

Εφαρμογές σε Χωρικά Δίκτυα

Ελευθέριος Τιάκας

Νοέμβριος 2016

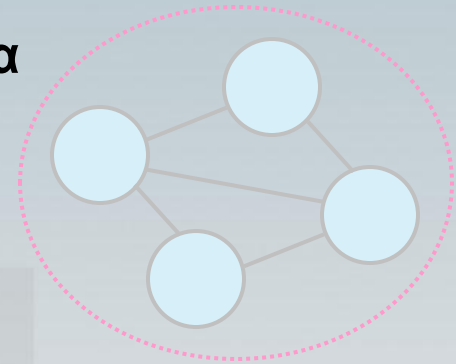


**Αριστοτέλειο
Πανεπιστήμιο
Θεσσαλονίκης**

**Εργαστήριο Τεχνολογίας και
Επεξεργασίας Δεδομένων
Τμήμα Πληροφορικής Α.Π.Θ.**

Σύνοψη Παρουσίασης

- **Σημαντικά Κίνητρα και Εφαρμογές σε Χωρικά Δίκτυα**
- **Βασικά Ερωτήματα και Προβλήματα**
- **Βασική Μεθοδολογία**
- **Ερωτήματα Ομοιότητας Τροχιών σε Χωρικά Δίκτυα**
- **Ερωτήματα Ομοιότητας Τροχιών με Τοποθεσίες σε Χωρικά Δίκτυα**
- **Εκτίμηση της Επιλεξιμότητας Ερωτημάτων σε Χωρικά Δίκτυα**
- **Ομαδοποίηση Κόμβων Γράφου με Μεταβατική Ομοιότητα**
- **Ιδέες για Μελλοντική Έρευνα**



Σημαντικά Κίνητρα και Εφαρμογές

- Συστήματα Πλοήγησης - Navigation Systems (map destination, shortest routes, route planning, location-based services, etc.)



map destination



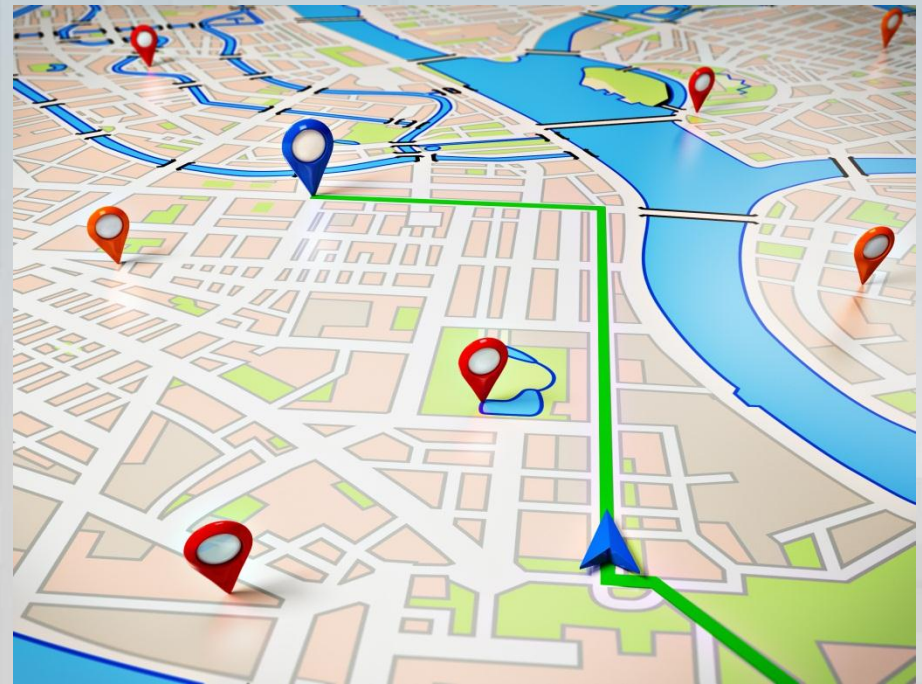
shortest routes

Σημαντικά Κίνητρα και Εφαρμογές

- Συστήματα Πλοήγησης - Navigation Systems (map destination, shortest routes, route planning, location-based services, etc.)



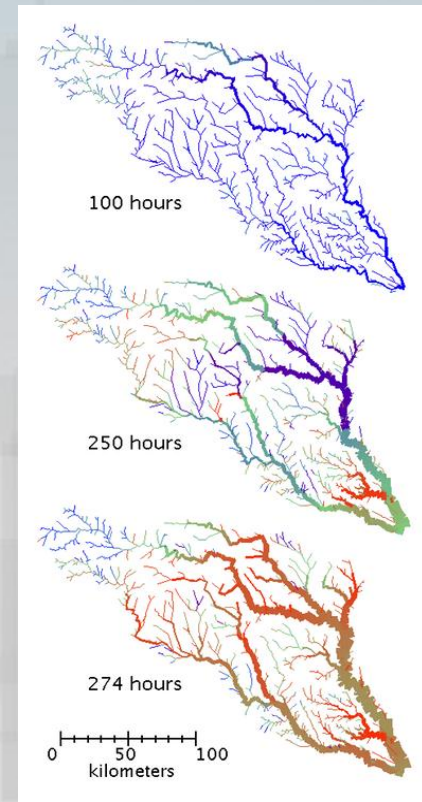
route planning



location-based services

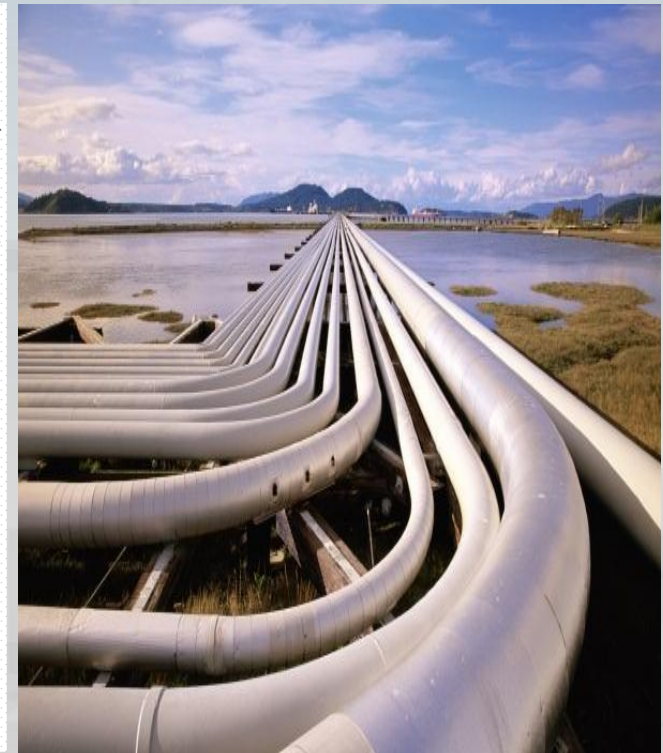
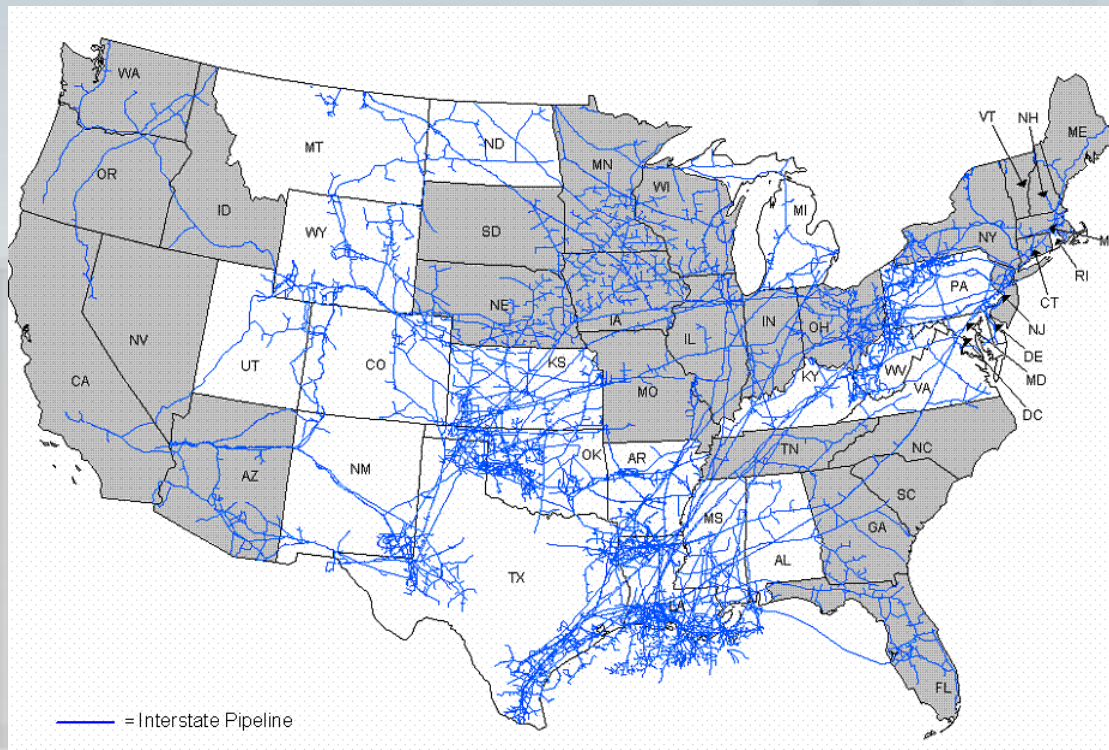
Σημαντικά Κίνητρα και Εφαρμογές

- Συστήματα Μεταφοράς - Transportation Systems (road, river, railway networks, traffic analysis, etc.)



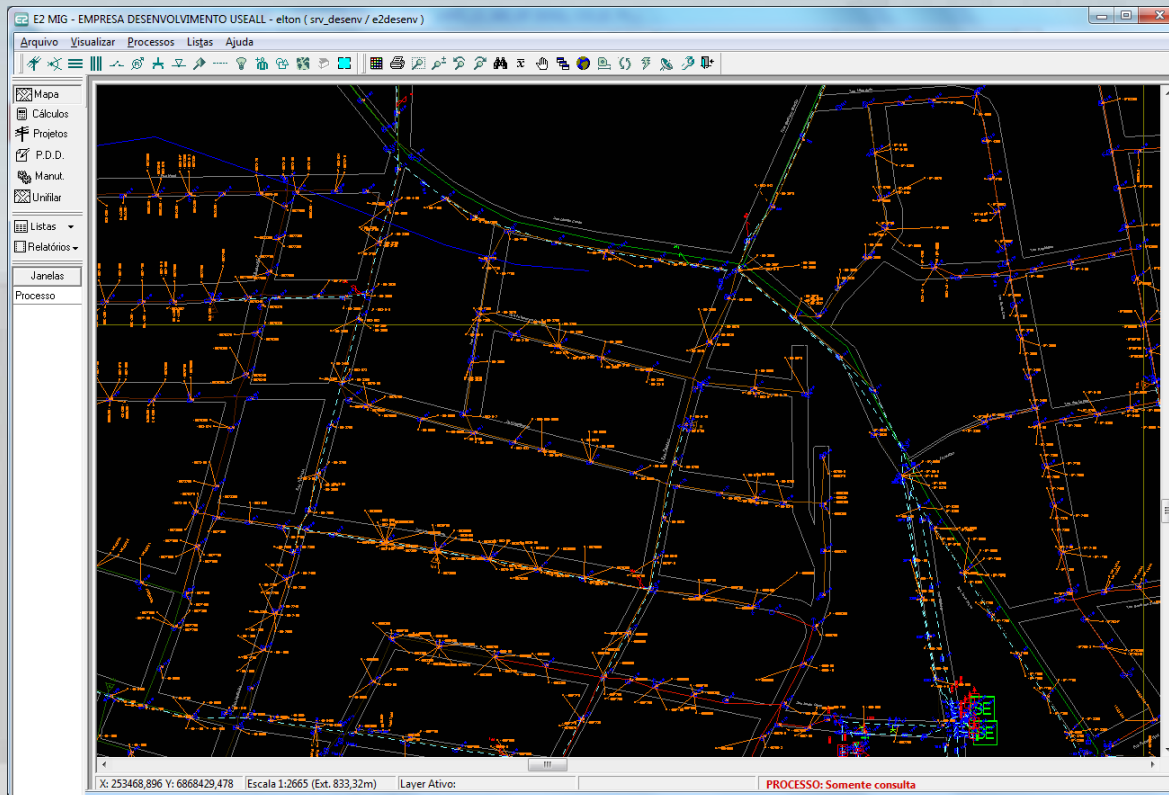
Σημαντικά Κίνητρα και Εφαρμογές

- Συστήματα Υποδομών (oil, electrical power network, natural gas, telephone lines, water-sewer systems etc.)



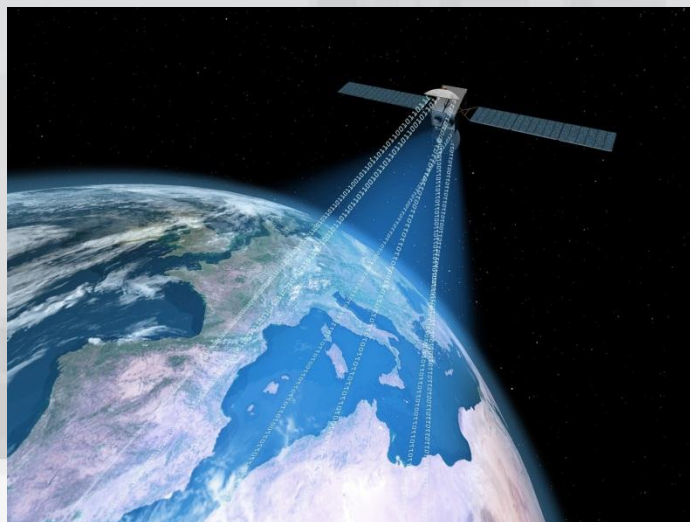
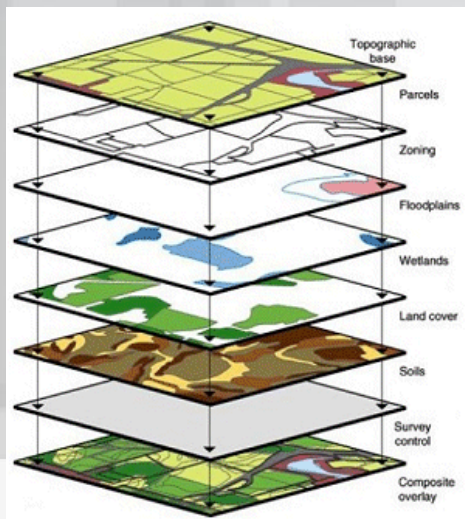
Σημαντικά Κίνητρα και Εφαρμογές

- Συστήματα Υποδομών (oil, electrical power network, natural gas, telephone lines, water-sewer systems etc.)



