος αυτός παράγει οδικά δίκτυα που έχουν σχεδόν τη μορφή του πραγματικού οδικού δικτύου.

Οι ιδιότητες των στοιχείων παρακολούθησης, όπως η συχνότητα δειγματοληψίας και η ταχύτητα του οχήματος, επηρεάζουν ιδιαίτερα το παραγόμενο αποτέλεσμα.

Ο έλεγχος της μεθόδου με τη χρήση πρόσθετων πηγών δεδομένων παρακολούθησης, τα οποία θα έχουν διαφορετικούς ρυθμούς δειγματοληψίας, καθώς και γεωγραφικές τοποθεσίες, επειδή ο αλγόριθμος έχει πολλές παραμέτρους που πρέπει να είναι συντονισμένες σε μια συγκεκριμένη ρύθμιση, θα ήταν καλό να μελετηθούν σε μεταγενέστερες εργασίες.

### 2.1.7 SEGMENTATION-BASED ROAD NETWORK CONSTRUCTION

Η τελευταία εργασία που θα αναπτύξουμε και είναι γραμμένη από τους Sophia Karagiorgou, Dieter Pfoser και Dimitrios Skoutas (2013) προτείνει μια νέα μέθοδο που έχει την ικανότητα να μετατρέπει κινητές τροχιές σε ένα ιεραρχικό δίκτυο μεταφορών. Χρησιμοποιεί έναν βελτιωμένο αλγόριθμο για την κατασκευή του χάρτη χωρίζοντας τα δεδομένα εισόδου σε τμήματα βάσει των τύπων κίνησης που έχουν. Τα παραγόμενα ιεραρχικά στρώματα του οδικού δικτύου συνδυάζονται σε ένα ενιαίο δίκτυο. Η διαμέριση αυτή αντιμετωπίζει τις προκλήσεις που προκαλούνται από τον θόρυβο, τον χαμηλό ρυθμό δειγματοληψίας των τροχιών και προβλέπει έναν μηχανισμό για να μπορεί ο χάρτης να ενημερώνεται αυτόματα.

Η περιγραφή της νέας αυτής προσέγγισης στο πρόβλημα της δημιουργίας ή της κατασκευής ενός χάρτη βασίζεται στην τμηματοποίηση του συνόλου δεδομένων των

τροχιών, χρησιμοποιώντας τα προφίλ των ταχυτήτων και κατασκευάζοντας χωριστά τμήματα του χάρτη. Συνεχίζοντας, συγχωνεύει όλα τα παραπάνω σε ένα ενιαίο οδικό δίκτυο.

## 2.2 ΑΝΑΛΥΣΗ ΑΛΓΟΡΙΘΜΩΝ

Ήδη στον τομέα της κατασκευής δικτύων υπάρχουν μελέτες και σχετικές εργασίες (αναπτύχθηκαν παραπάνω) με μοντέλα τα οποία θα χρησιμοποιήσουμε για να μπορέσουμε να δημιουργήσουμε, με τα δεδομένα που έχουμε, έναν θαλάσσιο χάρτη που θα βασίζεται στις κινήσεις των πλοίων. Οι μελέτες αυτές βασίζονται στις κινήσεις των τροχιών των αντικειμένων μέσω εκπομπής σημάτων GPS με στόχο τη λεπτομερή παρακολούθηση της κίνησης που διαγράφουν τα οχήματά μας. Τα ίχνη μας δίνουν τη δυνατότητα να κατανοούμε πλήρως το πώς κινείται ένα αντικείμενο (στην περίπτωση μας, πλοίο). Με τη συνεχή καταγραφή των ιχνών από το GPS που λαμβάνουμε παρατηρούμε σημαντικά στοιχεία για την πορεία που ακολουθεί ένα κινητό αντικείμενο. Έτσι, μπορούμε να δούμε προς ποια κατεύθυνση κινείται, την απόσταση που διανύει, την ταχύτητα που έχει, το πότε σταματάει για μικρό χρονικό διάστημα (pause), αν στρίβει, ή αν κάνει στάση μεγάλης διάρκειας (stop). Σύμφωνα με τα παραπάνω στοιχεία, θα μπορέσουμε να αναπαραστήσουμε γραφικά την κίνηση των πλοίων και να δημιουργήσουμε με τη βοήθεια των ήδη υπαρχόντων αλγορίθμων ένα θαλάσσιο δίκτυο που θα δώσει τη δυνατότητα να κατανοήσουμε και να εξελίξουμε τα θαλάσσια δίκτυα με βάση την εκπομπή των ιχνών των πλοίων μέσω αυτόματων συστημάτων εντοπισμού θέσης.

Οι αλγόριθμοι που θα χρησιμοποιηθούν βρίσκονται στο ([www.mapconstruction.org](http://www.mapconstruction.org)). Ο στόχος και ο σκοπός της εργασίας αυτής είναι η ανάλυση περισσότερων του ενός μοντέλου, ώστε να μπορέσουμε να συγκρίνουμε τα αποτελέσματα και να καταλήξουμε στον αλγόριθμο που θα προσεγγίσει με μεγαλύτερη ακρίβεια το θαλάσσιο δίκτυο που θέλουμε να ανακαλύψουμε.

Πριν αρχίσουμε με την εξαγωγή των αποτελεσμάτων, θα αναλύσουμε διεξοδικά τους αλγορίθμους αυτούς, δηλαδή τα τεχνικά χαρακτηριστικά, καθώς και το πώς λειτουργεί ο καθένας από αυτούς.

### 2.2.1 ΠΡΩΤΟΣ ΑΛΓΟΡΙΘΜΟΣ

Θα αρχίσουμε με τον αλγόριθμο που σχεδίασαν οι **M. Ahmed** και **C. Wenk**. (2012)

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ  
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

Ο συγκεκριμένος αλγόριθμος βασίζεται στην εισαγωγή τροχιών. Αρχικά ξεκινάει με ένα άδειο γράφημα  $G$  και στη συνέχεια το ενημερώνει με την προσθήκη ενός στοιχείου κάθε φορά. Ο αλγόριθμος αποτελείται από τρία κύρια στάδια:

- 1) Υπολογίζει τα τμήματα που ταυτίζονται (για τα οποία έχουμε μία αντιπροσωπευτική ακμή στο  $G$ ), καθώς και τα μη ταυτιζόμενα τμήματα της καμπύλης,
- 2) Ενημερώνει το  $G$  για τα ταυτιζόμενα τμήματα,
- 3) Δημιουργεί ή ενημερώνει νέες ακμές και κορυφές για τα μη ταυτιζόμενα τμήματα.

Στάδιο 1:

Για το πρώτο στάδιο εισάγουμε έναν απλό μερικό αλγόριθμο απόστασης Fréchet. Υπολογίζουμε την απόσταση Fréchet για το γράφημα  $G$  με την καμπύλη  $I$ . Παρόμοια, ο στόχος είναι να υπολογίσουμε ένα Ιμονότονο μονοπάτι, με το ελάχιστο σταθμισμένο μήκος κάτω από τη  $L_1$ -νόρμα από οποιοδήποτε αριστερό τελικό σημείο σε οποιοδήποτε δεξί τελικό σημείο στην επιφάνεια του ελεύθερου χώρου. Επειδή, όμως, το αντικείμενο είναι μονόπλευρο, μας ενδιαφέρει να ταιριάζουμε την καμπύλη με τη γραφική παράσταση που μας επιτρέπει να παρακάμπουμε τμήματα του γραφήματος. Ως εκ τούτου, μετράμε το μήκος του μονοπατιού μόνο εντός των μαύρων τμημάτων και μόνο κατά μήκος της κατεύθυνσης της καμπύλης. Αυτό δικαιολογείται στο ότι δεν υπάρχουν δύο οδικά τμήματα με λιγότερη απόσταση από  $3\epsilon$  το ένα από το άλλο, αν δεν έχουν μια κοινή κορυφή.

Στάδιο 2:

Στο συγκεκριμένο στάδιο ενημερώνουμε κάθε άκρη του γραφήματος ότι η καμπύλη μετατοπίζεται. Χρησιμοποιούμε τον αλγόριθμο ελάχιστης σύνδεσης σε ένα σύνολο των κορυφών που υπολογίζεται ως συνδυασμός του ταυτιζόμενου τμήματος της καμπύλης με την πολυγωνική καμπύλη που αντιπροσωπεύει την άκρη.

Στάδιο 3:

Στο τελευταίο στάδιο της διαδικασίας προσθέτουμε κάθε μη ταυτιζόμενο τμήμα της καμπύλης ως νέα άκρη στο γράφημα  $G$ .

Για να είναι σωστός και να λειτουργεί ο αλγόριθμος εξαρτάται από τις ακόλουθες παραδοχές που έχουν γίνει για το γράφημα του οδικού δικτύου:

- Καλή κορυφή: Η κορυφή με βαθμό μικρότερο ή ίσο του 2 και η μικρότερη γωνία μεταξύ δύο άκρων στην κορυφή να είναι μεγαλύτερη ή ίση του  $\pi/2$ .
- Κακή κορυφή: με βαθμό μεγαλύτερο του 2 και η μικρότερη γωνία μεταξύ δύο άκρων στην κορυφή να είναι μικρότερη του  $\pi/2$ .
- Καλό οδικό κομμάτι:
  - a. όλες οι κορυφές είναι καλές εκτός από τις τερματικές,

- b. δεν υπάρχει άλλο κομμάτι δρόμου που να είναι μέσα σε απόσταση  $3\varepsilon$ ,
- c. το μήκος κάθε κομματιού του δρόμου είναι μεγαλύτερο του

$3\varepsilon / \sin a_1 + 3\varepsilon / \sin a_2$  όπου  $a_1$  και  $a_2$  είναι οι ελάχιστες γωνίες των δύο τερματικών κορυφών.

- Παραδοχές για τα δεδομένα εισόδου:
  - a. Κάθε διαδρομή εισόδου βρίσκεται με την απόσταση Fréchet.
  - b. Για κάθε οδικό κομμάτι υπάρχει ένα μονοπάτι εισόδου  $l$  έτσι ώστε,  $l^{\varepsilon/2}$  να περιέχει το  $\gamma$  ή για κάθε οδικό κομμάτι  $\gamma$  υπάρχει ένα υπομονοπάτι  $l_1 \in l$  τέτοιο ώστε, το  $\gamma$  και το  $l_1$  να έχουν απόσταση Fréchet μικρότερη ή ίση του  $\varepsilon/2$ .

## 2.2.2 ΔΕΥΤΕΡΟΣ ΑΛΓΟΡΙΘΜΟΣ

Τα βήματα που βασίζεται ο αλγόριθμος των Biagioni-Eriksson(1012) είναι τα ακόλουθα πέντε:

- 1) **Εκτίμηση πυκνότητας:** Το πλήρες σύνολο των ίχνων του GPS περνά μέσω ενός πυρήνα εκτίμησης πυκνότητας (KDE) για να παράγει ένα ενιαίο δείγμα της εκτίμησης πυκνότητας για την εν λόγω περιοχή.
- 2) **Αρχική γενιά χάρτη:** Οι κεντρικοί δρόμοι εξάγονται χρησιμοποιώντας νέες αποχρώσεις του γκρι στη σκελετοποίηση του αλγορίθμου μας.
- 3) **Ταιριαστά ίχνη στον χάρτη:** Χρησιμοποιείται η μέθοδος Viterbi για να συνδέσει κάθε σημείο του δείγματος GPS στα αρχικά ίχνη με μια άκρη στον αρχικό χάρτη, σταθμισμένη με τη μέση πυκνότητα κάτω από κάθε άκρη.
- 4) **Βελτίωση της τοπολογίας:** Εδώ τα ταυτιζόμενα ίχνη μελετούνται για να αφαιρέσουν τις άκρες χαμηλής εμπιστοσύνης, συγχωνεύοντας τους περιττούς κόμβους και βρίσκουν μία λογική άκρη μεταβάσεων.
- 5) **Βελτίωση της γεωμετρίας:** Τέλος, η αμετάβλητη τοπολογία σε αυτό το βήμα ευθυγραμμίζει τις διασταυρώσεις, αφαιρεί τις στροφές και ταιριάζει τμήματα του δρόμου για να μετατρέψει τον προκύπτοντα τοπολογικά-ακριβή οδικό χάρτη σε έναν γεωμετρικά-ακριβή οδικό χάρτη. Τον τελικό, ουσιαστικά, χάρτη.

### 2.2.3 ΤΡΙΤΟΣ ΑΛΓΟΡΙΘΜΟΣ

Σύμφωνα με τους Cao-Krumm(2009) η μεγαλύτερη πρόκληση για την επίλυση του προβλήματος της δημιουργίας ενός οδικού δικτύου είναι να εξηγηθεί ο θόρυβος στα ίχνη του GPS. Η προσέγγιση αυτή, λοιπόν, λειτουργεί σε δύο στάδια.

**Βήμα 1:** Αποσαφήνιση των ιχνών του GPS για να ελαχιστοποιηθεί η επίδραση των σφαλμάτων του. Συγκεκριμένα, ομαδοποιούμε τα κοντινά ίχνη του GPS (που προέρχονται πιθανόν από τον ίδιο δρόμο) προσομοιώνοντας τη φυσική έλξη μεταξύ τους.

**Βήμα 2:** Με βάση τα αποτελέσματα από το Βήμα 1, χρησιμοποιούμε ένα απλό γράφημα του αλγορίθμου ώστε να συμπεράνουμε τη δομή του οδικού δικτύου.

Είναι σημαντικό να σημειωθεί ότι το Βήμα 1 είναι το πιο κρίσιμο μέρος της προσέγγισης αυτής. Όσο λιγότερος είναι ο θόρυβος στα ίχνη του GPS, τόσο ευκολότερο είναι να εξαχθεί η δομή του οδικού δικτύου. Επιπλέον, πιστεύουμε ότι η τεχνική αποσαφήνισης που χρησιμοποιείται στο στάδιο 1, έχει γενική εφαρμοσιμότητα στα αυτόματα συμπεράσματα των χαρακτηριστικών του δρόμου και της κυκλοφορίας, όπως τα όρια ταχύτητας, τα είδη δρόμων καθώς και τις δομές του μονοπατιού.

Για να διευκρινίσουμε τα ίχνη από το GPS, θα προσομοιώσουμε τη φυσική έλξη μεταξύ των διαφορετικών ιχνών του GPS, έτσι ώστε τα ίχνη που προέρχονται από τον ίδιο δρόμο να ομαδοποιούνται. Θα προσομοιώσουμε δύο τύπους δυνάμεων έλξης. Και για τους δύο τύπους δυνάμεων, κάθε κόμβος-ίχνος προσωρινά “ξεπαγώνει” από την αρχική του θέση και επιτρέπεται να κινηθεί σε σχέση με τις δυνάμεις που του ασκούνται. Έπειτα, το σημείο κινείται, ξαναπαγώνει και ένα άλλο σημείο κάνει την ίδια διαδικασία, και ούτω καθεξής για όλα τα σημεία. Οι δύο τύποι φυσικής έλξης είναι οι ακόλουθοι:

**Τύπος 1:** Για κάθε κόμβο-ίχνος A, έχουμε προσομοιώσει μια δύναμη έλξης από κάθε ένα από τα άλλα τμήματα του ίχνους. Συγκεκριμένα, πρώτα βρίσκουμε την ορθογώνια κατεύθυνση προς την κίνηση του οχήματος στη θέση A. Στη συνέχεια, σχεδιάζουμε μια γραμμή μέσα από το A σε αυτή την κατεύθυνση. Όλα τα τμήματα του ίχνους του GPS διασταυρώνονται με τη γραμμή που δημιουργεί μία δύναμη έλξης που ενεργεί στο A.

**Τύπος 2:** Ενώ η ελκτική δύναμη του τύπου 1 μετριάξει σαφώς την επίδραση του θορύβου των ιχνών του GPS, η προσομοίωση από μόνη της δεν επαρκεί για τον αλγόριθμό μας. Μόνο με την ελκτική δύναμη μεταξύ των ιχνών, όλα τα ίχνη τελικά θα ομαδοποιηθούν, χωρίς να έχει σημασία πόσο μακριά βρίσκονταν αρχικά. Έτσι, θα

προσομοιώσουμε μια δύναμη έλξης από την αρχική θέση του κάθε κόμβου. Ωστόσο, δεν θα πρέπει να είναι μια βαρυτική δύναμη. Διαισθητικά, όσο μακρύτερα είναι ο κόμβος από την αρχική του θέση, τόσο ισχυρότερη πρέπει να είναι η δύναμη. Ως εκ τούτου, η δύναμη τύπου 2 λειτουργεί όπως ακριβώς και η δύναμη του ελατηρίου.

Για κάθε κόμβο-ίχνος, υπολογίζουμε τη συνισταμένη των δυνάμεων που ασκούνται. Στη συνέχεια, μετακινούμε κάθε κόμβο προς την κατεύθυνση της συνισταμένης δύναμης, κάνοντας κάθε φορά ένα μικρό βήμα. Επαναλαμβάνουμε την όλη διαδικασία μέχρι να σταθεροποιηθεί.

#### 2.2.4 ΤΕΤΑΡΤΟΣ ΑΛΓΟΡΙΘΜΟΣ

Ο συγκεκριμένος αλγόριθμος είναι των J. Davies – A. Beresford – A. Hopper.(2006)

Η εφαρμογή που βασίζεται ο συγκεκριμένος αλγόριθμος έχει να κάνει με την μετατροπή των ιχνών του GPS από τα πολλαπλά οχήματα σε έναν οδικό χάρτη. Η διαδικασία περιλαμβάνει τέσσερα βασικά στάδια:

- 1) Δημιουργείται ένα ιστογράμμα 2D που περιλαμβάνει τον αριθμό των ενημερωμένων ιχνών του GPS που βρέθηκαν σε κάθε κελί.
- 2) Βρίσκει τις ακραίες θέσεις του δρόμου.
- 3) Υπολογίζει τις θέσεις των κεντρικών δρόμων.
- 4) Καθορίζει την κατεύθυνση του ταξιδιού που επιτρέπεται κατά μήκος του εκάστοτε δρόμου.

Ο συγκεκριμένος αλγόριθμος παρουσιάζει κάποια σημαντικά τεχνικά ζητήματα στην εξαγωγή αποτελεσμάτων. Πρώτον, επειδή έχει ακανόνιστα όρια, τα τμήματα που δημιουργούνται (μέσω του γραφήματος Voronoi) είναι εξίσου ακανόνιστα. Δεύτερον, ερμηνεύει τις οδικές γέφυρες ως σταυροδρόμι. Αυτό οφείλεται στο γεγονός ότι έχουν απορριφθεί τα δεδομένα υψομέτρου κατά τον σχηματισμό του ιστογράμματος 2D. Τρίτον, παρατηρούμε ότι κάποιοι κόμβοι είναι στραβοί. Αυτό συμβαίνει λόγω της αρχικής μετατροπής των ιχνών του GPS. Τέλος, ορισμένα ζεύγη των κοντινών διασταυρώσεων έχουν συγχωνευθεί.

#### 2.2.5 ΠΕΜΠΤΟΣ ΑΛΓΟΡΙΘΜΟΣ

Ο αλγόριθμος που προτείνεται από τους S. Karagiorgou – D. Pfoser(2012) για την εξαγωγή του οδικού δικτύου περιλαμβάνει τρία βασικά βήματα:

- 1) τον προσδιορισμό των διασταυρώσεων,

- 2) τη σύνδεση των διασταυρώσεων,
- 3) τη μείωση του γραφήματος του δικτύου.

Η προσέγγιση γίνεται με τη χρήση της ταχύτητας σε συνδυασμό με μία αλλαγή της κατεύθυνσης του οχήματος. Ο αλγόριθμος ανίχνευσης-διασταύρωσης σαρώνει όλες τις τροχιές χρησιμοποιώντας τον τρόπο «θέση ανά θέση και άκρη με άκρη», λαμβάνοντας υπόψιν τις παραπάνω προϋποθέσεις και καταγράφει όλες τις θέσεις που τις πληρούν. Στη συνέχεια, ομαδοποιεί τις στροφές με ένα σύστημα 8 στροφών, στο οποίο οι μονοί αριθμοί χρησιμοποιούνται για τις κλειστές στροφές ενώ οι ζυγοί για τις ανοιχτές στροφές. Με βάση αυτό το μοντέλο οι στροφές ομοδοποιούνται σε δύο κατηγορίες:

- 1) χωρική εγγύτητα,
- 2) ομοιότητα της στροφής.

Για κάθε κόμβο διασταύρωσης που δημιουργείται καταγράφουμε δύο ιδιότητες. Πρώτον, ένα βάρος για τον κόμβο που προκύπτει από το άθροισμα των βαρών που καταγράφονται για όλες τις στροφές που αποτελούν τη συστάδα.

Δεύτερον, οι επιτρεπόμενοι ελιγμοί για κάθε κόμβο καταγράφονται και δίνεται μια διασταύρωση. Ουσιαστικά, προκύπτει ποιες είναι οι πιθανές στροφές, όπως προέρχονται από τα δεδομένα εντοπισμού του GPS.

Το αποτέλεσμα αυτού του σταδίου είναι η δημιουργία ενός οδικού δικτύου που συνδέει τους κόμβους (διασταυρώσεις) με τις συνδέσεις (τμήματα τροχιάς). Ο αλγόριθμος είναι απλός, δεδομένου ότι εξετάζει ουσιαστικά όλες τις τροχιές με βάση το αν αυτές περιέχονται στο δείγμα των στροφών, σημειώνοντας τα αντίστοιχα τμήματα της τροχιάς τους. Στη συνέχεια γίνεται συμπίκνωση των συνδέσεων. Ο αλγόριθμος λειτουργεί με τρία βήματα. Τα τρία βήματα του αλγορίθμου περιλαμβάνουν:

- 1) sorting existing link samples,
- 2) τη χρήση μιας προσωρινής περιοχής αποθήκευσης γύρω από τα δείγματα σύνδεσης για τον προσδιορισμό σχετικών τμημάτων τροχιάς,
- 3) τη ρύθμιση της γεωμετρίας των συνδέσεων με βάση τη γεωμετρία των τροχιών.

## 2.2.6 ΕΚΤΟΣ ΑΛΓΟΡΙΘΜΟΣ

Εδώ παρουσιάζεται ενός ακόμη αλγόριθμος των S. Karagiorgou – D. Pfoser – D. Skoutas (2013) με την ονομασία TraceConflation. Ο αλγόριθμος λειτουργεί ως εξής:

Πρώτον, τα δείγματα ταξινομούνται με τη σειρά, σύμφωνα με την αλλαγή της κατεύθυνσης μεταξύ των εισερχόμενων και εξερχόμενων ακμών. Στη συνέχεια, τα

δείγματα που δείχνουν μια παρόμοια κίνηση, σε όρους απόλυτης κατεύθυνσης και χωρικής εγγύτητας, ομαδοποιούνται σε συστάδες στροφής. Οι συστάδες στροφής κατασκευάζονται από κάτω προς τα πάνω, βρίσκοντας για κάθε δείγμα στροφής το σύνολο των δειγμάτων του πλησιέστερου γείτονα. Οι συστάδες στροφής που προέρχονται από διαφορετικές κατευθύνσεις της κίνησης (αριστερή στροφή εναντίον δεξιάς στροφής) αλλά αφορούν χωρικά την ίδια διασταύρωση, ομαδοποιούνται μαζί για να παράγουν έναν ενιαίο κόμβο διασταύρωσης. Αυτή η βελτιωμένη μέθοδος οδηγεί σε κόμβους διασταύρωσης που τοποθετούνται με μεγαλύτερη ακρίβεια.

Το τελικό βήμα είναι η σύντηξη των παραγόμενων στρωμάτων για τις διάφορες κατηγορίες ταχύτητας. Αυτό γίνεται σταδιακά, ξεκινώντας από τα υψηλότερα επίπεδα ταχύτητας και καταλήγοντας στα χαμηλότερα επίπεδα ταχύτητας. Η διαίσθηση για αυτό είναι ότι τα υψηλότερα επίπεδα ταχύτητας αντιστοιχούν στις λεωφόρους και στις εθνικές οδούς και μπορούν να αναπαραχθούν με μεγαλύτερη ακρίβεια. Η σύντηξη περιλαμβάνει τρία στάδια:

- 1) την εύρεση αντιστοιχιών κόμβου-διασταύρωσης μεταξύ των διαφόρων στρωμάτων του δικτύου,
- 2) την εισαγωγή νέων κόμβων διασταύρωσης πάνω στις υπάρχουσες συνδέσεις του υψηλότερου στρώματος,
- 3) την εισαγωγή νέων συνδέσεων των χαμηλότερων στρωμάτων για ασυνήθιστα τμήματα του οδικού δικτύου.

### 2.2.7 ΕΒΔΟΜΟΣ ΑΛΓΟΡΙΘΜΟΣ

Οι Edelkamp και Schroedl ήταν οι πρώτοι που πρότειναν μια προσέγγιση κατασκευής χάρτη με βάση τη μέθοδο k-means. Ο αλγόριθμός τους με τη μέθοδο pointclustering δημιουργεί οδικά τμήματα με βάση τα δεδομένα παρακολούθησης, αντιπροσωπεύοντας την κεντρική γραμμή του δρόμου και χρησιμοποιώντας μια fittedspline και παρουσιάζει τα μονοπάτια που βρίσκει. Τα μονοπάτια βρέθηκαν από την ομαδοποίηση των δεδομένων με βάση την απόστασή τους από την κεντρική γραμμή του δρόμου.