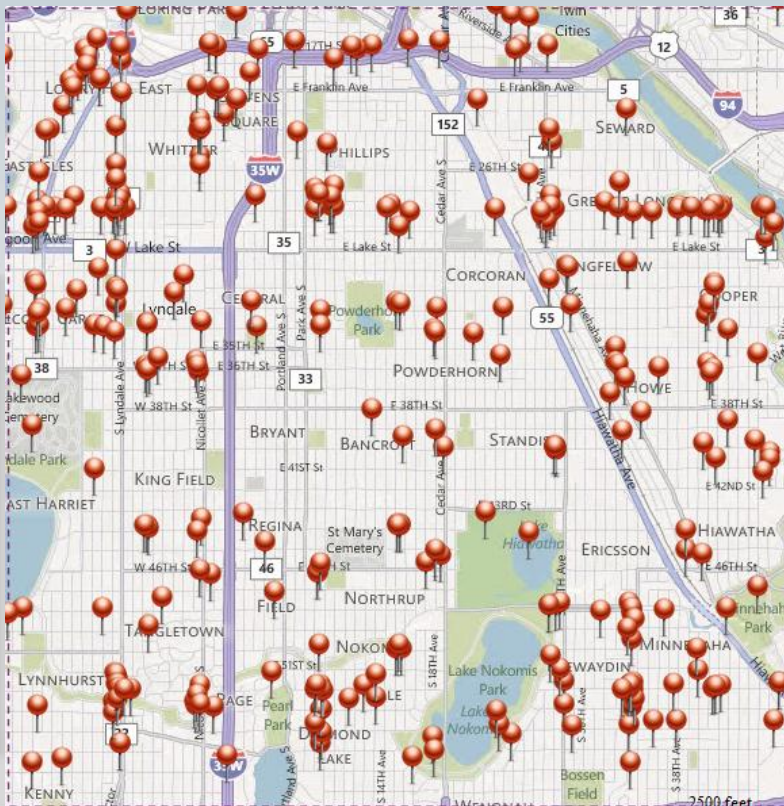
Σημαντικά Κίνητρα και Εφαρμογές

- Πολλοί διαφορετικοί Χάρτες και Συστήματα GIS.
- Μεγάλες ποσότητες δεδομένων (GPS, Sensors, δορυφόροι) για ροές (flows) και τροχιές (trajectories).
- Μεγάλη ζήτηση αγοράς αυτόματων συστημάτων και συστημάτων ελέγχου από εταιρείες, από κυβερνητικούς οργανισμούς αλλά και από ιδιώτες.
- Εκτενής έρευνα σε ιστορικά στοιχεία τέτοιων βάσεων δεδομένων.



Βασικά Ερωτήματα σε Χωρικά Δίκτυα

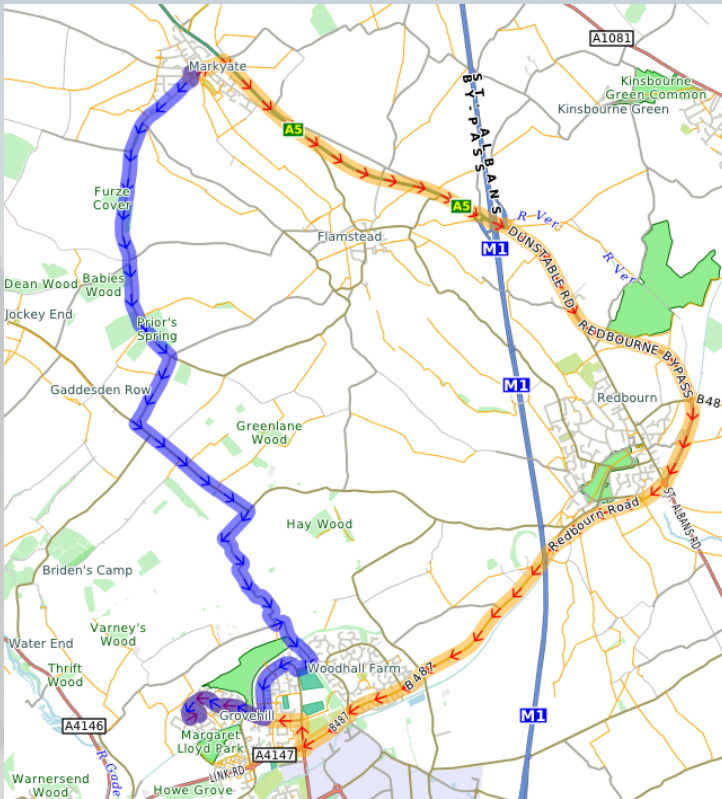
- Εύρεση βέλτιστης ομάδας χωρικών αντικειμένων με κριτήρια ως προς τον χώρο (απόσταση) ή τον χρόνο ή ακόμα και τον συνδυασμό χώρου-χρόνου:



- π.χ. τα κοντινότερα βενζινάδικα σε ακτίνα 5km (range), τα κοντινότερα ξενοδοχεία στα αξιοθέατα μιας πόλης (collocation) κ.α.
- Τα κριτήρια συνήθως αφορούν ελαχιστοποίηση των αποστάσεων (shortest-paths)
- Η επεξεργασία των ερωτημάτων στηρίζεται σε τεχνικές spatial data mining.

Βασικά Ερωτήματα σε Χωρικά Δίκτυα

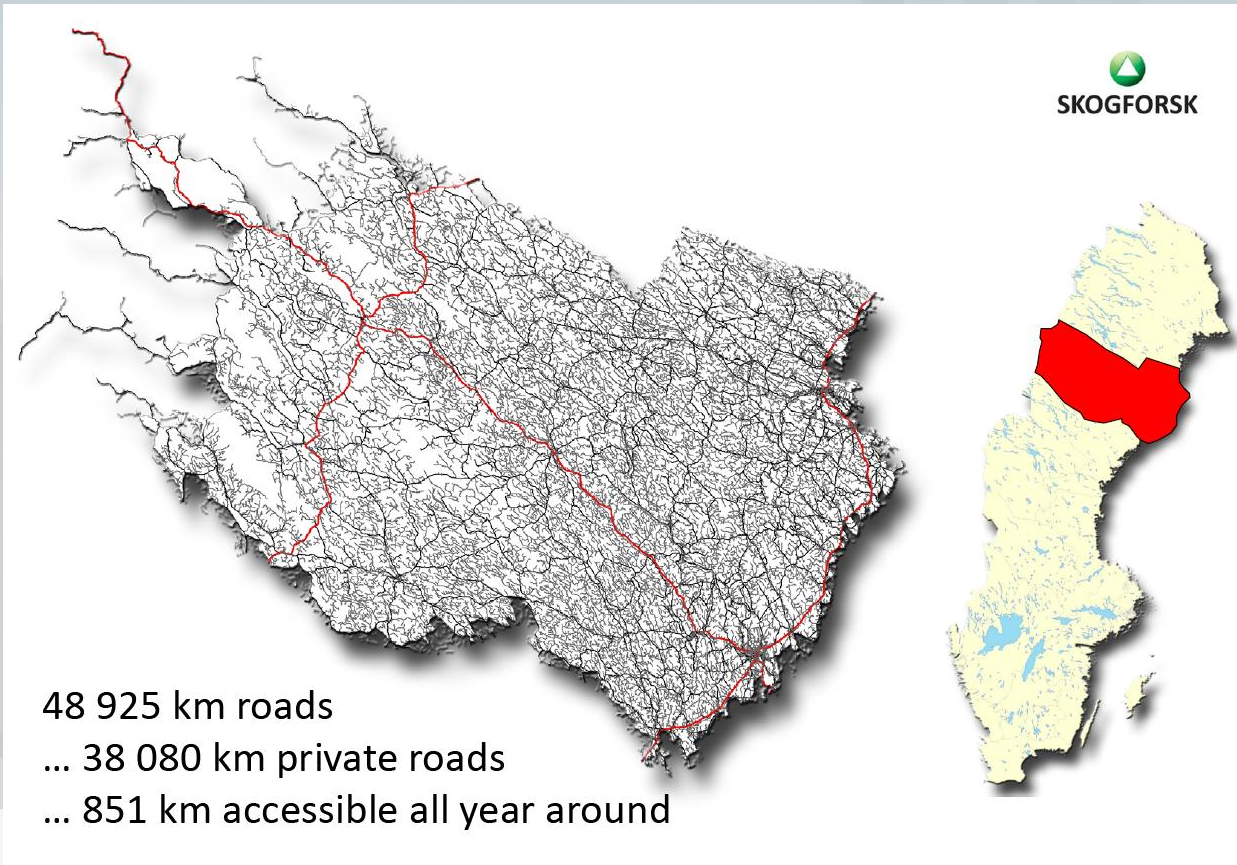
- Εύρεση βέλτιστης διαδρομής σε μορφή τροχιάς (trajectory) ως προς τον χώρο (απόσταση) ή τον χρόνο ή ακόμα και τον συνδυασμό χώρου-χρόνου:



- Η ελάχιστη διαδρομή σε απόσταση είναι η μπλε αλλά η ταχύτερη είναι η κίτρινη.
- Οι τροχιάς δημιουργούνται από την περιοδική καταγραφή της θέσης μέσω GPS ή άλλων συσκευών.
- Δημιουργούνται πολύ μεγάλες βάσεις δεδομένων ακόμα και σε μικρά δίκτυα (>1T).
- Χρειάζονται ευέλικτες δομές indexing για τη διαχείρισή τους.

Βασικά Ερωτήματα σε Χωρικά Δίκτυα

- Εύρεση βέλτιστης διαδρομής σε μορφή τροχιάς (trajectory) όταν εμφανίζονται αστάθμητοι παράγοντες και αβεβαιότητα (ερωτήματα ακόμα και πάνω σε δυναμικά χωρικά δίκτυα):



Βασικά Ερωτήματα σε Χωρικά Δίκτυα

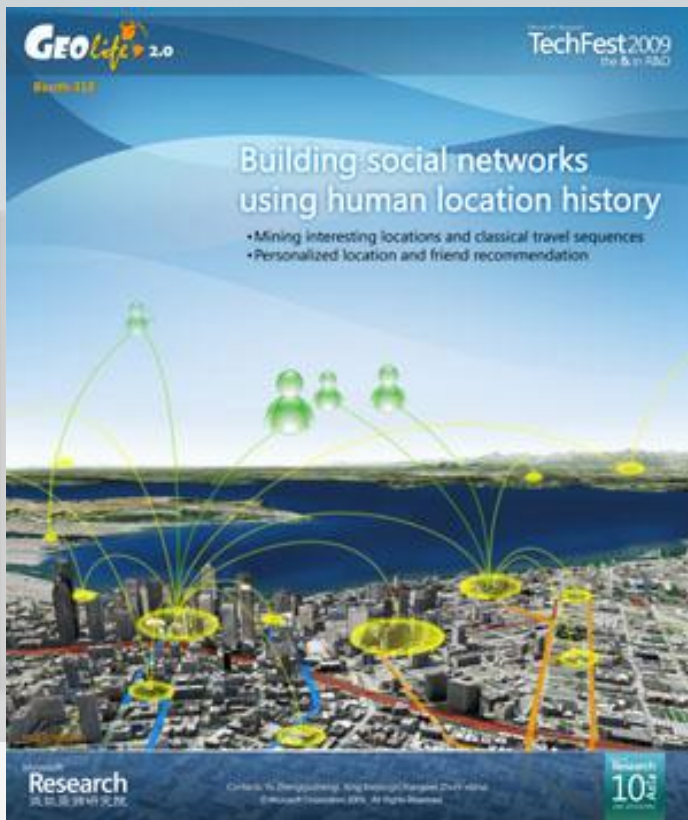
- Εύρεση βέλτιστης διαδρομής σε μορφή τροχιάς (trajectory) η οποία να διέρχεται από καθορισμένα σημεία ενδιαφέροντος (π.χ. αξιοθέατα, σημεία παράδοσης προϊόντων ή εμπορευμάτων κλπ.)



- Τελείως διαφορετικό πρόβλημα.
- Τα σημεία ενδιαφέροντος καθορίζουν επιπλέον περιορισμούς.
- Ακόμα πιο σύνθετο το ερώτημα αν παίζει ρόλο και η προτεραιότητα επίσκεψης.
- Χρειάζονται ειδικές δομές και αλγόριθμοι για την επεξεργασία τους.

Βασικά Ερωτήματα σε Χωρικά Δίκτυα

- Εύρεση βέλτιστης διαδρομής σε μορφή τροχιάς (trajectory) η οποία να χρησιμοποιεί καταγεγραμμένες διαδρομές (ιστορικό) άλλων (travel recommendation, sharing life experiences & friend recommendation, κλπ.)

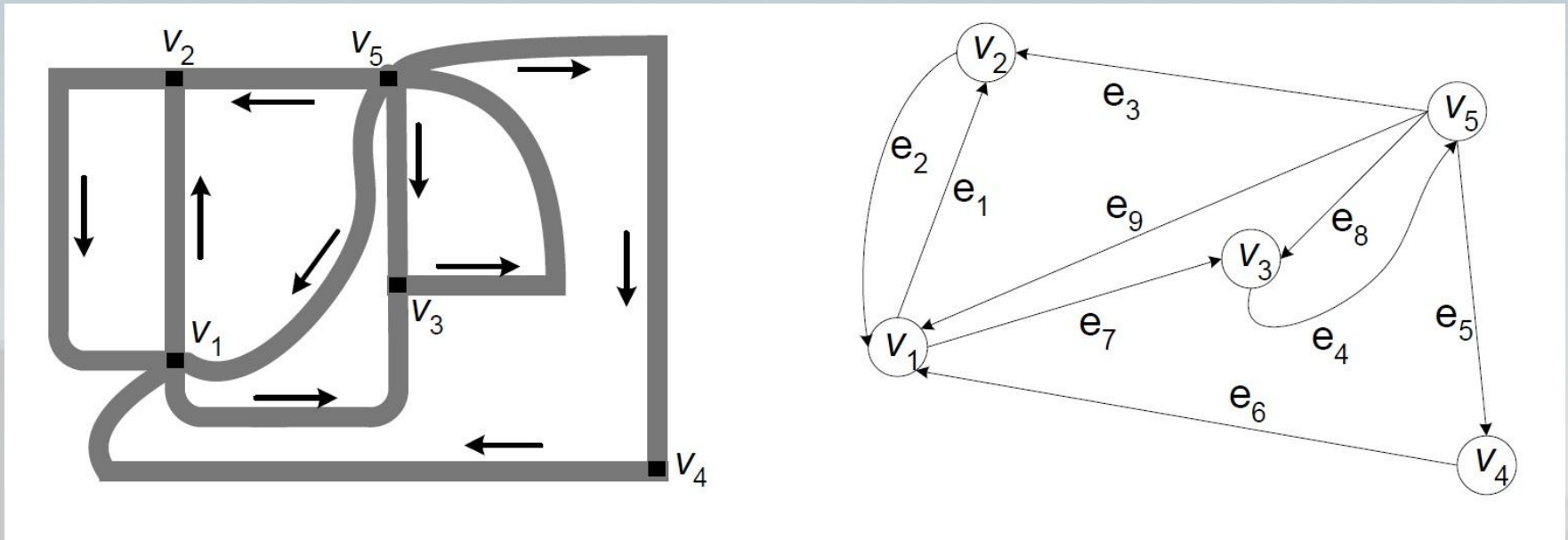


- Συνδυασμός Χωρικών και Κοινωνικών Δικτύων (Social Networks).