Mahmuda Ahmed · Sophia Karagiorgou · Dieter Pfoser · Carola Wenk (A Comparison and Evaluation of Map Construction Algorithms Using Vehicle Tracking Data) (2014)

### 2.2.8 ΟΓΔΟΟΣ ΑΛΓΟΡΙΘΜΟΣ



Ο αλγόριθμος των Geetal. είναι μια προσέγγιση ομαδοποίησης σημείου που εφαρμόζει τοπολογικά εργαλεία για να εξάγει την υποκείμενη δομή του γραφήματος. Η κύρια ιδέα αυτού του αλγορίθμου είναι να αποσυντεθούν τα δεδομένα εισόδου σε ομάδες που το καθένα θα αντιστοιχεί σε ένα μόνο τμήμα στο υποκείμενο γράφημα. Οι συγγραφείς υποθέτουν ότι το σύνολο των σημείων εισόδου είναι πυκνό και ο αλγόριθμός τους χρειάζεται μόνο έναν πίνακα αποστάσεων ή μία περιοχή του γραφήματος των σημείων που χρησιμοποιήσαμε ως είσοδο. Στη συνέχεια, ορίζεται μια συνάρτηση της περιοχής του γραφήματος η οποία κατανέμει σε κάθε σημείο του γραφήματος geodesic απόσταση σε ένα αυθαίρετο σημείο βάσης. Χρησιμοποιούν το γράφημα Reeb για να μοντελοποιήσουν τις συνδεδεμένες συνιστώσες του επιπέδου της αντίστροφης συνάρτησης. Τέλος, υπάρχει ένας κανονικός τρόπος για να μετρηθεί η σπουδαιότητα των χαρακτηριστικών στο γράφημα Reeb η οποία τους επιτρέπει να απλοποιηθεί εύκολα το γράφημα που προκύπτει. Παρέχουν εγγυήσεις εκτέλεσης, καθώς και μερικής ποιότητας εγγυήσεις για τις αντιστοιχίες των κύκλων. Μια ενσωμάτωση για τα άκρα, στη συνέχεια, λαμβάνεται με τη χρήση ενός κύριου αλγορίθμου καμπύλης που ταιριάζει μια καμπύλη προς τα σημεία που συμβάλλουν στην άκρη.

Mahmuda Ahmed · Sophia Karagiorgou · Dieter Pfoser · Carola Wenk (A Comparison and Evaluation of Map Construction Algorithms Using Vehicle Tracking Data) (2014)

## 2.3 ΚΑΤΗΓΟΡΙΕΣ ΑΛΓΟΡΙΘΜΩΝ

Οι παραπάνω αλγόριθμοι κατασκευής χαρτών οδικών δικτύων χωρίζονται σε τρεις κύριες κατηγορίες:

- 1) Pointclustering (συμπεριλαμβάνει τους k-means αλγορίθμους καθώς και τους KDE)
- 2) incremental track insertion (σταδιακή εισαγωγή τμημάτων)
- 3) intersection linking (συνδέσεις διασταυρώσεων).

### 2.3.1 POINT CLUSTERING

Οι αλγόριθμοι σε αυτή την κατηγορία εικάζουν ότι η είσοδος αποτελείται από ένα σύνολο σημείων τα οποία, στη συνέχεια, συγκεντρώνονται με πολλούς διαφορετικούς τρόπους για να αποκτήσουν τα τμήματα του δρόμου που τελικά δημιουργούν έναν οδικό χάρτη. Το σημείο εισόδου είτε περιλαμβάνει το σύνολο όλων των μετρήσεων κατά την πρώτη είσοδο είτε ένα πυκνό δείγμα όλων των μονοπατιών εισόδου. Εδώ, οι

διαδρομές εισόδου υποτίθεται ότι είναι συνεχείς καμπύλες που λαμβάνονται από παρεμβολές (συνήθως τμηματικά-γραμμικά) μεταξύ των μετρήσεων.

Ένας τύπος προσέγγισης, με επικεφαλής τους Edelkamp και SchrodI, χρησιμοποιεί τον αλγόριθμο k-means για να συγκεντρώσει το σύνολο των σημείων εισόδου, χρησιμοποιώντας μέτρα απόστασης (π.χ. Ευκλείδεια απόσταση) και ενδεχομένως οχήματα μέτρησης, ως προϋπόθεση για την εισάγωγησημαδιών σε σταθερές αποστάσεις κατά μήκος μιας διαδρομής.

Μια άλλη προσέγγιση χρησιμοποιεί την μέθοδο KDE για να μετατρέψει πρώτα το σημείο εισόδου σε μία density-based διακριτή εικόνα. Οι περισσότεροι από τους αλγορίθμους KDE λειτουργούν καλά είτε όταν τα δεδομένα είναι συχνά (ένα ανά δευτερόλεπτο) είτε όταν υπάρχει πλεόνασμα δεδομένων. Σε αυτή την κατηγορία ανήκει ο αλγόριθμος των Biagioni και Eriksson, καθώς και του Davies et al. Σε γενικές γραμμές, οι KDE αλγόριθμοι έχουν έναν δύσκολο χρόνο για να ξεπεράσουν το πρόβλημα των θορύβων, όταν συσσωρεύονται σε μια περιοχή. Οι δύο αλγόριθμοι που χρησιμοποιούνται σε αυτή την κατηγορία είναι και οι δύο KDE-based, αλλά χρησιμοποιούν πολύ διαφορετικές προσεγγίσεις για να εξάγουν τον χάρτη από την εκτίμηση της πυκνότητας του πυρήνα.

### 2.3.2 INCREMENTAL TRACK INSERTION

Οι αλγόριθμοι σε αυτή την κατηγορία κατασκευάζουν ένα οδικό χάρτη με βαθμιαία εισαγωγή τμημάτων σε έναν αρχικά άδειο χάρτη. Τα μέτρα απόστασης και οι κλάσεις των οχημάτων χρησιμοποιούνται για να γίνουν προσθήκες και διαγραφές κατά τη σταδιακή κατασκευή του χάρτη. Ένας από τους πρώτους αλγορίθμους αυτής της κατηγορίας ομαδοποιεί τα τμήματα απλώς για να βελτιώσει τον υπάρχοντα χάρτη και όχι για να τον υπολογίσει από την αρχή. Οι Cao-Krumm εισήγαγαν για πρώτη φορά ένα διευκρινιστικό βήμα, στο οποίο τροποποιούν τα τμήματα εισόδου εφαρμόζοντας τη φυσική έλξη στην ομάδα για να τοποθετήσουν τα ίδια τμήματα μαζί. Στη συνέχεια, τοποθετούν σταδιακά κάθε κομμάτι με τη χρήση τοπικών κριτηρίων, όπως η απόσταση και η κατεύθυνση. Οι Ahmed-Wenk παρουσίασαν μία αυξητική μέθοδο που χρησιμοποιεί την απόσταση Frechet για να ταιριάζει μερικώς τα κομμάτια στον χάρτη. Οι αλγόριθμοι των Cao-Krumm και Ahmed-Wenk χρησιμοποιούν πολύ διαφορετικές προσεγγίσεις για τη σταδιακή εισαγωγή τμημάτων. Είναι οι πιο γνωστοί αλγόριθμοι στη συγκεκριμένη κατηγορία.

### 2.3.3 INTERSECTION LINKING

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ  
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ**

Σε σχέση με το σημείο ομαδοποίησης, η σύνδεση διασταυρώσεων προσεγγίζει πρώτα τις κορυφές της διασταύρωσης του οδικού χάρτη και σε ένα δεύτερο βήμα συνδέει αυτές τις διασταυρώσεις μαζί με τον προσδιορισμό των κατάλληλων τμημάτων του δρόμου. Ενώ ένας χάρτης τελικά εξάγεται, η προσέγγισή τους λειτουργεί καλύτερα για καλά ευθυγραμμισμένους χάρτες και χρησιμοποιεί συχνά δείγματα δεδομένων των 1s ή 5s. Η μέθοδος των Karagiorgou – Pfoser στηρίζεται στην ανίχνευση αλλαγών στην κατεύθυνση της κίνησης για να συναχθούν κόμβοι διασταύρωσης και στη συνέχεια κάνουν "ομαδοποίηση" των τροχιών γύρω τους για να δημιουργήσουν τις άκρες του χάρτη.

*Table 1: Κατηγορίες Αλγορίθμων*

<b>ALGORITHMS</b>	<b>POINT CLUSTERING</b>	<b>INCREMENTAL TRACK INSERTION</b>	<b>INTERSECTION LINKING</b>
<b>AHMED AND WENK</b>		√	
<b>BIAGIONI AND ERIKSSON</b>	√		
<b>CAO AND KRUMM</b>		√	
<b>DAVIES ET AL.</b>	√		
<b>EDELKAMP AND SCHRODL</b>	√		
<b>GE ET AL.</b>	√		
<b>KARAGIORGOU AND PFOSER</b>			√

## ΚΕΦΑΛΑΙΟ 3

Στο κεφάλαιο αυτό θα προσπαθήσουμε να δώσουμε αρχικά τα βασικά μέτρα ποιότητας που χρησιμοποιούνται για τη σύγκριση των χαρτών στους παραπάνω αλγορίθμους. Θα συγκρίνουμε τα μέτρα αυτά και εν συνεχεία θα αξιολογήσουμε τους αλγορίθμους.

### 3 ΜΕΤΡΑ ΠΟΙΟΤΗΤΑΣ ΓΙΑ ΣΥΓΚΡΙΣΗ ΧΑΡΤΩΝ

Υπάρχουν δύο βασικά συστατικά για την αξιολόγηση της ποιότητας της κατασκευής ενός χάρτη:

- 1) η διαθεσιμότητα του αληθούς χάρτη εδάφους  $G$  ως μέρος των δεδομένων αναφοράς,
- 2) η ποιοτική σύγκριση που χρησιμοποιείται για την αξιολόγηση της ομοιότητας μεταξύ του κατασκευασμένου χάρτη  $C$  και του αληθούς χάρτη εδάφους  $G$ .

#### 3.1 ΜΕΤΡΑ ΠΟΙΟΤΗΤΑΣ

Στις παρακάτω παραγράφους θα αναπτύξουμε με ακρίβεια και περισσότερες λεπτομέρειες τα μέτρα ποιότητας που θα χρησιμοποιήσουμε για να συγκρίνουμε τους διάφορους αλγορίθμους. Χρησιμοποιούμε την `DirectedHausdorffDistance`, το μέτρο απόστασης `path-baseddistance`, το μέτρο απόστασης με βάση τα συντομότερα μονοπάτια (`shortestpath-baseddistance`) και τέλος το μέτρο `graph-samplingbaseddistance`. Τα δύο πρώτα μέτρα δεν έχουν ξαναχρησιμοποιηθεί για τη συγκριτική αξιολόγηση των κατασκευών οδικού δικτύου.

##### 3.1.1 DIRECTED HAUSDORFF DISTANCE

Η `Directed Hausdorff distance` δύο σειρών των σημείων  $A, B$  ορίζεται ως:

$$\vec{d}(A,B) = \max_{a \in A} \min_{b \in B} d(a, b).$$

Εδώ, το  $d(a, b)$  είναι συνήθως η Ευκλείδεια απόσταση μεταξύ των δύο σημείων που έχουμε πάρει. Διαισθητικά, η `Directed Hausdorff distance` εκχωρεί κάθε σημείο  $a$  που είναι ο πλησιέστερος γείτονας του  $b \in B$  και παίρνει το μέγιστο από όλες τις αποστάσεις μεταξύ των επιλεγμένων σημείων. Για να συγκρίνουμε δύο γραφικές παραστάσεις, εντοπίζουμε κάθε γράφημα ως το σύνολο των σημείων που καλύπτεται από όλες τις κορυφές και τα άκρα του. Εάν η `Directed Hausdorff distance` από τη

γραφική παράσταση  $C$  έως τη γραφική παράσταση  $G$  είναι το πολύ  $\varepsilon$ , αυτό σημαίνει ότι για κάθε σημείο οποιασδήποτε άκρης ή κορυφής του  $C$ , υπάρχει ένα σημείο στο  $G$  σε απόσταση το πολύ  $\varepsilon$ . Η ισοδύναμα, κάθε σημείο της  $C$  περιέχεται στο άθροισμα Minkowski του  $G$  με ένα δίσκο ακτίνας  $\varepsilon$ . Αυτό το μέτρο απόστασης δίνει μια ιδέα σχετικά με τη χωρική απόσταση για γραφικές παραστάσεις. Εάν η  $C$  είναι η κατασκευασμένη γραφική παράσταση και το  $G$  είναι το αληθινό έδαφος, όσο μικρότερη είναι η απόσταση από την  $C$  στο  $G$ , τόσο πιο κοντά είναι η γραφική παράσταση της  $C$  στο  $G$ .

### 3.1.2 PATH BASED DISTANCE

Από την άλλη η path-based distance θεωρεί τα γραφήματα ως σύνολα μονοπατιών. Η απόσταση μεταξύ δύο σετ διαδρομών υπολογίζεται με το μέτρο του Hausdorff, ενώ η απόσταση Frechet χρησιμοποιείται για τον υπολογισμό της απόστασης μεταξύ δύο διαδρομών.

Για δύο καμπύλες  $f, g$ , η απόσταση Frechet ορίζεται ως:

$$\delta_F(f, g) = \inf_{\alpha, \beta: [0,1] \rightarrow [0,1]} \max_{t \in [0,1]} d(f(\alpha(t)), g(\beta(t))),$$

όπου τα  $\alpha, \beta$  κυμαίνονται πάνω από τη συνεχή, επιρριπτική και μη φθίνουσα αναπαραμετροποίηση. Ας υποθέσουμε ότι τα  $C$  και  $G$  είναι δύο επίπεδα γεωμετρικά γραφήματα και ας πούμε ότι το  $\pi_C$  είναι ένα σύνολο μονοπατιών που παράγεται από το  $C$ , και το  $\pi_G$  είναι ένα σύνολο μονοπατιών που παράγεται από το  $G$ . Η path-based distance ορίζεται ως:

$$\vec{d}_{C,G}(\pi_C, \pi_G) = \max_{p_C \in \pi_C} \min_{p_G \in \pi_G} \delta_F(p_C, p_G)$$

Στην ιδανική περίπτωση, τα  $\pi_C$  και  $\pi_G$  θα πρέπει να είναι το σύνολο όλων των διαδρομών στις  $C$  και  $G$ , οι οποίες όμως έχουν εκθετικό μέγεθος.

### 3.1.3 SHORTEST PATH BASED DISTANCE

Οι Karagiorgou και Pfoser (2012) προτείνουν ένα μέτρο στο οποίο ουσιαστικά κάθε γράφημα είναι δείγμα, χρησιμοποιώντας τυχαία σύνολα των συντομότερων μονοπατιών. Λαμβάνοντας υπόψιν το κατασκευασμένο και το αληθινό έδαφος σαν δίκτυα  $C$  και  $G$  αντίστοιχα, ένα κοινό σύνολο από ζεύγη κόμβων (προέλευση, προορισμός) επιλέγεται και στα δύο, χρησιμοποιώντας την αναζήτηση του πλησιέστερου γείτονα, αν είναι απαραίτητο. Για όλα τα ζεύγη κόμβων, τα

συντομότερα μονοπάτια υπολογίζονται και στα δύο δίκτυα. Η γεωμετρική διαφορά / ομοιότητα μεταξύ των αντίστοιχων συντομότερων μονοπατιών χρησιμοποιείται για να εκτιμηθεί η ομοιότητα μεταξύ C και G.

Η discreteFrechetdistance και η Μέση Κάθετη απόσταση χρησιμοποιούνται για να συγκρίνουν τα συντομότερα μονοπάτια. Όσο τα "παρόμοια" συντομότερα μονοπάτια στο κατασκευασμένο δίκτυο βρίσκονται στο δίκτυο του αληθινού εδάφους, τόσο υψηλότερη είναι και η ποιότητα του δικτύου. Τα αποτελέσματα της shortestpath-based distance μπορούν να εκτιμηθούν με τη γραφική αναπαράσταση της απόστασης όλων των μονοπατιών μεταξύ τους ή με σύγκριση των μέσων τιμών για ολόκληρο το σύνολο των διαδρομών.

### 3.1.4 GRAPH-SAMPLING BASED DISTANCE

Οι Biagioni και Eriksson (2012) εισήγαγαν ένα μέτρο graph-sampling based distance για να καταφέρουν να αξιολογήσουν τη γεωμετρία και την τοπολογία των κατασκευασμένων οδικών δικτύων που αντιπροσωπεύονται από γραφήματα. Η κύρια ιδέα είναι η εξής: ξεκινώντας από μια τυχαία θέση του δρόμου να διερευνήσει την τοπολογία των γραφημάτων με την τοποθέτηση δειγματικών σημείων σε κάθε γραφική παράσταση που εξέρχεται έχοντας, όμως, μέγιστη ακτίνα. Αυτό παράγει δύο σύνολα περιοχών, τα οποία είναι χωρικά δείγματα γειτονικών γραφημάτων. Η διαδικασία δειγματοληψίας επαναλαμβάνεται για διάφορες τοποθεσίες.

Για την αντιστοίχιση bottleneck, τα σημεία του δείγματος στο ένα γράφημα μπορούν να συνεχίσουν να θεωρούνται ως «μπίλια» και από το άλλο γράφημα να θεωρούνται ως "τρύπες". Διαισθητικά, αν μία μπίλια προσγειώνεται κοντά σε μια τρύπα, τότε πέφτει μέσα. Οι μπίλιες, οι οποίες είναι πάρα πολύ μακριά από τρύπες, παραμένουν εκεί που βρίσκονται, ενώ οι τρύπες χωρίς μπίλιες γύρω τους παραμένουν κενές. Αν ένα από τα γραφήματα είναι το αληθινό-έδαφος, αυτή η διαφορά αντιπροσωπεύει την ακρίβεια του άλλου γραφήματος. Μετρώντας τον αριθμό των μπιλιών που έχουν παραμείνει στη θέση τους (δεν έχουν τοποθετηθεί δηλαδή σε τρύπα) και των άδειων τρυπών, ποσοτικοποιεί την ακρίβεια του παραγόμενου οδικού δικτύου σε σχέση με το αληθινό-έδαφος λαμβάνοντας υπόψιν δύο μέτρα. Το πρώτο μέτρο σχετίζεται με το ποσοστό των ψεύτικων μπιλιών,

$$\text{spurious} = \text{spurious\_marbles} / (\text{spurious\_marbles} + \text{matched\_marbles})$$

ενώ το δεύτερο μέτρο είναι η αναλογία άδειων τρυπών,

$$\text{missing} = \text{empty\_holes} / (\text{empty\_holes} + \text{matched\_holes}).$$

Για να παραχθεί ένα συνδυασμένο μέτρο απόδοσης από αυτές τις δύο τιμές, χρησιμοποιείται το γνωστό F-σکور, το οποίο υπολογίζεται από τον ακόλουθο τύπο:

$$F - score = 2 * \frac{precision * recall}{precision + recall}$$

όπου,  $precision = 1 - \text{spurious}$  και  $recall = 1 - \text{missing}$ .

Όσο υψηλότερο είναι το F-σکور, τόσο καλύτερο είναι το ταίριασμα. Η επαναλαμβανόμενη τοπική δειγματοληψία σε τυχαία επιλεγμένα σημεία δίνει μια ακριβή εικόνα των τοπικών γεωμετριών και τοπολογιών σε ολόκληρη τη γραφική παράσταση.

Υπάρχει και μία τροποποιημένη έκδοση της μέθοδου, η οποία αγνοεί τμήματα του οδικού δικτύου στα οποία δεν θα μπορούσε να βρεθεί καμία αντιστοιχία μεταξύ παραγόμενων δικτύων και δικτύων αληθινού-εδάφους.

### 3.2 ΣΥΓΚΡΙΣΗ ΑΠΟΣΤΑΣΗΣ ΜΕΤΡΩΝ

Όλα τα μέτρα απόστασης που περιγράφονται στην προηγούμενη ενότητα αποδίδουν διαφορετικές ιδιότητες των γραφημάτων. Ανάλογα με το είδος ομοιότητας που θέλουμε, χρησιμοποιούμε και το αντίστοιχο μέτρο απόστασης.

Αν, λοιπόν, κάποιος ψάχνει τη χωρική μετατόπιση μεταξύ δύο γραφημάτων χωρίς να ενδιαφέρεται για την τοπολογία ή την ομοιότητα της διαδρομής, θα χρησιμοποιήσει την *Directed Hausdorff distance*. Αν, όμως, ενδιαφέρεται για παρόμοια μονοπάτια και στα δύο γραφήματα, τότε η καλύτερη λύση είναι η *shortest path-based distance*.

Από την άλλη πλευρά, τα δύο μέτρα απόστασης *graph-sampling based distance* και *path-based distance* μεγιστοποιούν τη χρήση της τοπολογίας σε γραφήματα, τα οποία θέλουμε να συγκρίνουμε.

Μεταξύ αυτών των τεσσάρων μέτρων, μόνο η *graph-sampling based distance* εξασφαλίζει αντιστοιχία ένα-προς-ένα.

## ΚΕΦΑΛΑΙΟ ΤΕΤΑΡΤΟ

### 4.1 ΑΞΙΟΛΟΓΗΣΗ ΤΩΝ ΑΛΓΟΡΙΘΜΩΝ

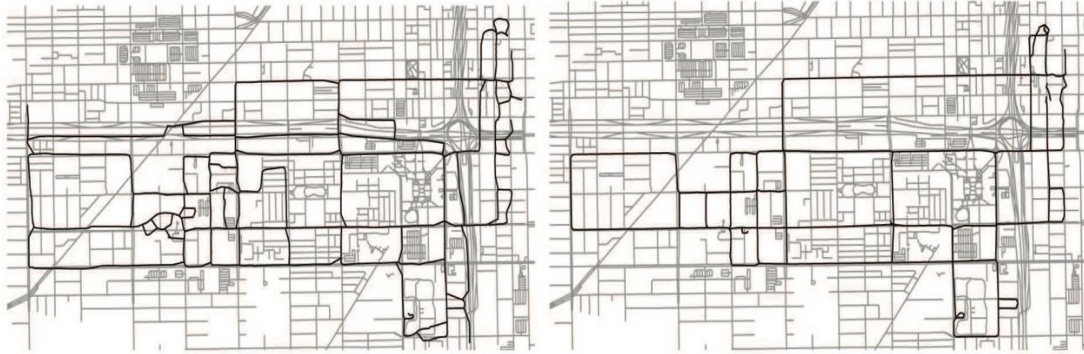
Μέσω μια πειραματικής μελέτης που παρουσιάστηκε στην εργασία MahmudaAhmed · SophiaKaragiorgou · DieterPfoser · CarolaWenk (AComparisonandEvaluationofMapConstructionAlgorithmsUsingVehicleTrackingData) (2014), θα προσπαθήσουμε να αξιολογήσουμε ποιος από τους παραπάνω αλγόριθμους είναι ο πιο αποτελεσματικός.

Κάθε ένας από τους αλγόριθμους που αναλύσαμε, χρησιμοποιεί διαφορετικές ρυθμίσεις παραμέτρων. Για τον αλγόριθμο των Ahmed και Wenk παρατηρούμε αρχικά ότι ανάλογα με τα δεδομένα που έχει να αναλύσει, διαφοροποιεί την απόσταση, δηλαδή το  $\epsilon$  που χρησιμοποιεί. Όσο μικρότερη είναι η περιοχή που θέλει να ανακαλύψει, τόσο μικραίνει την απόσταση  $\epsilon$  που χρησιμοποιεί. Με τον τρόπο αυτό, έχει καλύτερα αποτελέσματα με καλύτερη κάλυψη, αλλά υστερεί στην ακρίβεια της κατασκευής του χάρτη. Αυτό φαίνεται και από το παρακάτω γράφημα.

Παρατηρούμε ότι, ενώ έχει καλύψει το μεγαλύτερο μέρος της περιοχής, υπάρχουν κάποια σημεία, τα οποία αποκλίνουν από την πραγματική διαδρομή. Αυτό οφείλεται κυρίως στο γεγονός, ότι ο αλγόριθμος χρησιμοποιεί την απόσταση Fréchet μέσω της οποίας υπολογίζει ένα γράφημα. Προσπαθεί να ταιριάξει αυτό το γράφημα με τα τμήματα της καμπύλης που ταυτίζονται. Αφού το κάνει αυτό, προσπαθεί να δημιουργήσει νέες ακμές και κορυφές ή αν μπορέσει, να ενημερώσει τις ήδη υπάρχουσες για τα τμήματα που δεν έχουν ταυτοποιηθεί. Επειδή, όμως, δεν επιτρέπεται να υπάρχει κάποιο άλλο κομμάτι δρόμου σε μικρότερη απόσταση από αυτή του  $3\epsilon$  (ανάλογα με τι  $\epsilon$  χρησιμοποιεί κάθε φορά), έχει τις παραπάνω αποκλίσεις από το αληθινό γράφημα εδάφους.

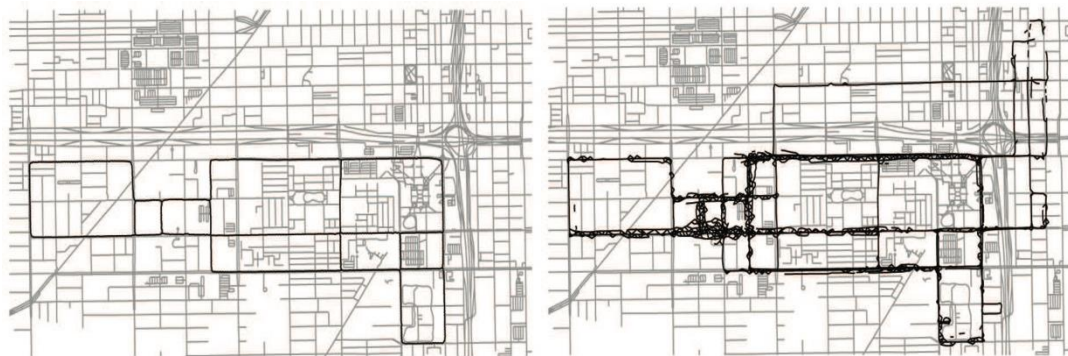
Ο παραπάνω αλγόριθμος είναι ένα απλός αυξητικός αλγόριθμος εισαγωγής αντικειμένου. Οι Ahmed και Wenk δίνουν θεωρητικές εγγυήσεις ποιότητας για τον χάρτη που δημιουργούν με τον αλγόριθμό τους, στον οποίο περιλαμβάνουν μια αντιστοιχία ένα-προς-ένα μεταξύ των καλά διαχωρισμένων τμημάτων του πραγματικού χάρτη και του χάρτη που εξάγουν, μεχρήση της απόστασης Fréchet μεταξύ των εν λόγω τμημάτων.