Βασική Μεθοδολογία

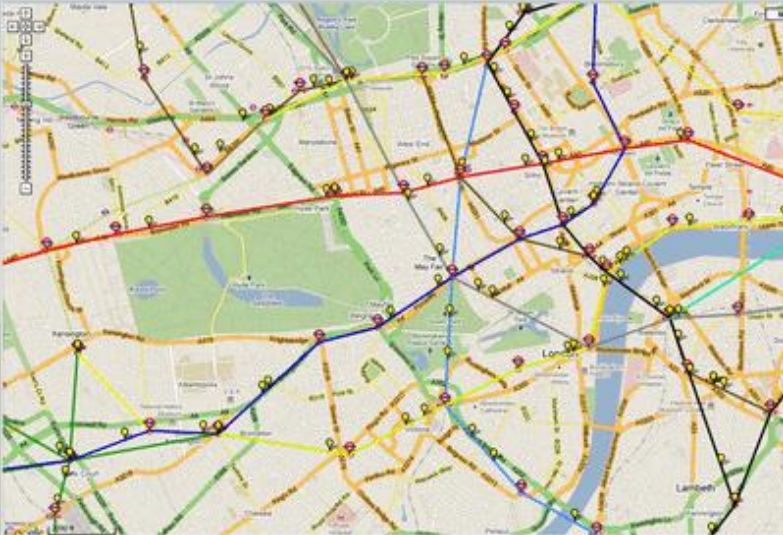
- Μοντελοποίηση του χωρικού δικτύου ως γράφου:



- Δημιουργία δομών για κόμβους, ακμές, βάρη, μονοπάτια, αποστάσεις (συνήθως **adjacency lists** και **distance matrices**).
- Υπολογισμός ελαχίστων αποστάσεων (**shortest path distances**), συνήθως με την μέθοδο **Dijkstra**.

Βασική Μεθοδολογία

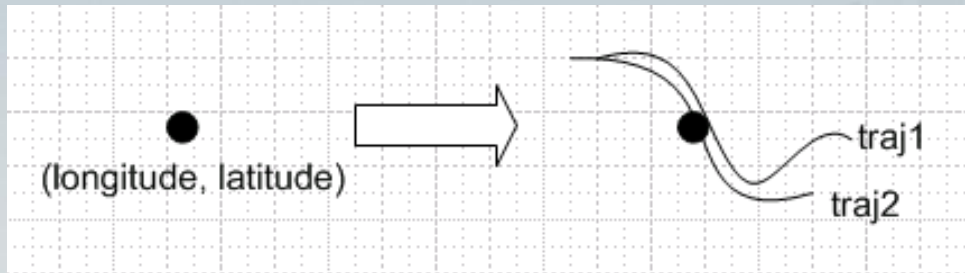
- Μοντελοποίηση των χωρικών αντικειμένων (σημεία ενδιαφέροντος, ομάδες, κατηγορίες κλπ.):



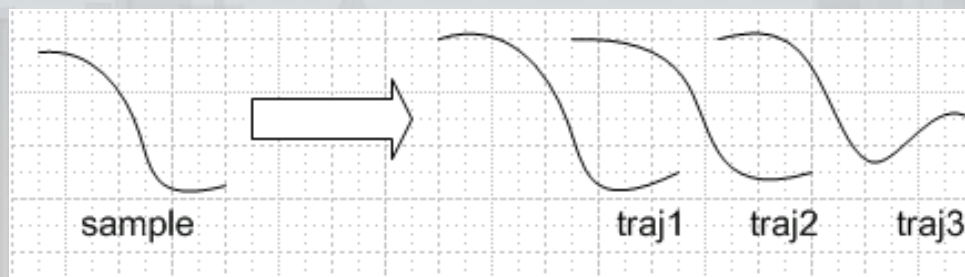
- Τα αντικείμενα μπορεί να μην βρίσκονται στους κόμβους, αλλά πάνω στις ακμές.
- Τα αντικείμενα μπορεί να κινούνται και να διαγράφουν τροχιές.
- Αν τα αντικείμενα είναι σταθερά χρησιμοποιούνται τεχνικές και αλγόριθμοι **data mining**.
- Χρησιμοποιούνται ειδικές δομές για την δεικτοδότησή τους (συνήθως τα **R-Trees** και οι παραλλαγές τους).

Βασική Μεθοδολογία

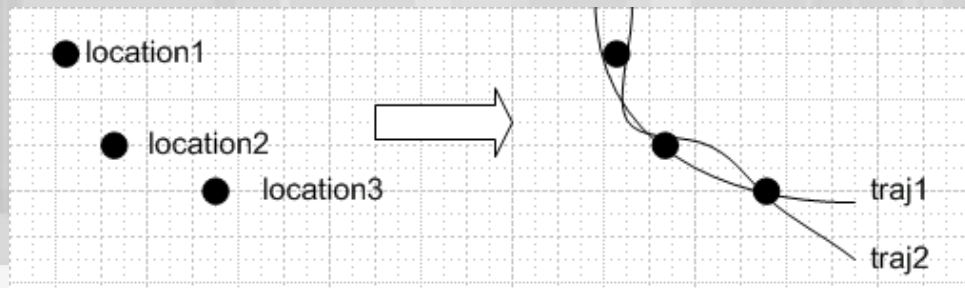
- Αναζήτηση **ομοιότητας** μεταξύ τροχιών ή μεταξύ σημείων ενδιαφέροντος και τροχιών με βάση καθορισμένα μέτρα ομοιότητας.



Frentzos et al. Geoinformatica-2007; Dfoser et al. VLDB-2000. (R-tree variants)

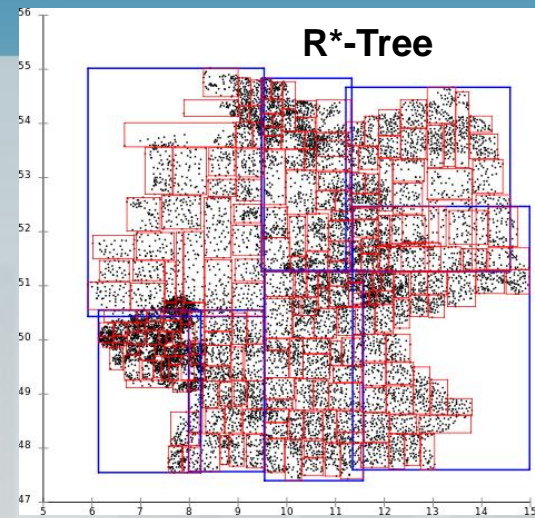
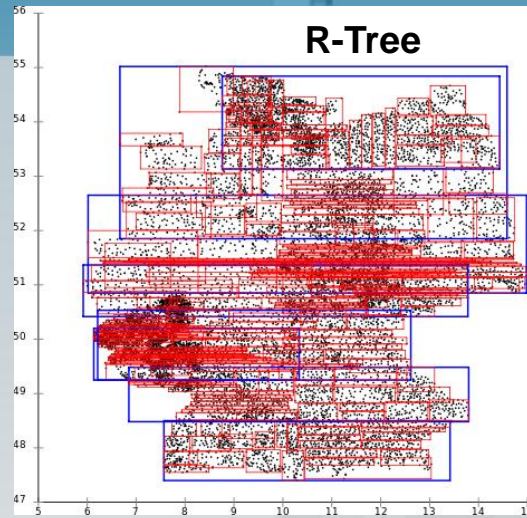
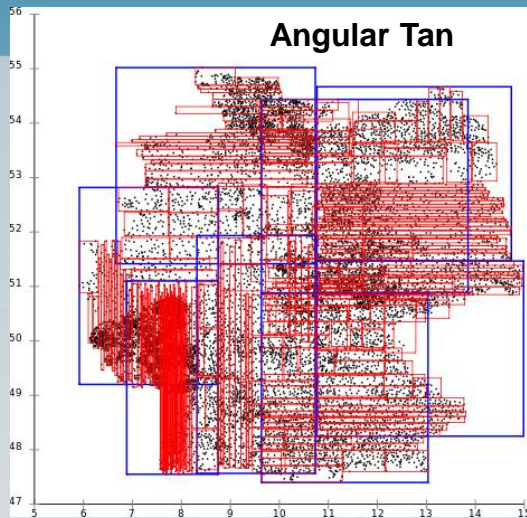


Chen et al, SIGMOD-2005; Vlachos et al, ICDE-2002; Yi et al, ICDE-1998, etc. (Similarity)



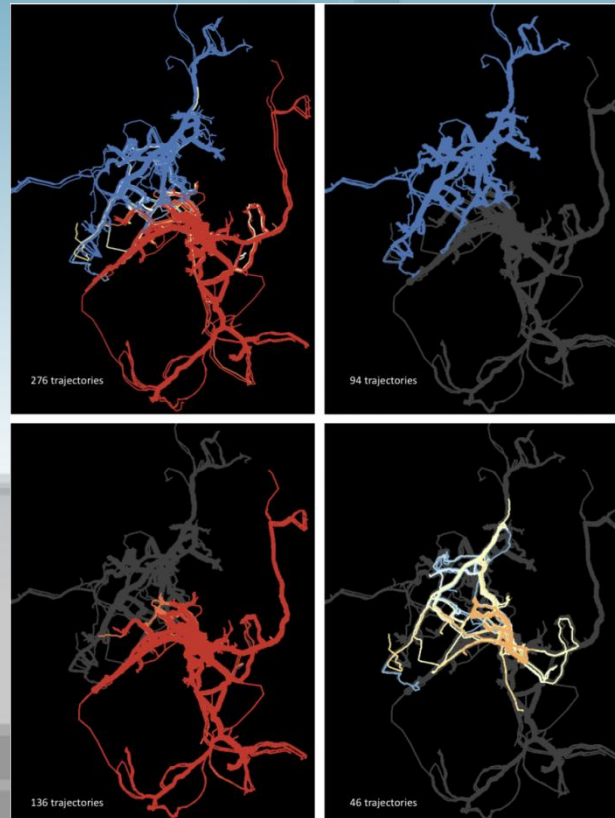
Chen et al, SIGMOD-2010; Wang et al, SIGMOD-2012; Shang et al, VLDB Journal 2013

Βασική Μεθοδολογία



- Πολλές φορές οι υπάρχουσες δομές (π.χ. **R-Trees**) δεν είναι αποδοτικές για την περίπτωση χωρικών δικτύων, αλλά μόνο για **Ευκλείδειους χώρους**.
- Αναγκαία η δημιουργία ευέλικτων δομών δεικτοδότησης (**indexing**) των αντικειμένων και των τροχιών τους.
- Δημιουργία αποδοτικών αλγορίθμων για την επεξεργασία των ερωτημάτων.
- Ελαχιστοποίηση του συνολικού κόστους επεξεργασίας (**CPU, I/O**).

Ερωτήματα Ομοιότητας Τροχιών σε Χωρικά Δίκτυα



Στην ομοιότητα τροχιών έγινε η μεγαλύτερη έρευνα τις τελευταίες δύο δεκαετίες.

Ερωτήματα Ομοιότητας Τροχιών σε Χωρικά Δίκτυα

Έρευνα και Βιβλιογραφία

- Μέθοδοι που αντιμετωπίζουν τις τροχιές ως **χρονοσειρές** ή **ακολουθίες** πραγματικών αριθμών:
 - Agrawal-1993, Faloutsos-1994, Yi-1998, Chan-1999, Keogh-2002, Chen-2004, Morse-2007, Sherkat-2008.
- Μέθοδοι που βασίζονται **μόνο στα χωρικά δεδομένα** των τροχιών:
 - Vlachos-2002, Chen-2005, Wang-2012, Lin-2005, Frentzos-2007.
- ❖ Κόστος υπολογισμών υψηλό, ειδικοί μηχανισμοί approximation για ελάττωσή του, η ομοιότητα αναζητείται κυρίως σε Ευκλείδειους χώρους (αρχικοί ή μετασχηματισμένοι χώροι) και γίνεται χρήση R-Trees και παραλλαγών τους.
- ❖ Χρήση ειδικών μέτρων ομοιότητας (π.χ. Dynamic Time Warping, Longest Common Sub-Sequence, Edit-Distance, Lp-norms) με υψηλό κόστος υπολογισμών, ομοιότητα σε Ευκλείδειους χώρους.