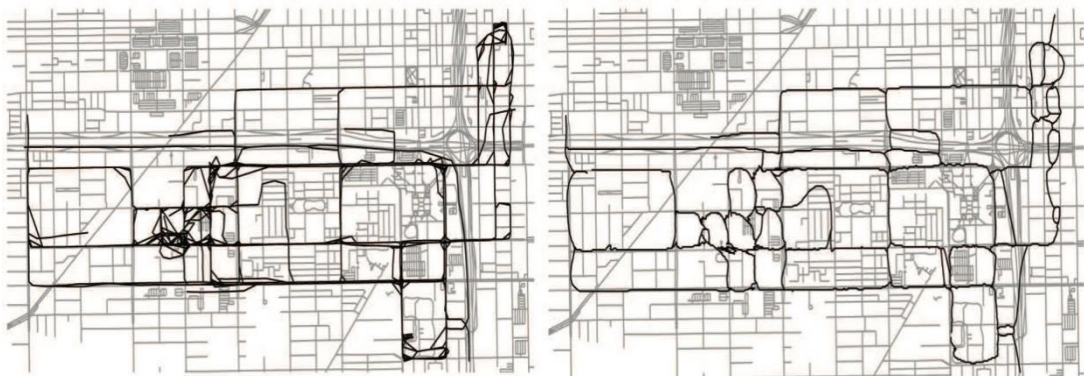


(a)Ahmed

(b)Biagioni

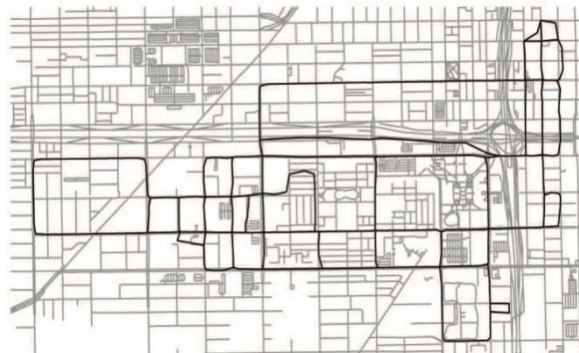


(c)Davies (d) Cao



(e) Edelkamm

(f) Ge



(g) Karagiorgou

Συγκριτικά, θα λέγαμε ότι το οδικό δίκτυο που εξάγεται από τον Davies σε σχέση με τα οδικά δίκτυα που εξάγουν οι υπόλοιποι αλγόριθμοι, έχει τέλεια ταύτιση με το αληθινό γράφημα. Η κάλυψη που παρέχουν, ωστόσο, οι άλλοι αλγόριθμοι είναι σαφώς καλύτερη από αυτή του Davies.

Παρόλου που οι αλγόριθμοι των Davies και Biagioniάνηκουν στην κατηγορία των pointclustering, βλέπουμε εντελώς διαφορετικά αποτελέσματα. Αυτό οφείλεται, κατά κύριο λόγο, στις διαφορετικές προσεγγίσεις που χρησιμοποιούν στους αλγόριθμούς τους.

Το καλύτερο αποτέλεσμα γραφικά φαίνεται να το παρουσιάζει ο αλγόριθμος της Karagiorgou. Παρόλου που σε δύο σημεία δεν έχει την σωστή τοποθέτηση το γράφημα πάνω στον χάρτη, τόσο η περιοχή κάλυψης, όσο και η ακρίβεια επιτυγχάνονται σε πολύ μεγάλο βαθμό.

Φαίνεται ότι οι αλγόριθμοι pointclustering με βάση την εκτίμηση της πυκνότητας, όπως των Biagioni και Davies, παράγουν χάρτες με χαμηλή πολυπλοκότητα (μικρός αριθμός κορυφών και ακμών), αλλά αποτυγχάνουν να ανακατασκευάσουν δρόμους, στους οποίους δεν κινούνται συχνά τα κινητά αντικείμενα. Ειδικότερα, ο χάρτης που δημιουργήθηκε μέσω του αλγορίθμου του Daviesetal. έχει σχετικά πολύ μικρή κάλυψη. Από την άλλη πλευρά, ο αλγόριθμος του Geetal. χρησιμοποίησε υποδείγματα όλων των αντικειμένων για να δημιουργήσει ένα πυκνό σύνολο, στο οποίο οφείλεται και η πολυπλοκότητα της κατασκευής των χαρτών του.

Οι αλγόριθμοι κατασκευής χάρτη, όπως του Ahmedetal. και του Caoetal. που βασίζονται στην incrementaltrackinsertion, αποτυγχάνουν να ομαδοποιήσουν τα αντικείμενα μαζί, όταν η μεταβλητότητα και το λάθος, που σχετίζονται με την είσοδο των αντικειμένων, είναι μεγάλα.

## 4.2 ΣΥΜΠΕΡΑΣΜΑ

Ένας καλός χάρτης, λοιπόν, είναι αυτός που μπορεί να συνδυάσει την ακρίβεια και την περιοχή κάλυψης. Εδώ, φαίνεται ότι οι KDE-basedpointclustering αλγόριθμοι, όπως των Biagioni και Davies, παράγουν χάρτες με χαμηλότερη πολυπλοκότητα (μικρότερος αριθμός κορυφών και ακμών) και συχνά αποτυγχάνουν να ανακατασκευάσουν δρόμους που δεν διασχίζονται αρκετά συχνά από τα οχήματα εισόδου. Από την άλλη πλευρά, ο αλγόριθμος που σχεδιάστηκε από τον Geχρησιμοποιεί σαν δείγμα όλα τα οχήματα για να δημιουργήσει ένα πολύ πυκνότερο σύνολο, που έχει σαν αποτέλεσμα την πολυπλοκότητα της κατασκευής του χάρτη.

Μια παρόμοια παρατήρηση μπορεί να γίνει για τους αλγόριθμους που βασίζονται στη σταδιακή εισαγωγή τμημάτων, όπως είναι οι αλγόριθμοι των Ahmed και Cao. Αποτυγχάνουν να συγκεντρώνουν τμήματα μαζί, όταν η μεταβλητότητα και το λάθος, που σχετίζονται με τις διαδρομές εισόδου, είναι μεγάλη.

Το αποτέλεσμα είναι οι κατασκευασμένοι οδικόι χάρτες να περιέχουν πολλαπλά άκρα για έναν μόνο δρόμο, που δημιουργεί μια μεγάλη κατασκευή, αλλά όχι κατ' ανάγκη και ένα πιο ακριβές οδικό δίκτυο.

Όσον αφορά την ποιότητα και την ακρίβεια μαζί, οι χάρτες που κατασκευάστηκαν με τη χρήση των αλγορίθμων Karagiorgou, Davies, και Biagioni έχουν γενικά μικρότερο path-based και DirectedHausdorffdistances και κατασκευάζουν χάρτες που μπορεί να θεωρηθούν πιο ακριβείς. Παρά το γεγονός ότι οι αλγόριθμοι που σχεδιάστηκαν από τον Ahmed και τον Ge παράγουν χάρτες με καλή κάλυψη και παρέχουν εγγυήσεις ποιότητας, οι path-based distances είναι μεγαλύτερες. Σε μια προσπάθεια να αξιολογηθούν και τα δύο, τόσο η ακρίβεια όσο και η κάλυψη, η shortest path-based distance δείχνει για τις περιπτώσεις των Davies και Cao καλής ποιότητας χάρτη, αλλά την ίδια στιγμή περιορισμένη κάλυψη περιοχής. Σε αυτή την αξιολόγηση, ο αλγόριθμος της Karagiorgou παράγει χάρτες που έχουν τόσο καλή κάλυψη, όσο και υψηλή ομοιότητα μονοπατιού. Δηλαδή, θα λέγαμε ότι ο αλγόριθμος της Karagiorgou προσφέρει υψηλή ακρίβεια, αλλά και μεγάλη κάλυψη της περιοχής.

Mahmuda Ahmed · Sophia Karagiorgou · Dieter Pfoser · Carola Wenk (A Comparison and Evaluation of Map Construction Algorithms Using Vehicle Tracking Data) (2014)

## ΚΕΦΑΛΑΙΟ ΠΕΜΠΤΟ

### 5 ΠΕΙΡΑΜΑΤΙΚΗ ΜΕΛΕΤΗ

Σε αυτό το κεφάλαιο θα προσπαθήσουμε να ανακαλύψουμε θαλάσσια δίκτυα χρησιμοποιώντας κάποιους από τους αλγορίθμους που είδαμε στα προηγούμενα κεφάλαια.

#### 5.1 ΠΡΟΣΕΓΓΙΣΗ

Η μελέτη μας έχει να κάνει με την προσπάθεια ανακάλυψης θαλάσσιων δικτύων μέσω των αλγορίθμων που αναλύουν ίχνη GPS. Η προσέγγιση αυτή θα αρχίσει με τη μελέτη των δεδομένων. Τα δεδομένα που λαμβάνουμε από τα συστήματα αυτόματου εντοπισμού έχουν την παρακάτω μορφή.

Table 2: AIS Δεδομένα

```
|0;null;215339000;1;2100;null;GAP_END;TURN;428661.2442;3843851.07  
16;2009-07-07 07:17:12.0;428661.2542;3843851.0816;2009-07-07  
07:17:13.0;null;GAP_START;null;448729.3687;3838745.8862;2009-07-  
07 08:03:07.0;448729.3787;3838745.8962;2009-07-07  
08:03:08.0;null;GAP_END;null;454131.3963;3837388.6277;2009-07-07  
08:15:25.0;454131.4063;3837388.6377;2009-07-07  
08:15:26.0;null;GAP_START;null;466648.0222;3834396.4319;2009-07-  
07 08:44:46.0;466648.0322;3834396.4419;2009-07-07  
08:44:47.0;null;GAP_END;null;478468.5765;3831552.2608;2009-07-07  
09:12:18.0;478468.5865;3831552.2708;2009-07-07  
09:12:19.0;null;GAP_END;null;434569.7177;3840032.9782;2009-08-11  
00:09:00.0;434569.7277;3840032.9882;2009-08-11 00:09:01.0;  
0;null;215379000;1;2100;null;GAP_END;TURN;357497.1312;3957595.72  
84;2009-06-02 23:51:43.0;357497.1412;3957595.7384;2009-06-02  
23:51:44.0;null;TURN;null;394164.4202;3956422.6459;2009-06-03  
01:11:58.0;394164.4302;3956422.6559;2009-06-03  
01:11:59.0;null;TURN;null;394202.3151;3956440.6832;2009-06-03  
01:12:04.0;394202.3251;3956440.6932;2009-06-03  
01:12:05.0;null;TURN;null;394268.7322;3956449.1365;2009-06-03  
01:12:16.0;394268.7422;3956449.1465;2009-06-03  
01:12:17.0;null;GAP_START;null;472578.9611;3955871.7057;2009-06-  
03 03:59:04.0;472578.9711;3955871.7157;2009-06-03
```

Απ' ότι παρατηρούμε, τα δεδομένα έχουν μία εντελώς διαφορετική απεικόνιση σε σχέση με αυτή που έχουν τα δεδομένα των GPS. Αρχίζοντας, λοιπόν, την προσέγγιση αυτή, το πρώτο βήμα που πρέπει να κάνουμε είναι η μετατροπή των παραπάνω

δεδομένων σε μορφές συμβατές με αυτές που δέχονται οι αλγόριθμοι που θα χρησιμοποιηθούν.

Με τη δημιουργία δύο αλγορίθμων που σχεδίασα στη γλώσσα προγραμματισμού pythonθα μετατρέψω τα δεδομένα από την αρχική τους μορφή στις αντίστοιχες μορφές που δέχονται οι αλγόριθμοι. Τα δεδομένα πρέπει να είναι στη μορφή (X Y timestamp) για τους αλγορίθμους των Ahmedκαι Karagiorgou, ενώ για τους υπολοίπους, θα πρέπει να έχουν 6 μεταβλητές. Αυτές πρέπει να είναι της μορφής (loc\_idXYtimestampprv\_loc\_idnext\_loc\_id).

Αφού ολοκληρωθεί η παραπάνω διαδικασία, θα πρέπει να ορίσουμε αντίστοιχες παραμέτρους σε κάθε αλγόριθμο, ώστε να έχουν αξία τα δεδομένα μας. Αυτό είναι ίσως το πιο δύσκολο κομμάτι αυτής της διαδικασίας. Οι αποστάσεις στη θάλασσα σε σχέση με τις αποστάσεις στην ενδοχώρα έχουν τεράστιες διαφορές. Η οριοθέτηση των διαδρομών, μέσω των αποστάσεων, γίνεται πολύ δύσκολη. Η ομαδοποίηση πλοίων που μπορεί να κινούνται σε αντίθετες κατευθύνσεις, καθώς και η μη ομαδοποίηση αυτών που κινούνται ομόρροπα λόγω των μεγάλων αποστάσεων που μπορεί να έχουν μεταξύ τους, ίσως δημιουργήσει τεράστιο πρόβλημα στην εξαγωγή χρήσιμων συμπερασμάτων.

## 5.2 DATASET

Η βάση δεδομένων που έχουμεαποτελείται από 1285 πλοία. Κάθε πλοίο στέλνει δεδομένα μέσω του συστήματος AIS στην παρακάτω μορφή:

Κάθε γραμμή (στο παραπάνω παράδειγμα έχω βαλει μία γραμμή ενός πλοίου) αποτελείται από ένα semantictrajectory με μορφή: profile\_id, trajectory\_tag, object\_id, traj\_id, srid και στη συνέχεια, με επαναλαμβανόμενο τρόπο, στέλνει τα παρακάτω δεδομένα:

defining\_tag, episode\_tag, activity\_tag, minx, miny, mint, maxx, maxy, maxt.

Table 3: AIS Δεδομένα 2

0	null	215171000	1	2100	null	GAP_END TURN	852977,7	3966368,389	02:25,0	852977,7	3966368	02:26,0
0	null	215213000	1	2100	null	GAP_END TURN	425043,8	3847083,015	29:27,0	425043,8	3847083	29:28,0

Διαβάζοντας και προσπαθώντας να ερμηνεύσουμε τα παραπάνω δεδομένα παρατηρούμε ότι:

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ  
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

- 1) τα έξι πρώτα κελιά, εκτός από το τρίτο που αναφέρεται στο `object_id`, τα οποία είναι μοναδικά για κάθε πλοίο λαμβάνουν την ίδια τιμή.

Συνεχίζοντας στις επόμενες δύο στήλες:

- 2) Στην έβδομη στήλη η τιμή ή καλύτερα το γνώρισμα που δίνεται στο πλοίο είναι το `GAP_END` ή το `GAP_START`. Μια πρώτη εκτίμηση που μπορούμε να κάνουμε είναι ότι το πλοίο βρίσκεται στον τελικό ή στον αρχικό του προορισμό,
- 3) στην επόμενη λαμβάνονται άλλες δύο "τιμές" εκτός από το `GAP_END` και το `GAP_START` οι οποίες είναι οι `TURN` και `null`,
- 4) οι τελευταίες 6 στήλες μας δίνουν από μία τριάδα `X Y` και `timestamp`, με διαφορά πάντα ενός δευτερολέπτου ανά δύο ζεύγη.

Στους αλγορίθμους που θα χρησιμοποιήσουμε χρειαζόμαστε μόνο τα στοιχεία των στηλών 9-14 και φυσικά όλα τα αντίστοιχα στοιχεία που στέλνει κάθε πλοίο ξεχωριστά. Αφού εξαγάγουμε τα νέα στοιχεία με την χρήση των προγραμμάτων μετατροπής βάσης που κατασκευάσα, θα πάρουμε τα παρακάτω αποτελέσματα.

```
630072.1963 4370116.0989 63415097949
630072.2063 4370116.1089 63415097950
628764.5039 4367173.8943 63415098462
628764.5139 4367173.9043 63415098463
626778.5859 4362841.6542 63415099211
```

Στην τελευταία στήλη αναφέρεται η ημερομηνία, χωρίς να υπολογίζονται τα δέκατα του δευτερολέπτου. Ενώ τρέχοντας και τον δεύτερο αλγόριθμο (αναφέρονται αναλυτικά στο παράρτημα), θα λάβουμε τα εξής αποτελέσματα:

```
1,630072.2063,4370116.1089,63415097950,None,2
2,628764.5139,4367173.9043,63415098463,1,3
3,626778.5959,4362841.6642,63415099212,2,4
4,622702.7243,4353954.4731,63415100771,3,5
5,620976.9737,4350179.7149,63415101445,4,6
6,618858.1127,4345517.6451,63415102259,5,7
```

Η πρώτη ιδιαιτερότητα που έχουμε σε σχέση με τα αρχικά δεδομένα είναι ότι έχουμε εισάγει `loc_id`. Το μεγαλύτερο πρόβλημα που παρουσιάζεται στα δεδομένα AIS που έχουμε είναι η έλλειψη του `loc_id`. Με έναν απλό τρόπο επιχειρούμε να "ξεγελάσουμε" τον αλγόριθμο χρησιμοποιώντας έναν αύξοντα αριθμό για την τοποθεσία του πλοίου. Όπως παρατηρείται στα δεδομένα που λαμβάνονται με

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ  
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

GPS, εκτός από τα X Y και timestamp, υπάρχει και η ακριβής θέση του κινητού αντικειμένου.

Table 4: GPS Δεδομένα

```
112090457,41.866969,-87.671384,1303742017.0,None,112090770
112090770,41.86678,-87.671336,1303742039.0,112090457,112090850
112090850,41.866713,-87.671036,1303742044.0,112090770,112090901
112090901,41.866716,-87.670759,1303742047.0,112090850,112090950
112090950,41.866725,-87.670418,1303742050.0,112090901,112090985
112090985,41.866733,-87.670168,1303742052.0,112090950,112091020
112091020,41.866739,-87.669907,1303742054.0,112090985,112091053
112091053,41.866742,-87.669644,1303742056.0,112091020,112091090
112091090,41.866748,-87.669382,1303742058.0,112091053,112091129
112091129,41.866748,-87.669124,1303742060.0,112091090,112091166
112091166,41.866748,-87.668867,1303742062.0,112091129,112091202
112091202,41.866741,-87.668618,1303742064.0,112091166,112091236
112091236,41.866737,-87.668365,1303742066.0,112091202,112091275
112091275,41.866734,-87.668117,1303742068.0,112091236,112091337
112091337,41.866734,-87.66776,1303742071.0,112091275,112091394
112091394,41.866747,-87.667399,1303742074.0,112091337,112091453
112091453,41.866767,-87.667047,1303742077.0,112091394,112091517
112091517,41.866774,-87.666755,1303742080.0,112091453,112091617
112091617,41.866776,-87.66649,1303742085.0,112091517,112091765
112091765,41.866877,-87.666277,1303742092.0,112091617,None
```

Η βασική διαφορά των δεδομένων αυτών είναι κυρίως ο ρυθμός λήψης τους. Ένα αυτοκίνητο στέλνει σήμα σε μικρά και τακτά χρονικά διαστήματα. Αυτό βοηθάει στη συλλογή στοιχείων σχετικά με την κίνηση που κάνει το όχημα και την απόσταση που διανύει, με αποτέλεσμα να έχουμε την ικανότητα να αξιοποιήσουμε στον μέγιστο βαθμό τέτοιου είδους δεδομένα.

Στα πλοία και ειδικότερα σε αυτά που εκπέμπουν AIS δεδομένα αυτό δεν ισχύει. Η διαφορά μεταξύ δύο σημάτων μπορεί να είναι τεράστια, χωρίς να ξέρουμε πού βρίσκεται το πλοίο, αφού απουσιάζει η loc\_id.

### 5.3 ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΛΓΟΡΙΘΜΩΝ

Με τα δεδομένα που είχαμε, καταφέραμε να τρέξουμε με πολλές αλλαγές τρεις αλγόριθμους. Τον αλγόριθμο του Ahmed, του Cao και του Edelkamp.

Ξεκινώντας από τον αλγόριθμο του Ahmed, τα αποτελέσματα που έβγαλα για τις τροχιές των 1285 πλοίων είναι τα παρακάτω:

Table 5 Ahmed-1

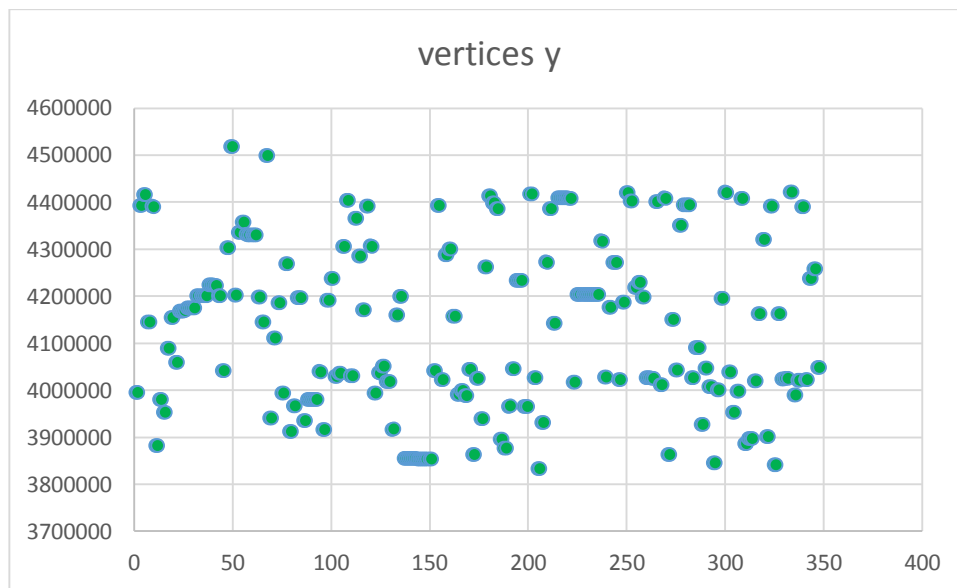


Table 6 Ahmed-2

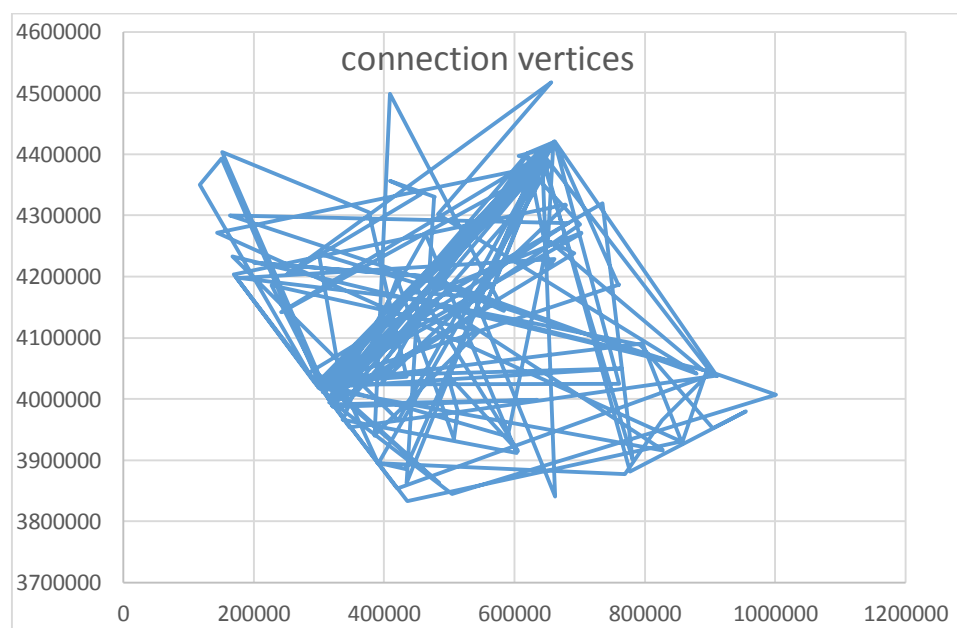
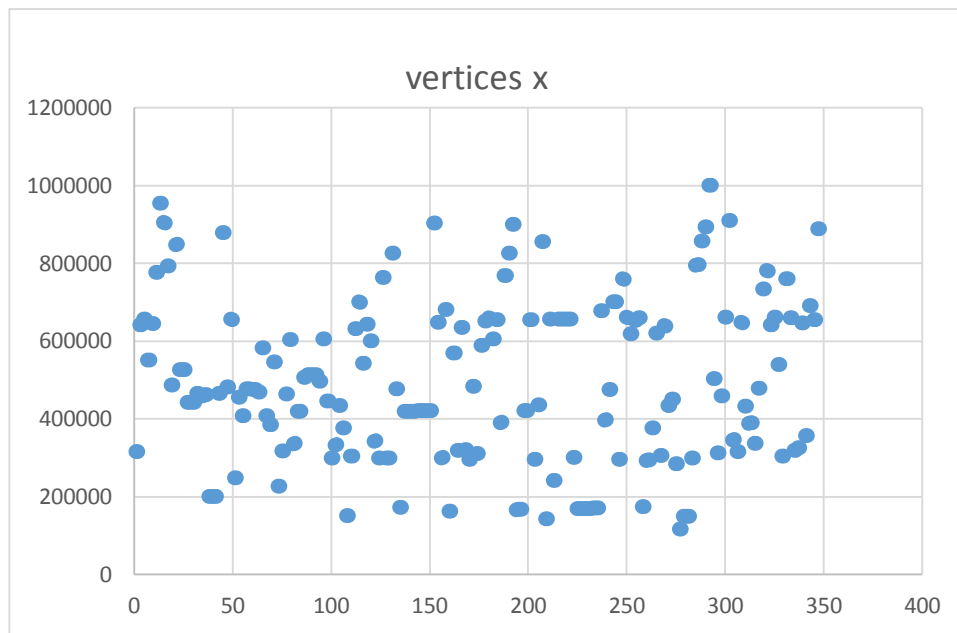




Table 7 Ahmed-3



Ενώ, λοιπόν, ο αλγόριθμος τρέχει και διαβάζει όλα τα δεδομένα των πλοίων και εξάγει δύο αρχεία των edges.txt και των vertices.txt, στη συνέχεια σταματάει τη διαδικασία χωρίς να εμφανίσει κάποιο σφάλμα. Οι παράμετροι που δώσαμε, ώστε να καταφέρει να τρέξει ο αλγόριθμος, είναι οι ακόλουθοι:

```
EPS=1000.0#epsilon in meters  
HAS_ALTITUDE=false #if input file has altitude information  
ALT_EPS=50.0#minimum altitude difference in meters between two streets
```

To false σημαίνει ότι τα δεδομένα που δώσαμε είναι της μορφής (X Y timestamp). Η διαδικασία, ωστόσο, σταματάει εκεί. Δεν εξάγει ο αλγόριθμος τίποτα περισσότερο. Η παραπάνω διαδικασία δεν ήταν ιδιαίτερα χρονοβόρα σε σχέση με τη μεγάλη βάση δεδομένων που είχαμε.