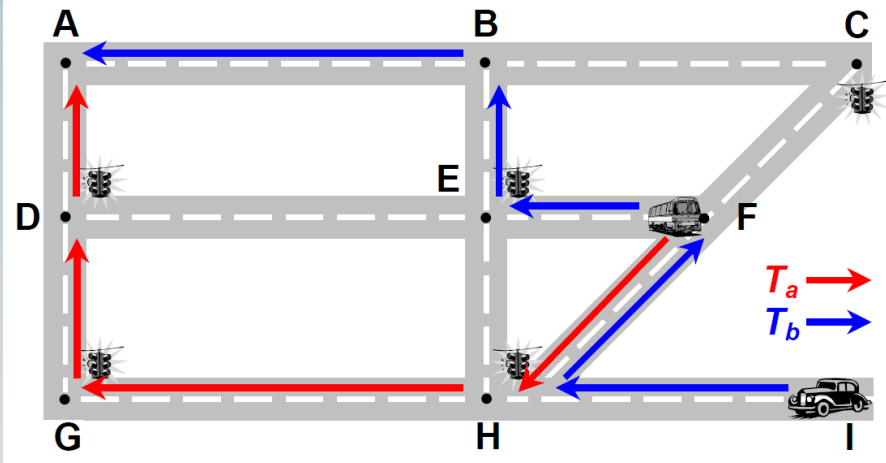
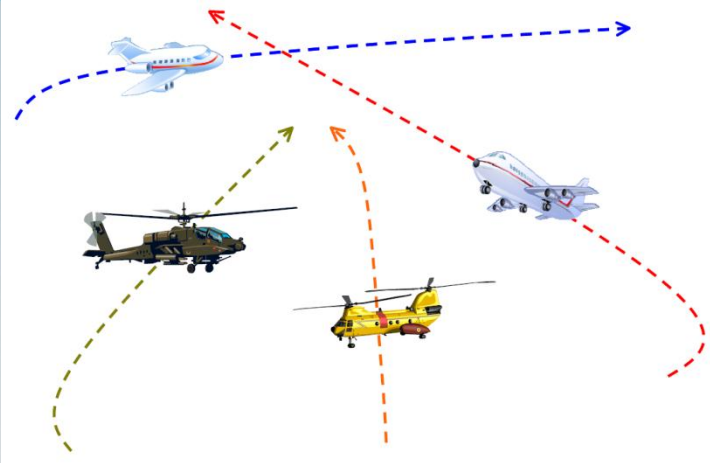
Ερωτήματα Ομοιότητας Τροχιών σε Χωρικά Δίκτυα

Έρευνα και Βιβλιογραφία

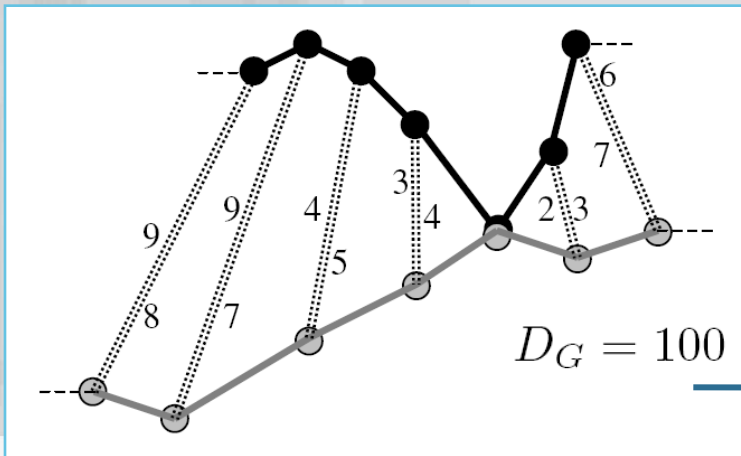
- Μέθοδοι που βασίζονται και σε χωρικά και σε χρονικά δεδομένα αλλά προβάλλονται σε Ευκλείδειους χώρους με παρόμοια προβλήματα:
 - Cai-2004 (Chebyshev-based distances), Frentzos-2007.
- Μέθοδοι που βασίζονται σε Χωρικά Δίκτυα:
 - Tiakas-2006, Tiakas-2009.
- ❖ Η ομοιότητα τροχιών δεν έχει μελετηθεί επαρκώς πάνω σε χωρικά δίκτυα, αλλά σε ελεύθερους (κυρίως ευκλείδειους) χώρους. Η πραγματικότητα όμως είναι διαφορετική.
- ❖ Οι περισσότερες εργασίες που προτάθηκαν για ομοιότητα τροχιών σε δίκτυα μελέτησαν το πρόβλημα δυαδικά, δηλαδή πότε δύο τροχιές είναι όμοιες ή ανόμοιες, χωρίς να εκφράζεται το πόσο όμοιες είναι. Η ποσοτική ομοιότητα τροχιών είναι σημαντική (cluster detection, data mining, outliers etc.).

Ερωτήματα Ομοιότητας Τροχιών σε Χωρικά Δίκτυα

Μέτρα Ομοιότητας



- Τροχιές κινούμενων αντικειμένων στον ελεύθερο ευκλείδειο χώρο και πάνω σε ένα χωρικό δίκτυο.



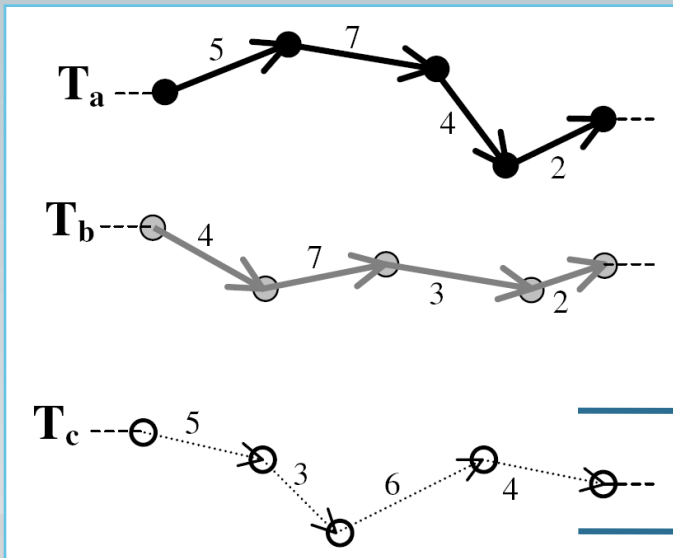
- Βασικό μέτρο ομοιότητας δύο τροχιών T_a , T_b με ίδιο μήκος περιγραφής m :

$$D_{net1}(T_a, T_b) = \frac{1}{m} \sum_{i=1}^m \frac{c(v_{ai}, v_{bi}) + c(v_{bi}, v_{ai})}{2D_G}$$

$$\frac{1}{7} \frac{(17+16+9+7+0+5+13)}{200} = \frac{67}{1400} = 0.047857$$

Ερωτήματα Ομοιότητας Τροχιών σε Χωρικά Δίκτυα

Μέτρα Ομοιότητας



- Βασικό μέτρο ομοιότητας δύο τροχιών T_a, T_b με ίδιο μήκος περιγραφής m στον χρόνο:

$$D_{time}(T_a, T_b) = \frac{1}{m-1} \sum_{i=1}^{m-1} \frac{|(T_a[i+1].t - T_a[i].t) - (T_b[i+1].t - T_b[i].t)|}{\max\{(T_a[i+1].t - T_a[i].t), (T_b[i+1].t - T_b[i].t)\}}$$

$$D_{time}(T_a, T_b) = \frac{1}{4} \left(\frac{1}{5} + \frac{0}{7} + \frac{1}{4} + \frac{0}{2} \right) = 0.1125$$

$$D_{time}(T_a, T_c) = \frac{1}{4} \left(\frac{0}{5} + \frac{4}{7} + \frac{2}{6} + \frac{2}{4} \right) = 0.35119$$

- Εναλλακτικό μέτρο ομοιότητας δύο τροχιών T_a, T_b με ίδιο μήκος περιγραφής m στον χώρο:

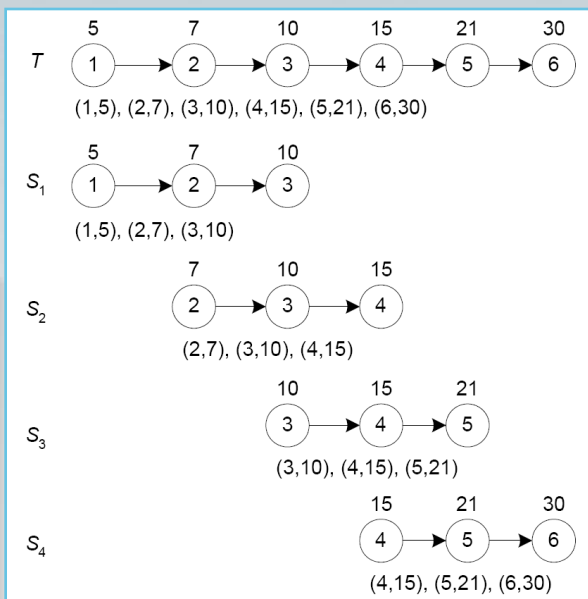
$$D_{net2}(T_a, T_b) = \frac{1}{m} \sum_{i=1}^m \frac{\sqrt{(x_{v_{ai}} - x_{v_{bi}})^2 + (y_{v_{ai}} - y_{v_{bi}})^2}}{D_G}$$

- Οι όμοιες τροχιές που επιστρέφονται με τη χρήση του μέτρου D_{net2} χρειάζονται έλεγχο και με το μέτρο D_{net1} (ύπαρξη false alarms).

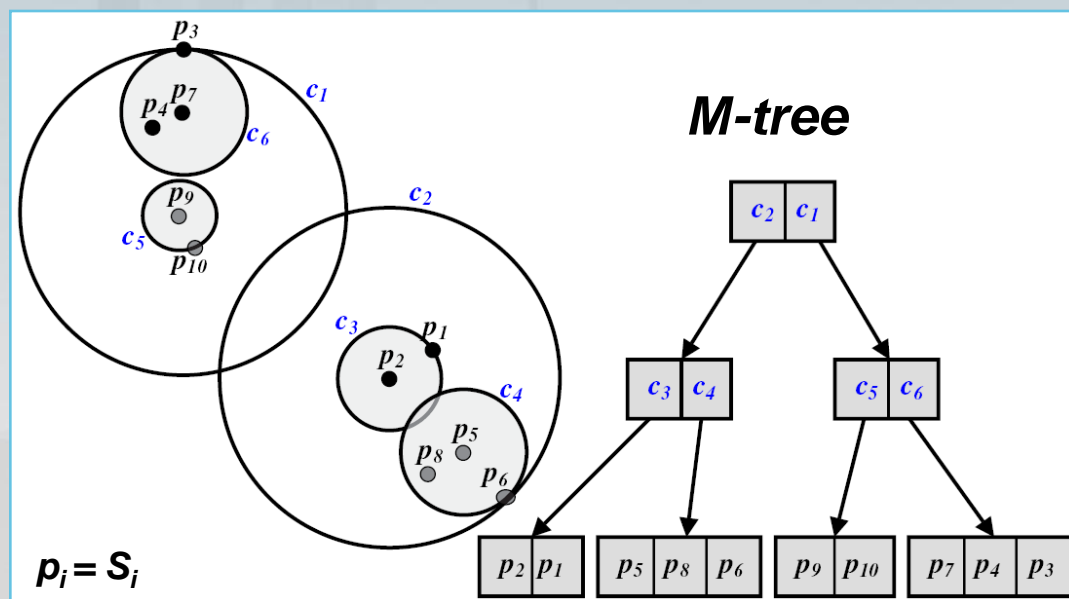
Ερωτήματα Ομοιότητας Τροχιών σε Χωρικά Δίκτυα

Αποσύνθεση – Δεικτοδότηση Τροχιών

- Αποσύνθεση τροχιών διαφορετικού μήκους σε τμήματα ίδιου μεγέθους μ :



- Δεικτοδότηση όλων των τμημάτων των τροχιών με **M-trees**:



- Χρήση ενός **M-tree** για κάθε διαφορετικό μέτρο ομοιότητας: (**NET-M-tree**, **TIME-M-tree**).
- Χρήση **LRU-Buffer** για τις υπολογισμένες αποστάσεις.

Ερωτήματα Ομοιότητας Τροχιών σε Χωρικά Δίκτυα

Επεξεργασία Ερωτημάτων Περιοχής

- Βασικές παράμετροι του ερωτήματος: (T_q) τροχιά αναφοράς, (E_{net}, E_{time}) ακτίνες αποστάσεων στο χώρο και στο χρόνο.
- Κάθε τροχιά T που θα βρίσκεται στα αποτελέσματα του ερωτήματος πρέπει να ικανοποιεί την εξής ιδιότητα:

$$\exists ts \subseteq T, \exists qs \subseteq T_q : D_{net}(ts, qs) \leq E_{net} \wedge D_{time}(ts, qs) \leq E_{time}$$

Μέθοδος Mtreell

- Για κάθε τμήμα τροχιάς qs της T_q εκτελείται ένα απλό ερώτημα περιοχής στα NET-M-Tree και TIME-M-Tree. Αποτελέσματα στα σύνολα C_{net} και C_{time} αντίστοιχα.
- $AS = C_{net} \cap C_{time}$ (id's τμημάτων).
- $AT =$ id's αντιστοιχών τροχιών από το AS .

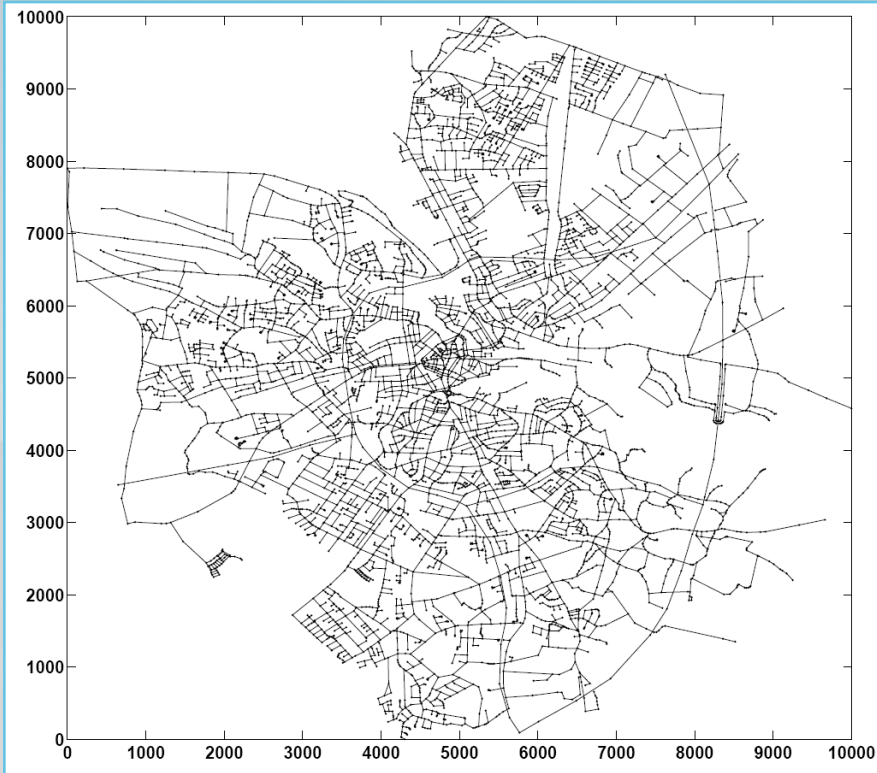
Μέθοδος Mtreeel

- Για κάθε τμήμα τροχιάς qs της T_q εκτελείται ένα απλό ερώτημα περιοχής μόνο στο NET-M-Tree. Αποτελέσματα στο σύνολο C_{net}
- Κάθε τμήμα του C_{net} ελέγχεται και στο χρόνο (αποτελέσματα στο AS).
- $AT =$ id's αντιστοιχών τροχιών από το AS .

Ερωτήματα Ομοιότητας Τροχιών σε Χωρικά Δίκτυα

Αξιολόγηση της Απόδοσης

■ Το οδικό δίκτυο του Oldenburg:



Βασικά στοιχεία

- 6105 κόμβοι
- 7135 ακμές
- Μέσος βαθμός κόμβων = 2,3
- Μέσο μήκος ακμών = 73,68 m
- Διάμετρος $D_G = 12.986$ m

Στοιχεία τροχιών

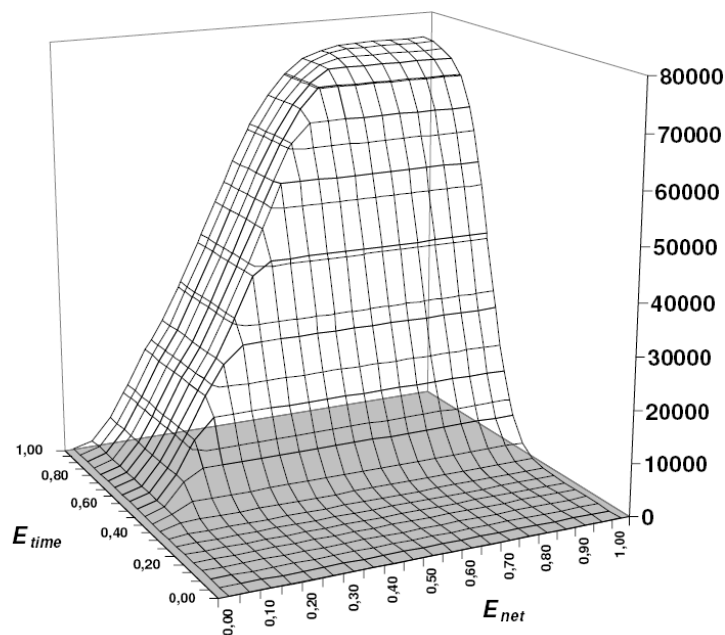
- 3797 τροχιές με μήκη 10-100
- 75144 τμήματα τροχιών ($\mu=10$)

Ερωτήματα Ομοιότητας Τροχιών σε Χωρικά Δίκτυα

Αξιολόγηση της Απόδοσης

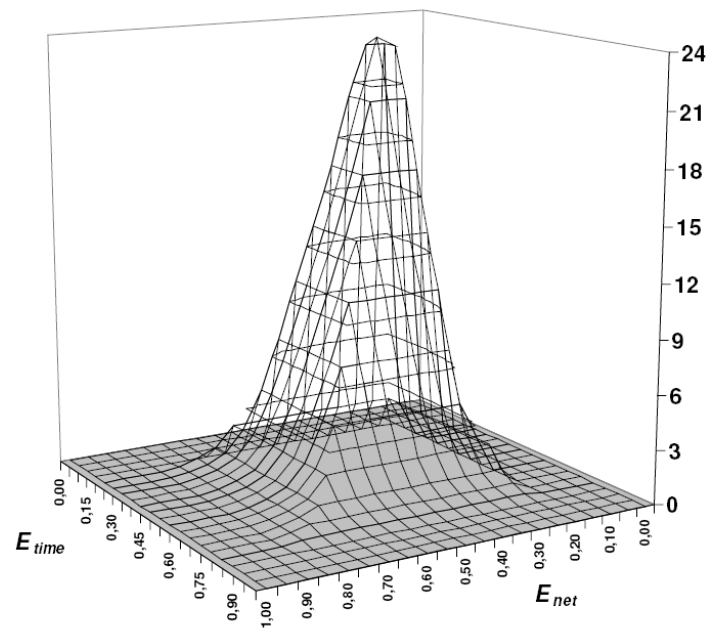
- Βασικά μεγέθη αποτελεσμάτων:

Number of (common-accepted) sub-trajectories (AS) using D_{net1} & D_{net2+1} (M-treel,II methods)



πλήθος τμημάτων στο AS

False alarms ratio (FA%) using D_{net2+1} (M-treel,II methods)

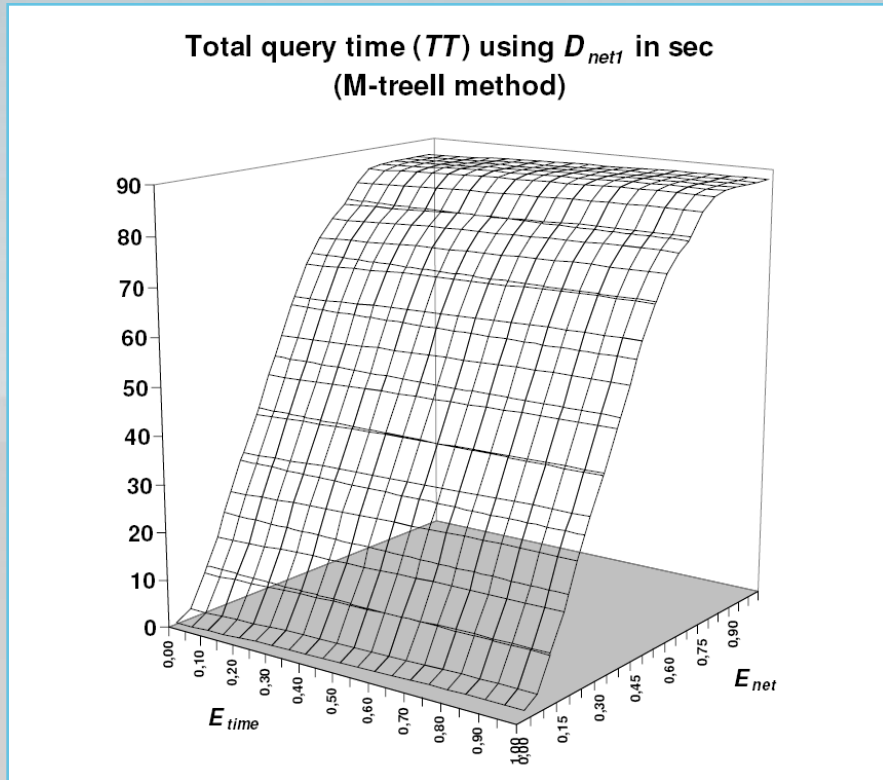


false alarms για το μέτρο D_{net2+1}
(κοντά στο 0,30 το μέγιστο)

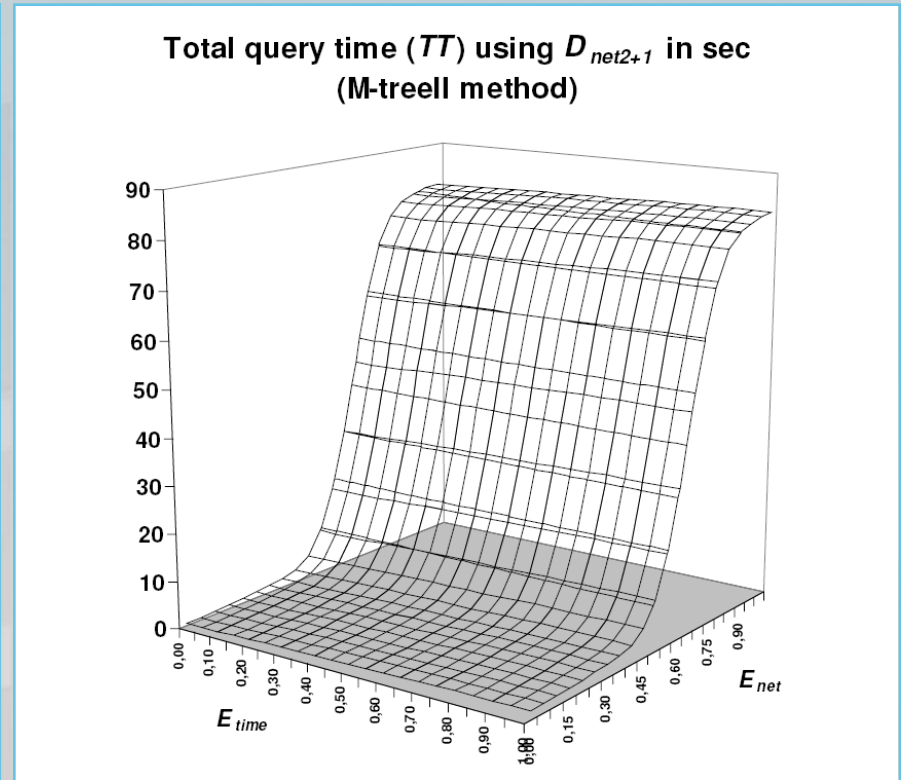
Ερωτήματα Ομοιότητας Τροχιών σε Χωρικά Δίκτυα

Αξιολόγηση της Απόδοσης

- Συνολικός χρόνος εκτέλεσης για την μέθοδο Mtrell:



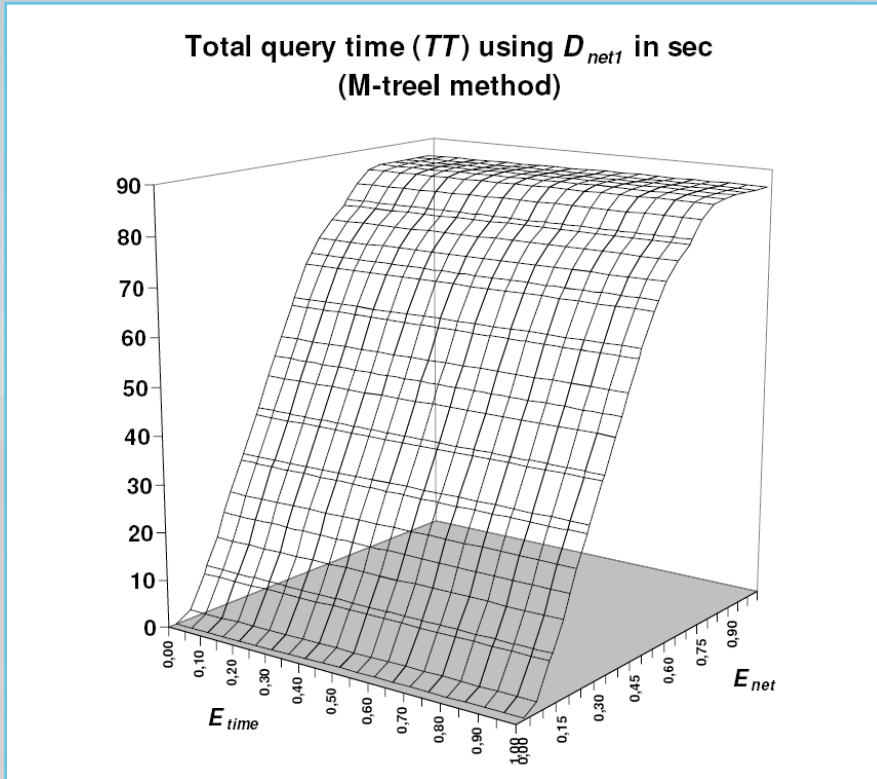
μέτρα D_{net1} , D_{time}



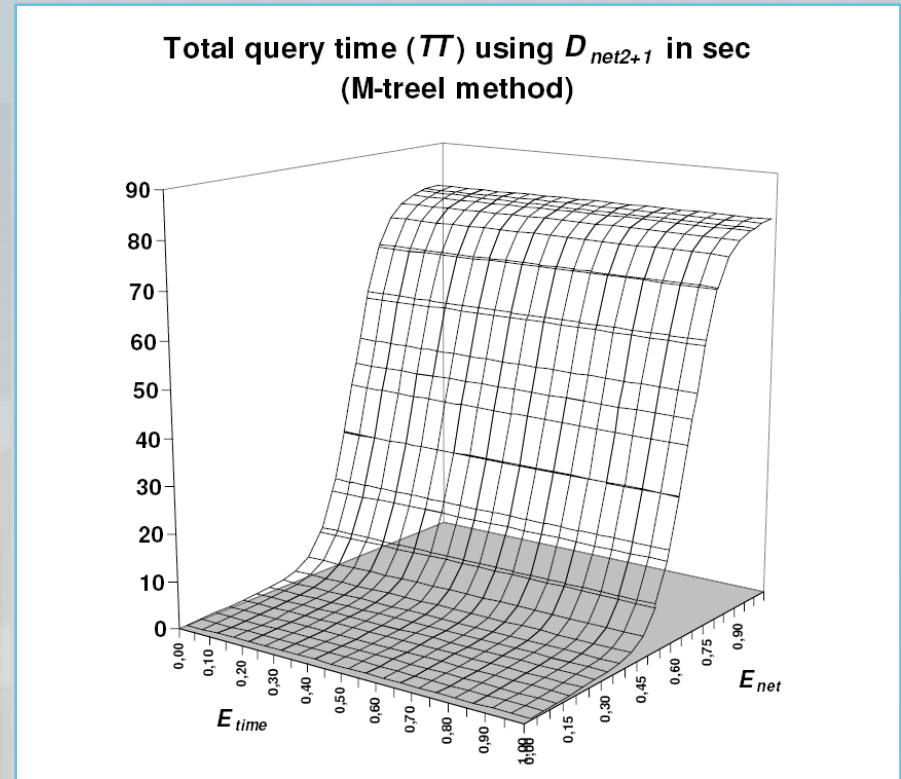
μέτρα D_{net2+1} , D_{time}

Ερωτήματα Ομοιότητας Τροχιών σε Χωρικά Δίκτυα Αξιολόγηση της Απόδοσης

- Συνολικός χρόνος εκτέλεσης για την μέθοδο Mtreel:

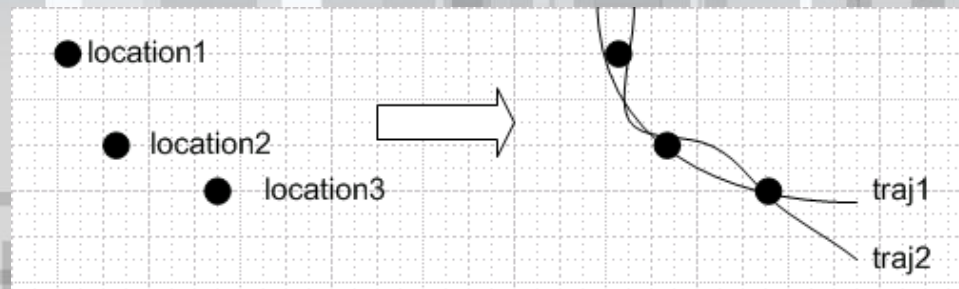


μέτρα D_{net1} , D_{time}



μέτρα D_{net2+1} , D_{time}

Ερωτήματα Ομοιότητας Τροχιών με Τοποθεσίες Χωρικών Δικτύων



Location-based Trajectory Similarity.

Ερωτήματα Ομοιότητας Τροχιών με Τοποθεσίες

Το πρόβλημα



- Εύρεση βέλτιστης τροχιάς η οποία να διέρχεται από καθορισμένα σημεία ενδιαφέροντος (τοποθεσίες)

Ερωτήματα Ομοιότητας Τροχιών με Τοποθεσίες

Έρευνα και Βιβλιογραφία

- Τελείως διαφορετικό πρόβλημα καθώς τα **σημεία ενδιαφέροντος (τοποθεσίες)** καθορίζουν επιπλέον περιορισμούς (με προτεραιότητα επίσκεψης ή χωρίς).
- Το πρόβλημα προτάθηκε αρχικά από τον Chen το 2010:
 - ❖ Οι τροχιές ορίζονται σε Ευκλείδειους χώρους, τα μέτρα ομοιότητας περιέχουν Ευκλείδειες αποστάσεις και γίνεται χρήση R-Trees.
 - ❖ Η αναζήτηση ομοιότητας γίνεται με Incremental k-NN τεχνικές από κάθε location.
- Βελτιώθηκε από τον Tang το 2011:
 - ❖ Οι τροχιές πάλι ορίζονται σε Ευκλείδειους χώρους και γίνεται χρήση R-Trees, αλλά η αναζήτηση ομοιότητας χρησιμοποιεί τον αλγόριθμο του Fagin (Threshold Algorithm).