Ο επόμενος αλγόριθμος που έτρεξε τα δεδομένα μας είναι του Cao. Η διαδικασία κράτησε πολλές ώρες (τουλάχιστον 36). Το αποτέλεσμα ήταν να εξάγει ο αλγόριθμος μία νέα βάση δεδομένων αποτελούμενη από τα edges και τα nodes και να τα παρουσιάζει γραφικά, όπως φαίνεται παρακάτω.

Table 8 Cao-1

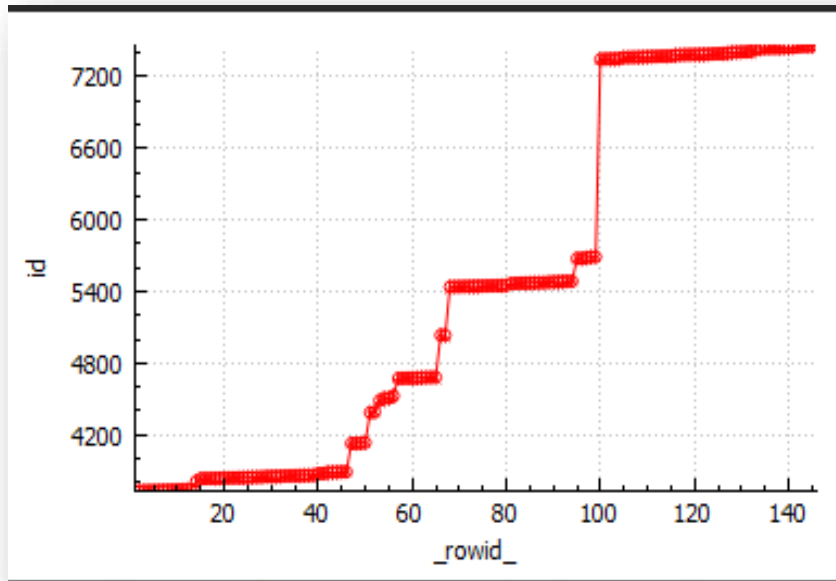


Table 9 Cao-2

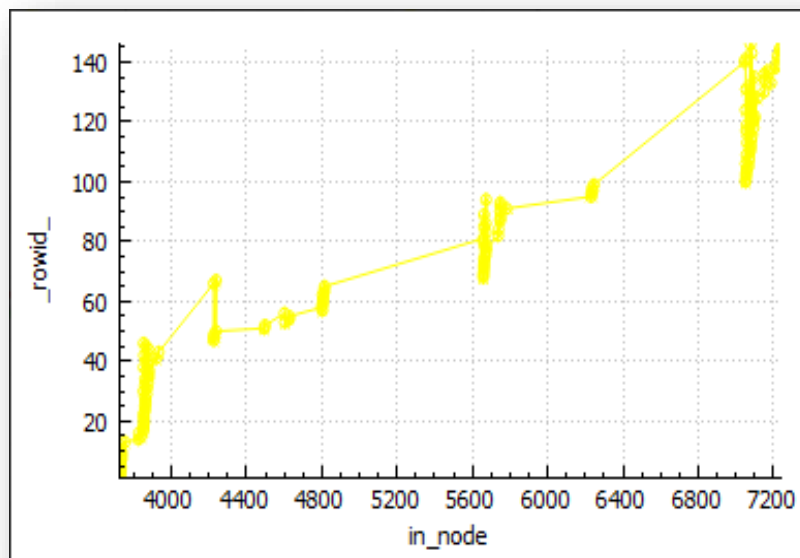
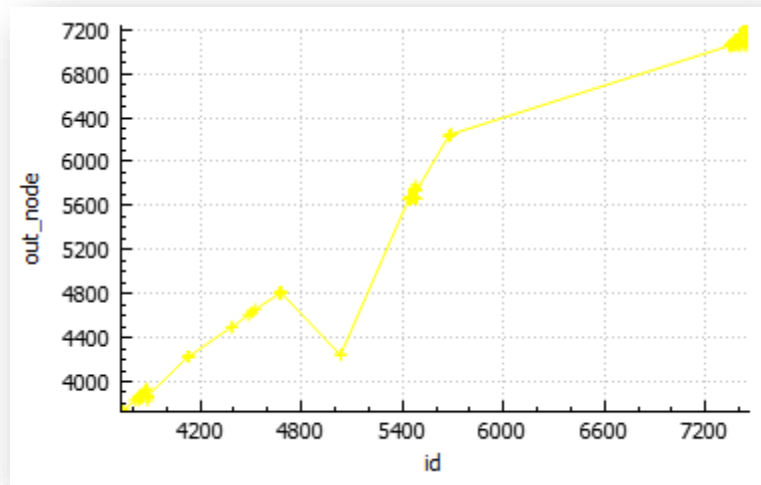


Table 10 Cao-3



Και μετά από την εξαγωγή των αποτελεσμάτων ο αλγόριθμος ολοκληρώθηκε.

Τρίτος και τελευταίος από τους αλγορίθμους που έτρεξαν είναι του Edelkamp. Και αυτός έκανε ακριβώς την ίδια διαδικασία με τον προηγούμενο. Εξήγαγε μία νέα βάση δεδομένων που περιελάμβανε και αυτή τα δύο παραπάνω αρχεία.

Table 11 Edelkamp-1

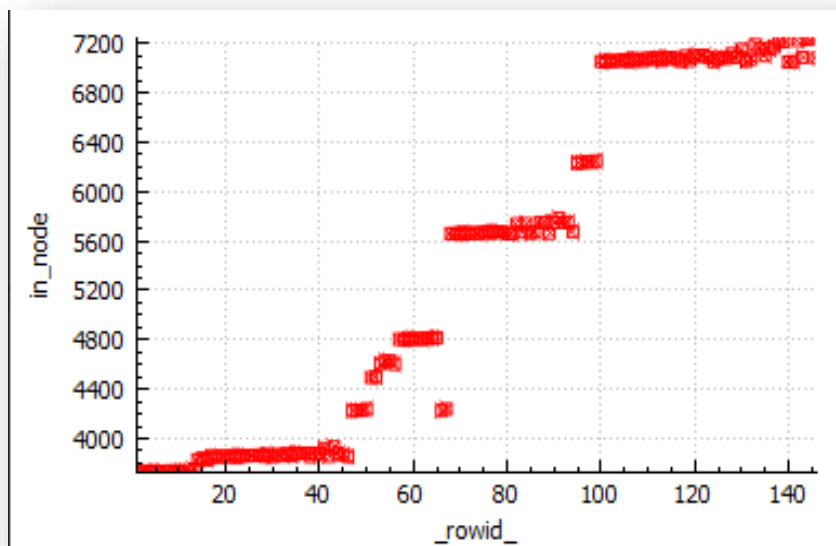
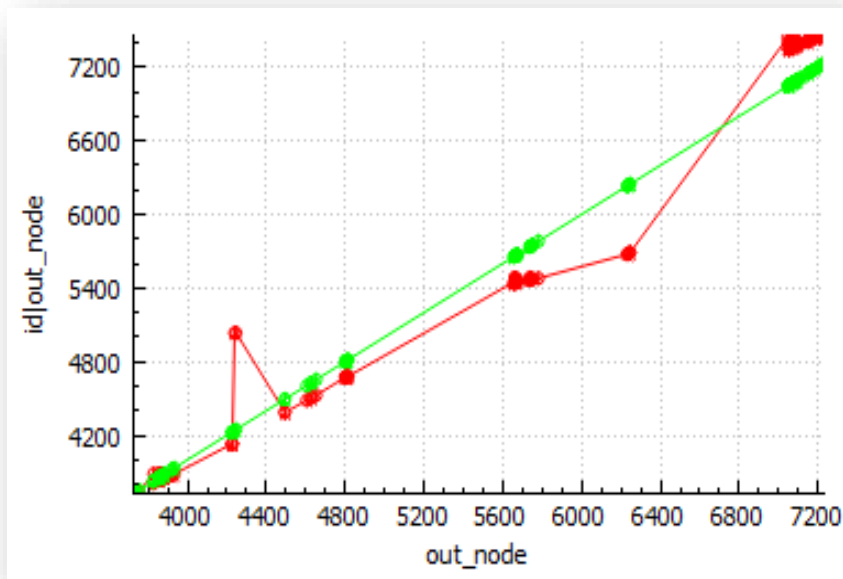


Table 12 Edelkamp-3



Τα παραπάνω είναι τα μόνα αποτελέσματα που εξήχθησαν, όσες μετατροπές και να έγιναν, είτε στα δεδομένα μας είτε στους αλγορίθμους που προσπαθήσαμε να χρησιμοποιήσουμε.

## 5.4 ΣΥΜΠΕΡΑΣΜΑΤΑ

Το βασικό πρόβλημα μεταξύ των δεδομένων AIS και των αλγορίθμων που χρησιμοποιούνται στην κατασκευή χαρτών οδικού δικτύου είναι η διαφορετικότητα των δεδομένων με `loc_id`. Μέσω αυτού και της συχνότητας των εκπομπών σήματος, η ανάλυσή τους με τους συγκεκριμένους αλγορίθμους γίνεται εύκολα και υπολογίζονται όλες οι παράμετροι που έχουν τεθεί. Τα AIS δεδομένα είναι μία ενδιαφέρουσα προοπτική, αφού σε σχέση με τα δεδομένα που προέρχονται από το GPS, παρέχουν μία γκάμα πληροφοριών. Ωστόσο, η ανάλυσή τους από τους επίγειους αλγορίθμους κατασκευής χαρτών παρουσιάζει πολλά προβλήματα. Ενδεικτικά, θα αναφέρουμε ότι με τη διαδικασία που ακολουθεί ο `tracebundle` (Karagiorgou) λαμβάνεται υπόψη η γωνία (στροφή) των κινητών αντικειμένων. Στα δεδομένα του αυτόματου συστήματος εντοπισμού, η μεγαλύτερη γωνία που σχηματίστηκε ήταν  $4^{\circ}$ . Αυτό είχε σαν αποτέλεσμα ο αλγόριθμος να τερματίζει, αφού όσες φορές και να επαναλάμβανε τη διαδικασία εντοπισμού σημείων που δημιουργούσαν συνθήκες στροφής, δεν έπαιρνε ποτέ κανένα αποτέλεσμα.

Κλείνοντας αυτή την προσπάθεια, θα λέγαμε ότι τα δεδομένα που λαμβάνονται από τα πλοία δεσχετίζονται ούτε ως δομή αλλά ούτε και στον τρόπο με τον οποίο συλλέγονται σε σχέση με τα δεδομένα των GPS. Η διαφορετικότητα των δύο συστημάτων εντοπισμού θέσης, καθώς και η μορφολογία των δεδομένων καθιστούν σχεδόν απαγορευτική την ανάλυσή τους από αλγορίθμους που σχετίζονται με την κατασκευή οδικών χαρτών.

## ΠΕΡΙΟΡΙΣΜΟΙ

Το βασικό πρόβλημα που αντιμετώπισα ήταν η έλλειψη των μοντέλων που χρησιμοποιούσαν οι αλγόριθμοι για να τρέξουν. Οι περισσότεροι αλγόριθμοι, αν και ελεύθεροι στο διαδίκτυο, χρησιμοποιούσαν μοντέλα μέσα στον αλγόριθμό τους τα οποία δε δίνονταν για εγκατάσταση στο αντίστοιχο πρόγραμμα, αλλά πολύ περισσότερο δεν κυκλοφορούσαν στο διαδίκτυο, ώστε να μπορέσω να τα εγκαταστήσω. Ο περιορισμός αυτός δε μου έδωσε τη δυνατότητα να μπορέσω να εξάγω τα αποτελέσματα τα οποία περίμενα.

Στο σημείο αυτό θα ήθελα να τόνισω τη βοήθεια της κ. Καραγιώργου, ώστε να μπορέσουμε να «τρέξουμε» τον δικό της αλγόριθμο με τα δεδομένα που είχαμε. Αυτό δεν κατέστη δυνατό, λόγω της ιδιομορφίας του αλγορίθμου και των δεδομένων, αλλά και μόνο η συμβολή της σε αυτή την προσπάθεια ήταν υπερπολύτιμη.