Ερωτήματα Ομοιότητας Τροχιών με Τοποθεσίες

Έρευνα και Βιβλιογραφία

- Ακόμα μία βελτίωση από τον Wang το 2012:
- ❖ Δίνεται δυνατότητα στον χρήστη να ορίσει **προτεραιότητες επίσκεψης**. Γίνεται χρήση του Heaviest Common Subsequence μέτρου. Οι τροχιές όμως πάλι ορίζονται σε Ευκλείδειους χώρους.
- Οι μόνες δουλειές σε Spatial Networks είναι του Shang (2012, 2013):
- ❖ Στο EDBT-2012 προτείνεται αλγόριθμος που συνδυάζει την αναζήτηση στον χώρο με textual attributes, ενώ το 2013 γίνεται συνδυασμός και στον χρόνο με κατάλληλα μέτρα (VLDB-Journal).
- ❖ Μεγάλο πρόβλημα το θέμα των πολλαπλών visits στα trajectory data και το ότι απαιτείται ο προϋπολογισμός όλων των ελαχίστων αποστάσεων του χωρικού δικτύου (**all-to-all shortest path distances**)

Ερωτήματα Ομοιότητας Τροχιών με Τοποθεσίες

Μέτρα Ομοιότητας

- Βασικό μέτρο ομοιότητας τροχιάς T στον **χώρο** με τις επιλεγμένες τοποθεσίες $q_j \in Q$:

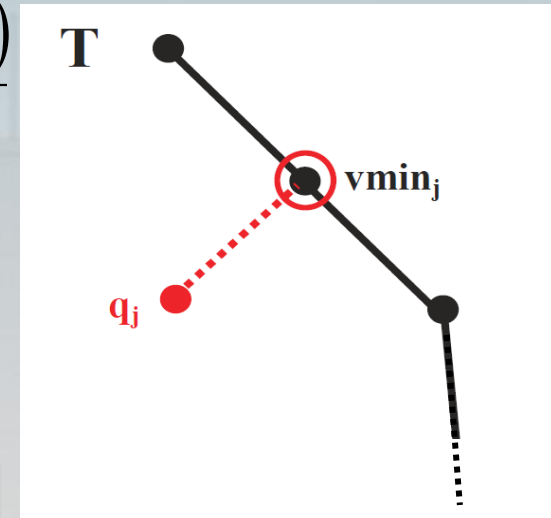
$$D_{net}(Q, T) = \frac{1}{m} \sum_{j=1}^m d(q_j, T) \quad d(q_j, T) = \min_{v_i \in T} \frac{c(q_j, v_i)}{D_G}$$

- Βασικό μέτρο ομοιότητας τροχιάς T στον **χρόνο** με τις επιλεγμένες τοποθεσίες $q_j \in Q$:

$$D_{time}(Q, T) = \frac{1}{m-1} \sum_{\substack{j=2, \dots, m \\ dt_j > qt_j}} \frac{dt_j - qt_j}{\max\{dt_j, qt_j\}}$$

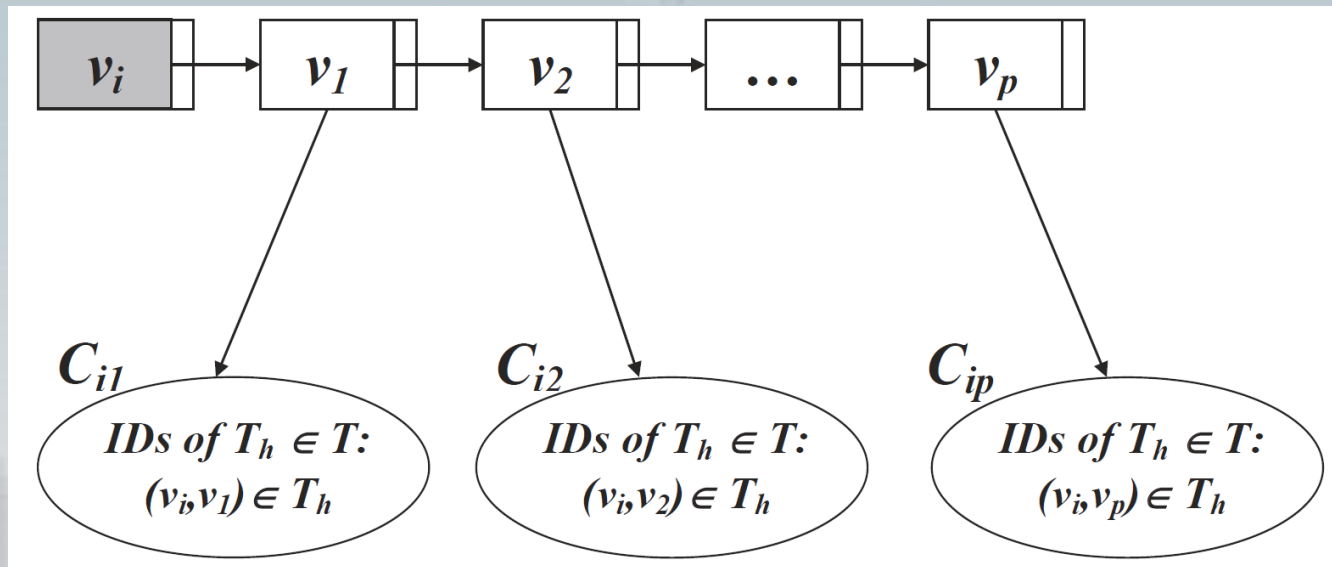
- qt_j : χρονικό διάστημα ανοχής καθορισμένο από τον χρήστη για την μετάβαση από την τοποθεσία q_{j-1} στην q_j (αν δεν υπάρχει τότε ∞).
- dt_j : χρονικό διάστημα μεταξύ των κοντινότερων κόμβων $vmin_{j-1}$, $vmin_j$ στις τοποθεσίες q_{j-1} , q_j αντίστοιχα.

$$D(Q, T) = a \cdot D_{net}(Q, T) + (1-a) \cdot D_{time}(Q, T)$$



Ερωτήματα Ομοιότητας Τροχιών με Τοποθεσίες

Δεικτοδότηση Τροχιών



- Χρήση δομής **adjacency list** για τον γράφο του χωρικού δικτύου.
- Επέκταση της δομής με κατάλληλα **Trajectory-ID Clusters**.
- ❖ **Linear space** για αποθήκευση γράφου και τροχιών, **Linear time** για την κατασκευή της δομής, Κατανομή των τροχιών σε πολλά μικρά clusters, Μείωση των πολλαπλών visits στα δεδομένα των τροχιών.

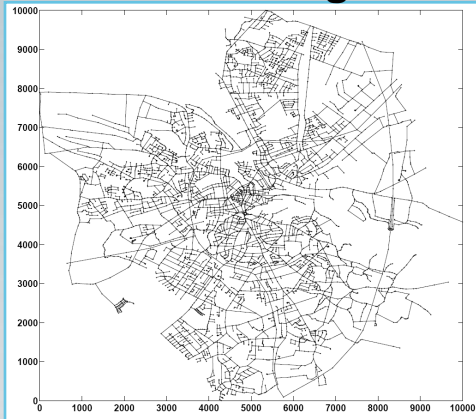
Ερωτήματα Ομοιότητας Τροχιών με Τοποθεσίες

Προοδευτικός Αλγόριθμος

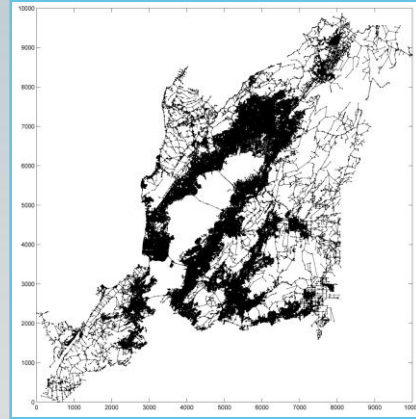
- Από κάθε μία τοποθεσία $q_j \in Q$ γίνεται **Dijkstra expansion** με **round-robin** εναλλαγή.
- Από κάθε edge που βρέθηκε στο expansion level και μέσω της προηγούμενης δομής επιλέγονται τα trajectory IDs από τα αντίστοιχα clusters.
- Σε κάθε trajectory που βρέθηκε υπολογίζονται οι αποστάσεις του D_{net} , D_{time} και D . Μετά τον υπολογισμό ένα flag γίνεται true ώστε να μην υπολογιστούν ξανά ακόμα και αν βρεθούν σε άλλο cluster.
- Οι top-k min αποστάσεις και τα αντίστοιχα trajectory IDs κρατούνται σε σωρό (k-μόνο στοιχείων) και ενημερώνονται.
- Ενημερώνεται ένα **threshold**:
$$L = \frac{a}{m} \sum_{j=1}^m \frac{c(q_j, v \min_j)}{D_G}$$
 όπου $v \min_j$ είναι ο κοντινότερος κόμβος στην q_j στο τρέχον επίπεδο ανάπτυξης.
- Κάθε τροχιά στο τρέχον top-k που έχει απόσταση D μικρότερη του L εμφανίζεται **προοδευτικά** στα αποτελέσματα.

Ερωτήματα Ομοιότητας Τροχιών με Τοποθεσίες Πειράματα και Αποτελέσματα

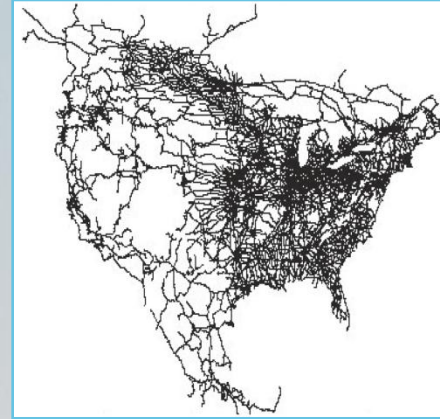
Oldenburg



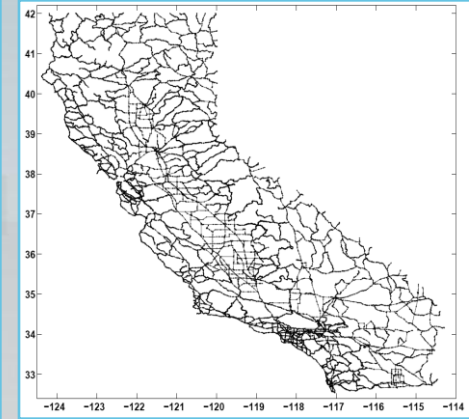
San Francisco



North America



California

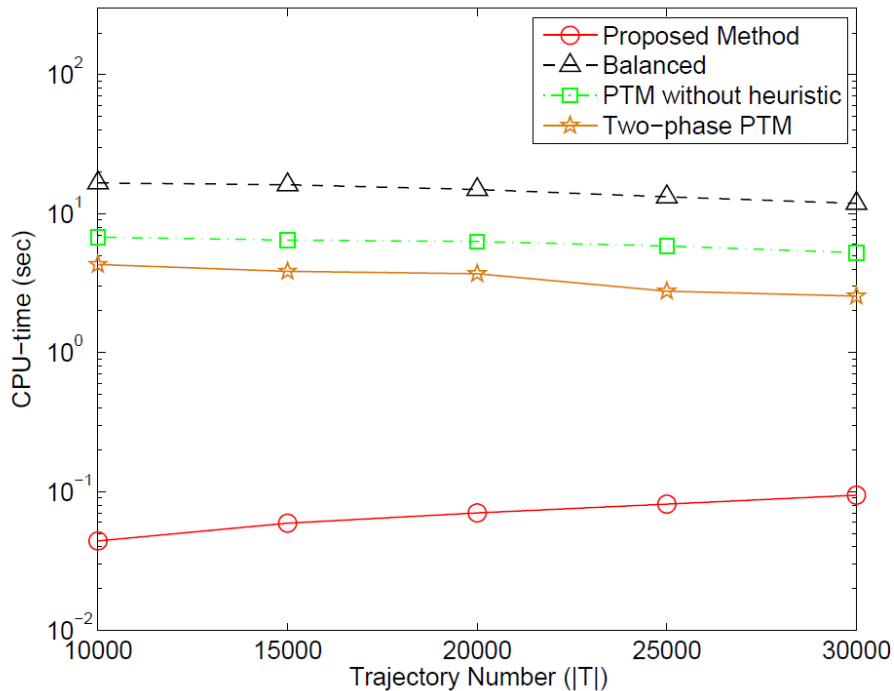


	OL	SF	NA	CAL
Nodes ($ V $)	6,105	174,956	175,813	1,965,206
Edges ($ E $)	7,035	223,001	179,179	5,533,214
Maximum Node Degree	5	8	14	12
Average Node Degree	2.3046683	2.5492238	2.0382907	5.6311796
Average Edge Weight	73.67902	8.782676	3.527082	0.0161862
Graph Diameter (D_G)	12,985.97	16,828.54	15,476.12	16.4288

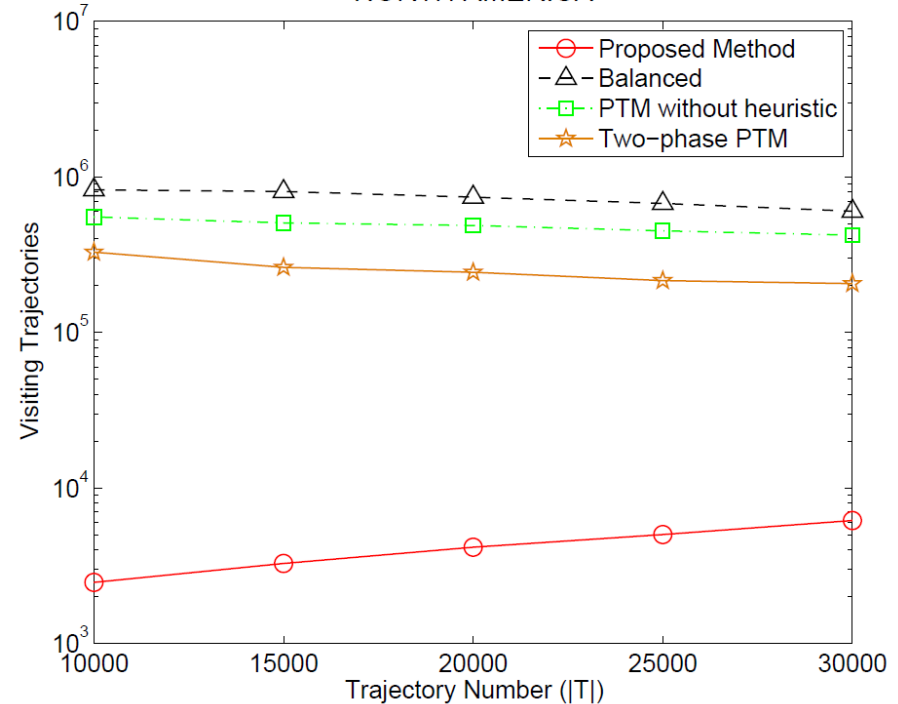
	OL	SF	NA	CAL
# of Trajectories ($ T $)	10,000	15,000	30,000	40,000
# of Trajectory Spatial Points ($ R $)	423,989	714,453	1,628,191	1,795,614
# of Trajectory Edges ($ RE $)	413,989	699,453	1,598,191	1,755,614