## ΠΑΡΑΡΤΗΜΑ

### 1 ΑΛΓΟΡΙΘΜΟΣ ΜΕΤΑΤΡΟΠΗΣ ΔΕΔΟΜΕΝΩΝ 1

```
import os
import io

path='./'
input_file='vessels_trajectories_1285.txt'
excelFiles = []

def create_trip_Files(id):
    print(id)
    file_name = "trip_"+id+".txt"
    print(file_name)
    if (not os.path.isfile(file_name)):
        print("creating file"+file_name)
        new_file=open(file_name,'w').close()

def populateFiles(id, line1, line2):
    file_name = "trip_"+id+".txt"
    print("Adding lines to file "+file_name)
    print(line1)
    print(line2)
    f = open(file_name,'a')
    f.write(line1+'\n') # python will convert \n to os.linesep
    f.write(line2+'\n')
    f.close()

def turn_date_to_number(dt):
    print(dt)
    date=dt.split(' ')[0]
    time=dt.split(' ')[1]
    print(date, time)
    year=int(date.split('-')[0])
    month=int(date.split('-')[1])
    day=int(date.split('-')[2])
    hour=int(time.split(':')[0])
    minute=int(time.split(':')[1])
    second=int(time.split(':')[2].split('.')[0])
    # print(year , month, day, hour, minute, second)
```

```
result = year*31556926 + month*2629743 + day*86400 + hour*3600 + minute*60
+ second
#t=(mat(:,3)*31556926+mat(:,4)*2629743.83+mat(:,5)*86400+mat(:,6)*3600+mat(:,7)
*60+mat(:,8))
return str(result)

def createNewData(file):
print(file);
fh = open(file,'r')
#get data from each line
line_num = 0
for line in fh:
line_num = line_num+1
# print(line)
elements = line.strip('\n').split(';')
print(elements)

#ship_ID
object_ID=elements[2]

max_length = len(elements)

#POP last item if it is empty
if(elements[max_length-1] == ""):
elements.pop(max_length-1)

new_length = len(elements)
#check if data is OK
if (new_length-5)%9 != 0:
print("Error in your data at line ",line_num)
else:
print("Data length is OK. Proceeding with reading data...")

#creating files for each Object_ID if file not exists
create_trip_Files(object_ID)

#populate file
for iin range (5,new_length,9):
# print("i=" , i)
dim_x1 = elements[i+3]
dim_y1 = elements[i+4]
dim_t1 = elements[i+5]
```

```
dim_x2 = elements[i+6]
dim_y2 = elements[i+7]
dim_t2 = elements[i+8]
line1= dim_x1+' '+dim_y1+' '+turn_date_to_number(dim_t1)
line2= dim_x2+' '+dim_y2+' '+turn_date_to_number(dim_t2)
print(line1)
populateFiles(object_ID, line1, line2)

#test date to number
#time1 = turn_date_to_number(dim_t1)
#print(time1)

for file in os.listdir(path):
#look for .dat files in path
if os.path.isfile(path+file) & file.endswith('.dat'):
    excelFiles.append(os.path.join(path, file))

#read excel files
#for f in excelFiles:
# createNewData(f)
createNewData(input_file)
```

Ο παραπάνω αλγόριθμος μετατρέπει τα δεδομένα στη μορφή (X Y TIMESTAMP)



## 2 ΑΛΓΟΡΙΘΜΟΣ ΜΕΤΑΤΡΟΠΗΣ ΔΕΔΟΜΕΝΩΝ 2

```
import os
import io

path='./'
input_file='vessels_trajectories_1285.txt'
excelFiles = []

def create_trip_Files(id):
    print(id)
    file_name = "trip_"+id+".txt"
    new_file=open(file_name,'w').close()

def populateFiles(id, line1, line2):
    file_name = "trip_"+id+".txt"
    print("Adding lines to file "+file_name)
    print(line1)
    print(line2)
    f = open(file_name,'a')
    # f.write(line1+'\n') # python will convert \n to os.linesep
    f.write(line2+'\n')
    f.close()

def turn_date_to_number(dt):
    print(dt)
    date=dt.split(' ')[0]
    time=dt.split(' ')[1]
    print(date, time)
    year=int(date.split('-')[0])
    month=int(date.split('-')[1])
    day=int(date.split('-')[2])
    hour=int(time.split(':')[0])
    minute=int(time.split(':')[1])
    second=int(time.split(':')[2].split('.')[0])
    # print(year , month, day, hour, minute, second)
    result = year*31556926 + month*2629743 + day*86400 + hour*3600 + minute*60
    + second
    #t=(mat(:,3)*31556926+mat(:,4)*2629743.83+mat(:,5)*86400+mat(:,6)*3600+mat(:,7)
    *60+mat(:,8))
    return str(result)
```

```
def createNewData(file):
print(file);
    fh = open(file,'r')
#get data from each line
    line_num = 0
for line in fh:
    line_num = line_num+1
# print(line)
    elements = line.strip('\n').split(';')
print(elements)

#ship_ID
    object_ID=elements[2]

    max_length = len(elements)

#POP last item if it is empty
if(elements[max_length-1] == ""):
    elements.pop(max_length-1)

    new_length = len(elements)
#check if data is OK
if (new_length-5)%9 != 0:
print("Error in your data at line ",line_num)
else:
print("Data length is OK. Proceeding with reading data...")

#creating files for each Object_ID if file not exists
    create_trip_Files(object_ID)

    trip_id =1
#populate file
for iin range (5,new_length,9):
# print("i=" , i)
    current_id = trip_id
    previous_id = current_id -1
    next_id = current_id + 1
if previous_id == 0:
    previous_id = 'None'
print((new_length-5)/9 , next_id)
if next_id > (new_length-5)/9:
    next_id = 'None'
```

```
dim_x1 = elements[i+3]
dim_y1 = elements[i+4]
dim_t1 = elements[i+5]
dim_x2 = elements[i+6]
dim_y2 = elements[i+7]
dim_t2 = elements[i+8]
line1=
str(current_id)+','+dim_x1+','+dim_y1+','+turn_date_to_number(dim_t1)+','+str(previous_id)+','+str(next_id)
line2=
str(current_id)+','+dim_x2+','+dim_y2+','+turn_date_to_number(dim_t2)+','+str(previous_id)+','+str(next_id)
print(line2)
populateFiles(object_ID, line1, line2)
trip_id = trip_id+1
#test date to number
#time1 = turn_date_to_number(dim_t1)
#print(time1)

for file in os.listdir(path):
#look for .dat files in path
if os.path.isfile(path+file) & file.endswith('.dat'):
    excelFiles.append(os.path.join(path, file))

#read excel files
#for f in excelFiles:
# createNewData(f)
createNewData(input_file)
```

Στον παραπάνω αλγόριθμο, εκτός από τα X, Y και το χρόνο (TIMESTAMP), που κρατάμε από τα αρχικά δεδομένα λόγω έλλειψης στοιχείων δημιουργούμε location\_id.

### 3 ΑΛΓΟΡΙΘΜΟΣ ΜΕΤΑΤΡΟΠΗΣ ΔΕΔΟΜΕΝΩΝ-3

```
import os
import io

path='./'
input_file='vessels_trajectories_1285.txt'
excelFiles = []
```

```
def create_trip_Files(id):
    print(id)
    file_name = "trip_" + id + ".txt"
    print(file_name)
    if (not os.path.isfile(file_name)):
        print("creating file" + file_name)
        new_file = open(file_name, 'w').close()

def populateFiles(id, line1, line2):
    file_name = "trip_" + id + ".txt"
    print("Adding lines to file " + file_name)
    print(line1)
    print(line2)
    f = open(file_name, 'a')
    # f.write(line1 + '\n') # python will convert \n to os.linesep
    f.write(line2 + '\n')
    f.close()

def turn_date_to_number(dt):
    print(dt)
    date = dt.split(' ')[0]
    time = dt.split(' ')[1]
    print(date, time)
    year = int(date.split('-')[0])
    month = int(date.split('-')[1])
    day = int(date.split('-')[2])
    hour = int(time.split(':')[0])
    minute = int(time.split(':')[1])
    second = int(time.split(':')[2].split('.')[0])
    # print(year, month, day, hour, minute, second)
    result = year * 31556926 + month * 2629743 + day * 86400 + hour * 3600 + minute * 60
    + second
    # t = (mat(:,3) * 31556926 + mat(:,4) * 2629743.83 + mat(:,5) * 86400 + mat(:,6) * 3600 + mat(:,7)
    * 60 + mat(:,8))
    return str(result)

def createNewData(file):
    print(file)
    fh = open(file, 'r')
    # get data from each line
    line_num = 0
```

```
for line in fh:
    line_num = line_num+1
# print(line)
    elements = line.strip('\n').split(';')
print(elements)

#ship_ID
    object_ID=elements[2]

    max_length = len(elements)

#POP last item if it is empty
if(elements[max_length-1] == ""):
    elements.pop(max_length-1)

    new_length = len(elements)
#check if data is OK
if (new_length-5)%9 != 0:
    print("Error in your data at line ",line_num)
else:
    print("Data length is OK. Proceeding with reading data...")

#creating files for each Object_ID if file not exists
    create_trip_Files(object_ID)

#populate file
for i in range (5,new_length,9):
    # print("i=" , i)
        dim_x1 = elements[i+3]
        dim_y1 = elements[i+4]
        dim_t1 = elements[i+5]
        dim_x2 = elements[i+6]
        dim_y2 = elements[i+7]
        dim_t2 = elements[i+8]
        line1= dim_x1+' '+dim_y1+' '+turn_date_to_number(dim_t1)
        line2= dim_x2+' '+dim_y2+' '+turn_date_to_number(dim_t2)
    print(line1)
    populateFiles(object_ID, line1, line2)

#test date to number
#time1 = turn_date_to_number(dim_t1)
#print(time1)
```

```
for file in os.listdir(path):  
    #look for .dat files in path  
    if os.path.isfile(path+file) & file.endswith('.dat'):  
        excelFiles.append(os.path.join(path, file))  
  
#read excel files  
#for f in excelFiles:  
#    createNewData(f)  
createNewData(input_file)
```

## ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

### ΞΕΝΕΣ ΑΝΑΦΟΡΕΣ

[1] Laurent ETIENNE, Thomas DEVOGELE and Alain BOUJU [ SPATIO-TEMPORAL TRAJECTORY ANALYSIS OF MOBILE OBJECTS FOLLOWING THE SAME ITINERARY] [The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38, Part II] (2010), pp 86-91

[2] Katerina TZAVELLA, Martin ULMKE [Semi-automatic sea lane extraction combining Particle Filtering (PF) and Geographic Information Systems (GIS)] (2014), Vogler, R., Car, A., Strobl, J. & Griesebner, G. (Eds.) (2014): GI\_Forum 2014. Geospatial Innovation for Society. © Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach. ISBN 978-3-87907-545-4. © ÖAW Verlag, Wien. eISBN 978-3-7001-7652-7, doi:10.1553/giscience2014s2. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

[3] Giuliana Pallotta, Michele Vespe, and Karna Bryan [Traffic Route Extraction and Anomaly Detection from AIS Data] (2013), This work relates to Department of the Navy Grant N62909-11-1-7040 issued by Office of Naval Research Global, pp1-4

[4] Published in Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96) Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise (DBSCAN) pp 226-229

[5] KC Li [High dimensional data analysis via the SIR/PHD approach] (2000), pp 15-16

[6] Mahmuda Ahmed · Sophia Karagiorgou · Dieter Pfoser · Carola Wenk (A Comparison and Evaluation of Map Construction Algorithms Using Vehicle Tracking Data) (2014 first online), GeoInformatica July 2015, Volume 19, Issue 3, pp 601–632

[7] Lili Cao , John Krumm (From GPS Traces to a Routable Road Map ) (2009),

GIS '09 Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems,  
pp 3-12

[8] James Biagioni, Jakob Eriksson (Map Inference in the Face of Noise and Disparity)(2012),

SIGSPATIAL '12 Proceedings of the 20th International Conference on Advances in Geographic Information Systems,  
pp 79-88

[9] Sophia Karagiorgou, Dieter Pfoser, Dimitrios Skoutas (Segmentation-Based Road Network Construction ) (2013),

SIGSPATIAL'13 Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems  
pp 470-473

[10] Jonathan J. Davies, Alastair R. Beresford, and Andy Hopper (Scalable, Distributed, Real-Time Map Generation )IEEE Pervasive Computing (Volume:5 , Issue: 4 )(2006) pp 47-54

[11] Sophia Karagiorgou, Dieter Pfoser (On Vehicle Tracking Data-Based Road Network Generation) (2012),

SIGSPATIAL '12 Proceedings of the 20th International Conference on Advances in Geographic Information Systems  
pp 89-98

[12] Mahmuda Ahmed, Carola Wenk (Construction Street-Maps from GPS Trajectories) (2012), In Proc. European Symp. on Algorithms, pp 60–71, 2012.

[13] STEFAN SCHROEDL, KIRI WAGSTAFF, SETH ROGERS, PAT LANGLEY, CHRISTOPHER WILSON (Mining GPS Traces for Map Refinement), Data Mining and Knowledge Discovery July 2004, Volume 9, Issue 1, pp 59-87(2004)



ΑΝΑΦΟΡΕΣ ΣΤΟ ΔΙΑΔΙΚΤΥΟ

[www.mapconstruction.org](http://www.mapconstruction.org)

[https://en.wikipedia.org/wiki/Automatic\\_identification\\_system](https://en.wikipedia.org/wiki/Automatic_identification_system)