Ερωτήματα Ομοιότητας Τροχιών με Τοποθεσίες Πειράματα και Αποτελέσματα

NORTH AMERICA



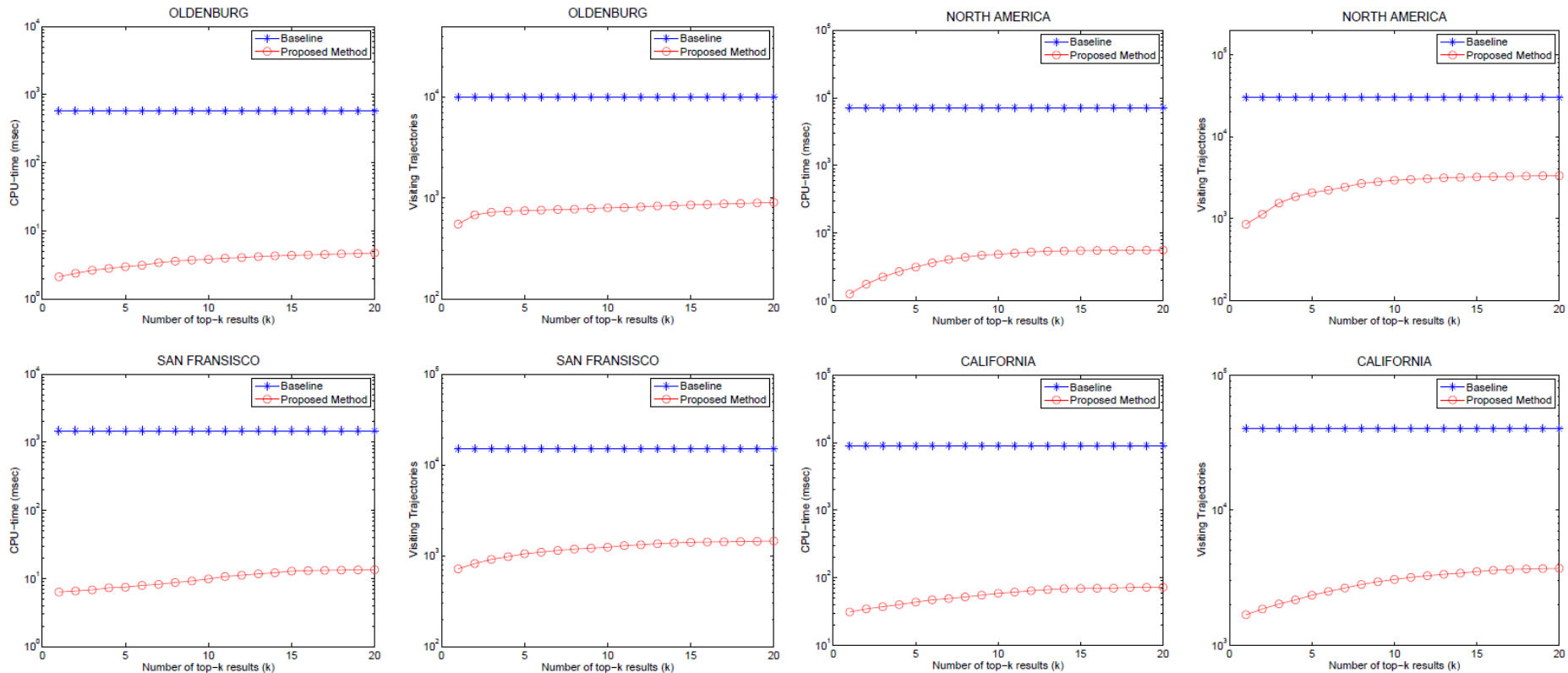
NORTH AMERICA



- Χρόνοι εκτέλεσης και πλήθος visits στις τροχιές (που αντιπροσωπεύει το κόστος I/O) τουλάχιστον **δύο τάξεις μεγέθους καλύτερα** σε σχέση με τις μεθόδους του Shang.

Ερωτήματα Ομοιότητας Τροχιών με Τοποθεσίες

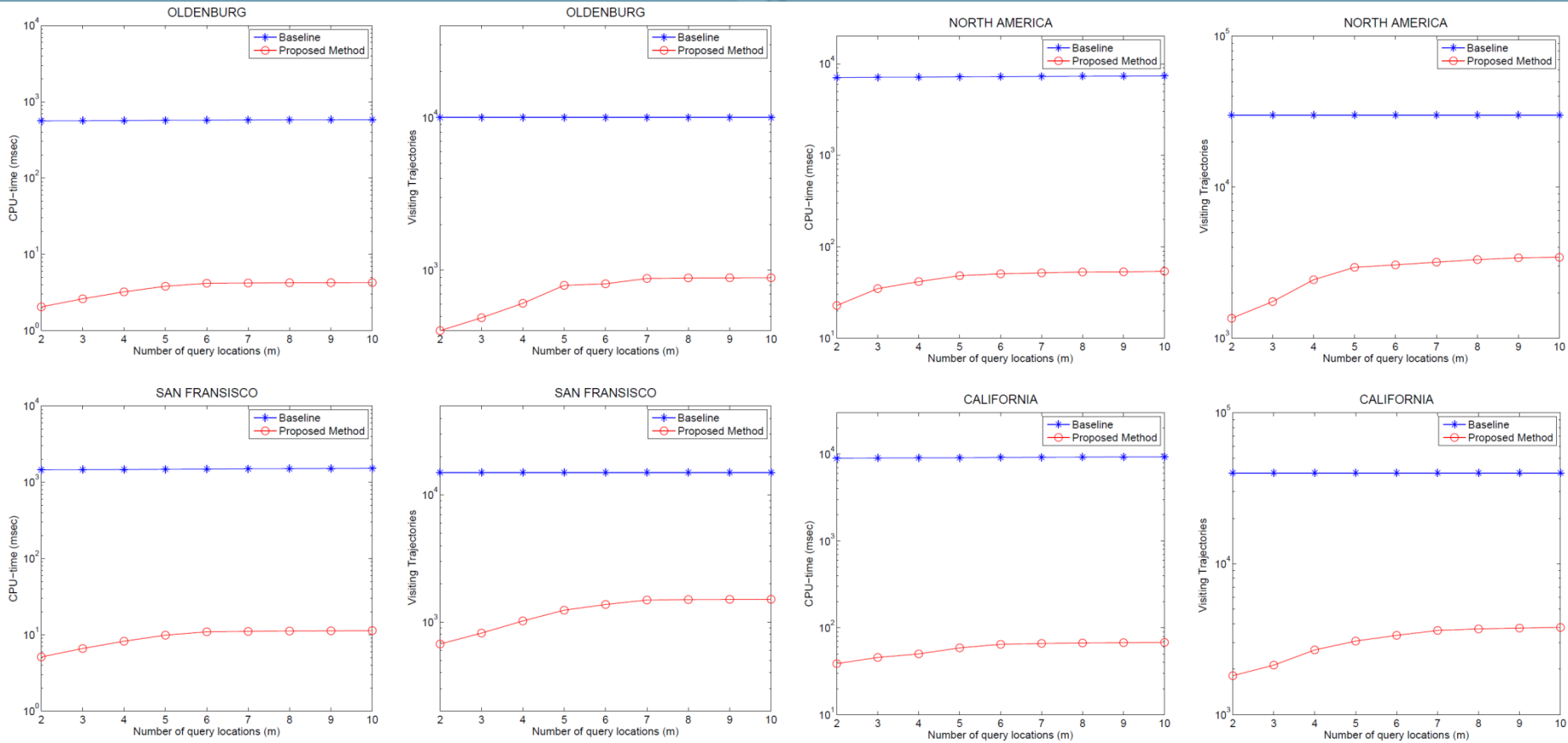
Πειράματα και Αποτελέσματα



- Η προοδευτική εμφάνιση των αποτελεσμάτων δίνει πολλά πλεονεκτήματα:
- ❖ Ο χρήστης παραλαμβάνει μία-μία τις τροχιές σε διαφορετικές χρονικές στιγμές χωρίς να περιμένει να πάρει όλο το top-k μαζί.

Ερωτήματα Ομοιότητας Τροχιών με Τοποθεσίες

Πειράματα και Αποτελέσματα



❖ Το πλήθος των τοποθεσιών επηρεάζει γραμμικά την απόδοση του αλγορίθμου.

Εκτίμηση της Επιλεξιμότητας Ερωτημάτων σε Χωρικά Δίκτυα

Oldenburg



Εκτίμηση της Επιλεξιμότητας ερωτημάτων σε Χωρικά Δίκτυα – Κίνητρα και Εφαρμογές

- Η εκτίμηση της επιλεξιμότητας ενός ερωτήματος είναι η εκτίμηση της περιοχής του χώρου των δεδομένων που πρόκειται αυτό να προσπελάσει κατά την εκτέλεσή του. Το μέγεθος αυτής της περιοχής έχει επίπτωση και στο συνολικό I/O κόστος του ερωτήματος.
- Δεν έχει γίνει εκτίμηση της επιλεξιμότητας σε χωρικά δίκτυα αλλά μόνο σε δεδομένα πάνω σε ευκλείδειους χώρους (R-Trees, Power Method, Hausdorff fractal dimension) κλπ.
- Σημαντικές εφαρμογές στην βελτιστοποίηση ερωτημάτων και στην εκτίμηση του κόστους I/O.
- Εφαρμογές στην πρόβλεψη τροχιών και θέσεων κινούμενων αντικειμένων σε δίκτυα.
- Εφαρμογές στην εκτίμηση του πλήθους των αντικειμένων που βρίσκονται γύρω από ένα προκαθορισμένο αντικείμενο.
- Εφαρμογές στη συνεχή αξιολόγηση ερωτημάτων (π.χ. συνεχής παρακολούθηση ενός συνόλου αντικειμένων γύρω από ένα κινούμενο αντικείμενο σε συγκεκριμένη απόσταση).

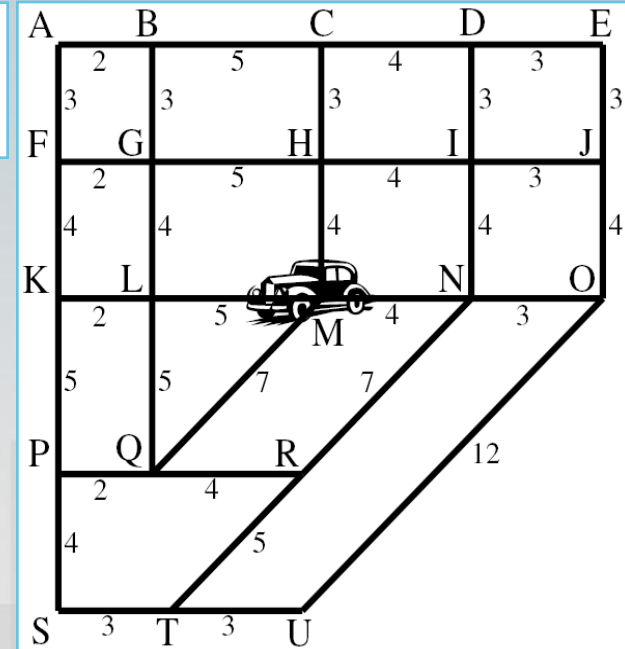
Εκτίμηση της Επιλεξιμότητας ερωτημάτων σε Χωρικά Δίκτυα – Ορισμός του προβλήματος

- Το πλήθος των κόμβων και των ακμών στην περιοχή ενδιαφέροντος από έναν κόμβο v_0 με βάση μία ακτίνα e είναι:

$$N(v_0, e) = |\{v \in V_G : d(v, v_0) \leq e\}|$$

$$E(v_0, e) = |\{(v_i, v_j) \in E_G : \min(d(v_i, v_0), d(v_j, v_0)) + w(v_i, v_j) \leq e\}|$$

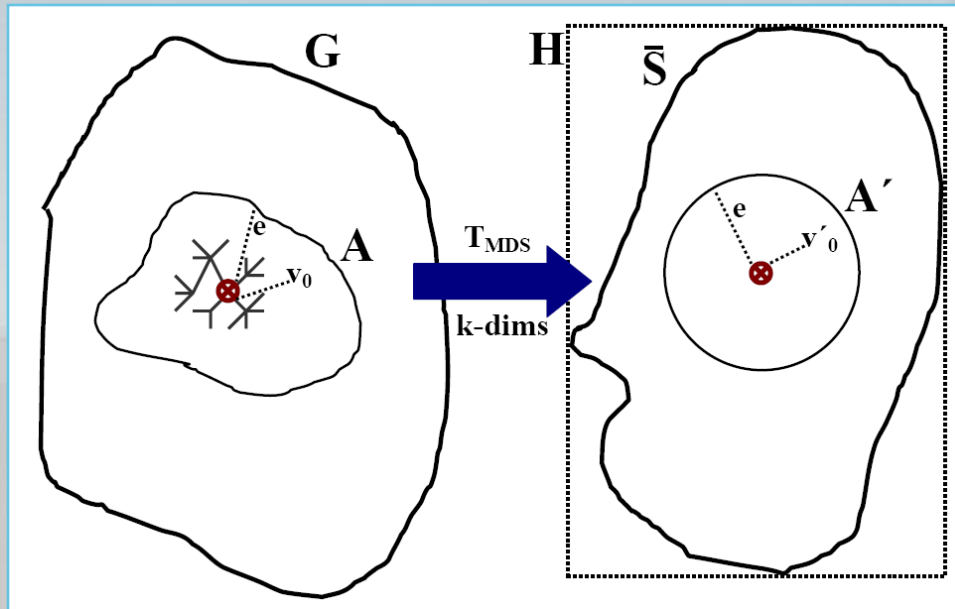
- Αναζητείται η εύρεση αποτελεσματικών εκτιμητών $\tilde{N}(v_0, e)$, $\tilde{E}(v_0, e)$.
- Το κόστος των περισσότερων βασικών και σύνθετων ερωτημάτων αναλύεται σε κόστη συγκεκριμένων ερωτημάτων περιοχής (π.χ. top-k, join, κλπ.) σε επίπεδο μηχανής.



e	$N(v_0, e)$	Κόμβοι	$E(v_0, e)$	Ακμές
0,1,2,3	1	M	0	—
4	3	M,N,H	2	MH, MN
5,6	4	M,N,H,L	3	MH, MN, ML

Εκτίμηση της Επιλεξιμότητας ερωτημάτων σε Χωρικά Δίκτυα – Η μέθοδος MDS-grid

- Μετασχηματισμός του δικτύου σε Ευκλείδειο χώρο k κατάλληλων διαστάσεων ώστε να διατηρούνται οι αρχικές αποστάσεις των κόμβων:



- Εφαρμογή ενός πλέγματος H με c^k υπερ-ορθογώνια κελιά.
- Ορισμός της αντίστοιχης περιοχής A' στον Ευκλείδειο χώρο.
- Μέτρηση της πυκνότητας CD σε κάθε κελί που βρίσκεται μέσα ή που τέμνεται με την περιοχή A' .
- Άθροισμα των πυκνοτήτων.

$$\tilde{N}(v_0, e) = \sum_{C(i_1, i_2, \dots, i_k) \subset A'} CD(i_1, i_2, \dots, i_k) + \frac{1}{2} \cdot \sum_{\substack{C(i_1, i_2, \dots, i_k) \cap A' \neq \emptyset \\ \wedge C(i_1, i_2, \dots, i_k) \not\subset A'}} CD(i_1, i_2, \dots, i_k)$$

Εκτίμηση της Επιλεξιμότητας ερωτημάτων σε Χωρικά Δίκτυα – Οι μέθοδοι Global και Local

- **Μέθοδος Global:** εκτίμηση με καθολικές παραμέτρους του δικτύου (μέσο βάρος ακμών, μέσος βαθμός κόμβων):

$$\tilde{N}(v_0, e) = \tilde{N}(e) = \frac{\overline{deg}}{2} \cdot \frac{e}{w} \cdot \left(\frac{e}{w} + 1\right) + 1$$

$$\tilde{E}(v_0, e) = \tilde{E}(e) = \overline{deg} \cdot \frac{e^2}{w^2}$$

- **Μέθοδος local:** εκτίμηση με τοπικές πυκνότητες κόμβων και ακμών με την εκλογή μίας σταθερής μικρής απόστασης e_c :

$$LND_v = |\{v_i \in V_G : d(v, v_i) \leq e_c\}|$$

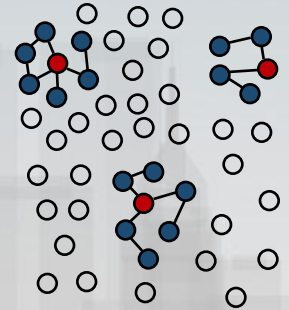
$$LED_v = |\{(v_i, v_j) \in E_G : \min(d(v, v_i), d(v, v_j)) + w(v_i, v_j) \leq e_c\}|$$

$$NLND_v = \frac{LND_v}{\max\{LND_{v_i}, v_i \in V_G\}}$$

$$NLED_v = \frac{LED_v}{\max\{LED_{v_i}, v_i \in V_G\}}$$

$$\tilde{N}(v_0, e) = NLND_{v_0} \cdot \left(\frac{\overline{deg}}{2} \cdot \frac{e}{w} \cdot \left(\frac{e}{w} + 1\right) + 1\right)$$

$$\tilde{E}(v_0, e) = NLED_{v_0} \cdot \overline{deg} \cdot \frac{e^2}{w^2}$$



(Κάθε κόμβος έχει τις δικές του τοπικές πυκνότητες κόμβων και ακμών)

Εκτίμηση της Επιλεξιμότητας ερωτημάτων σε Χωρικά Δίκτυα – Οι μέθοδοι Kernel και Binary

- **Μέθοδος Kernel:** εκτίμηση με τοπικές πυκνότητες πυρήνων κόμβων και ακμών με βάση ένα προκαθορισμένο εύρος ζώνης h :

$$LND_v = \sum_{v_i \in V_G} K\left(\frac{d(v, v_i)}{h}\right) \quad LED_v = \sum_{(v_i, v_j) \in E_G} \left(K\left(\frac{d(v, v_i)}{h}\right) + K\left(\frac{d(v, v_j)}{h}\right) \right)$$

- Χρήση κυρίως του Gaussian Kernel:
$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$
- **Μέθοδος Binary:** εκτίμηση με δυαδική κωδικοποίηση των αποστάσεων:
 - Εκλογή μίας απόστασης w_u ως μοναδιαίας ($w_u < \bar{w}$).
 - Μετασχηματισμός του δικτύου σε νέο (χωρίς βάρη), διαιρώντας κάθε ακμή σε μοναδιαίες με τη προσθήκη $|A_G|$ ενδιάμεσων κόμβων.
$$|A_G| \approx |E_G| \cdot \frac{\bar{w} - w_u}{w_u}$$
 - Υπολογισμός δυαδικών κωδικών c_v για κάθε κόμβο v με προσαρμογή υπερκύβων (μέθοδος Gupta).

Εκτίμηση της Επιλεξιμότητας ερωτημάτων σε Χωρικά Δίκτυα – Η μέθοδος Binary – Σύγκριση

- Η απόσταση **Hamming** κάθε ζεύγους κωδικών προσεγγίζει την απόσταση δικτύου των αντιστοίχων κόμβων τους.

$$H(c_{v_i}, c_{v_j}) = \sum_{bits=1} (c_{v_i} \oplus c_{v_j})$$

- Συνάρτηση προσήμου: $u(x) = 1$ όταν $x \geq 0$ αλλιώς $u(x) = 0$.

- Τύποι τελικών εκτιμήσεων:

$$\tilde{N}(v_0, e) = \sum_{v_i \in V_G} u \left(e - \frac{H(c_{v_0}, c_{v_i}) \cdot w_u}{2} \right)$$

$$\tilde{E}(v_0, e) = \sum_{(v_i, v_j) \in E_G} u \left(e - w_u \cdot \frac{\min\{H(c_{v_0}, c_{v_i}), H(c_{v_0}, c_{v_j})\}}{2} - w(v_i, v_j) \right)$$

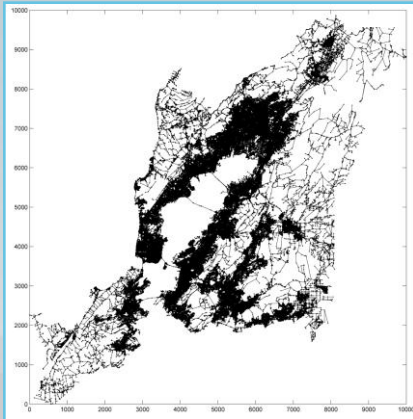
- Σύγκριση απαιτήσεων και **πολυπλοκότητας** των μεθόδων:

Μέθοδος	Χρόνος εκτιμήσεων (node/edge)	Απαιτούμενος χώρος	Χρόνος προϋπολογισμών
MDS-grid	$O \left(\prod_{i=1, \dots, k} \lceil \frac{2e}{dx_i} \rceil \right)$	$O(k \cdot V_G + c^k)$	$O(V_G ^3 + c^k)$
Global	$O(1)/O(1)$	$O(1)$	$O(V_G + E_G)$
Local	$O(1)/O(1)$	$O(V_G)$	$O \left(\frac{e_c}{D_G} \cdot V_G ^2 \cdot \log V_G \right)$
Kernel	$O(1)/O(1)$	$O(V_G)$	$O \left(\frac{k \cdot h}{D_G} V_G ^2 \cdot \log V_G \right)$
Binary	$O(V_G)/O(E_G)$	$O \left(V_G \cdot \sqrt{ V_G + E_G } \cdot \frac{\bar{w} - w_u}{w_u} \right)$	$O(n\sqrt{n}),$ $n = \left(V_G + E_G \cdot \frac{\bar{w} - w_u}{w_u} \right)$

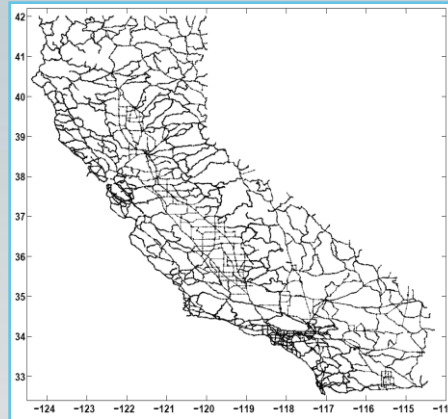
Εκτίμηση της Επιλεξιμότητας ερωτημάτων σε Χωρικά Δίκτυα – Πειραματικά Αποτελέσματα

- Δίκτυα που χρησιμοποιήθηκαν:

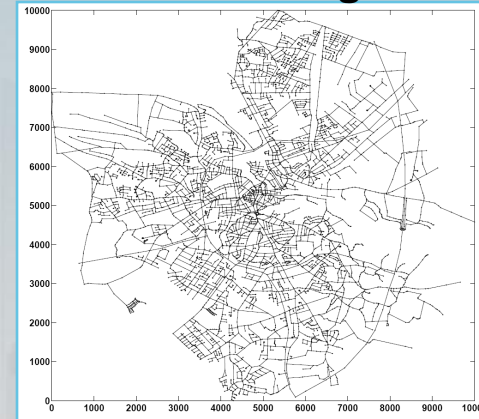
San Francisco



California



Oldenburg



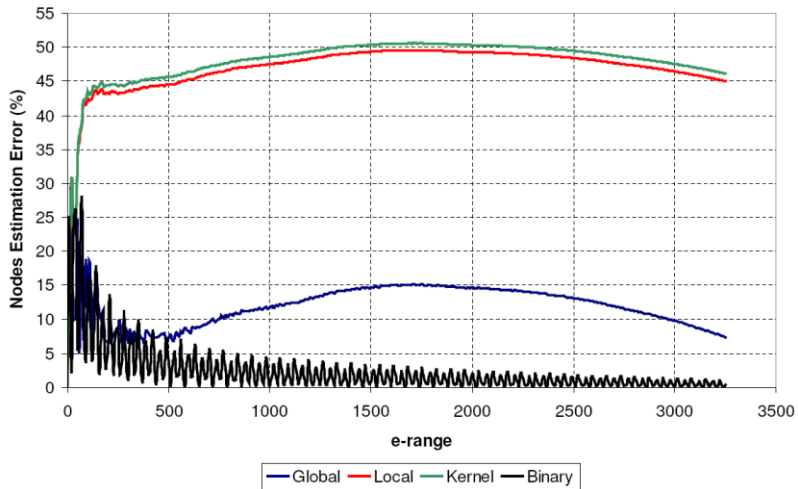
Παράμετρος	Oldenburg	California	San Francisco	Συνθετικό
$ V_G $	6105	21048	174956	250000
$ E_G $	7035	21693	223001	499000
deg	2,304668	2,061289	2,549224	3,992
\bar{w}	73,67902	0,01618624	8,782676	15,00156
D_G	12985,97	16,4288	16828,54	13366,2

- Επιλογή των τιμών των ρυθμίσιμων παραμέτρων των μεθόδων:

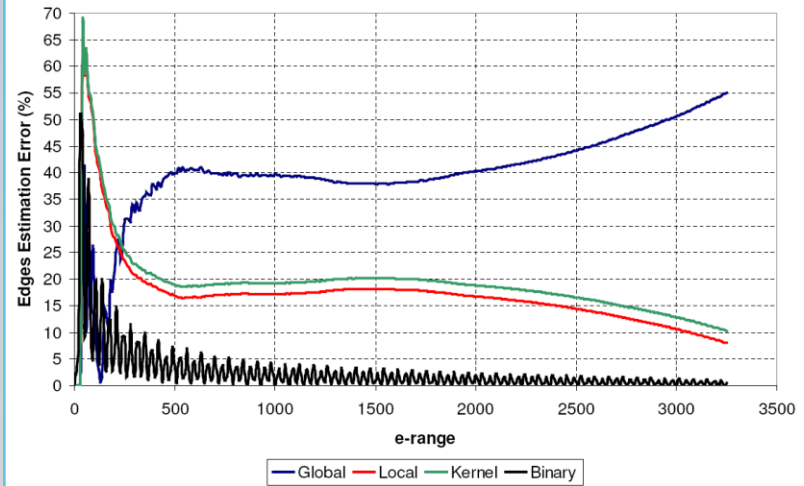
$$e_c = \sqrt{\frac{D_G}{4\bar{w}}} \quad h = \frac{\bar{w}}{3} \quad w_u = \frac{\bar{w}}{3}$$

Εκτίμηση της Επιλεξιμότητας ερωτημάτων σε Χωρικά Δίκτυα – Πειραματικά Αποτελέσματα

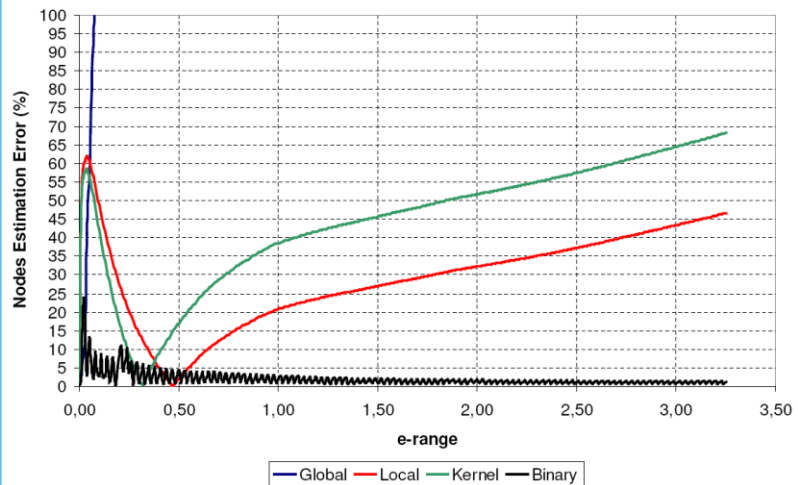
Estimation Accuracy in Oldenburg (Nodes)



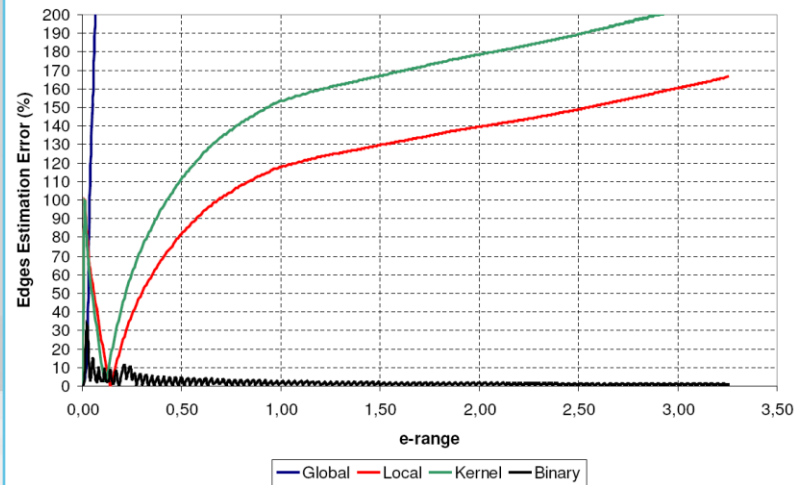
Estimation Accuracy in Oldenburg (Edges)



Estimation Accuracy in California (Nodes)

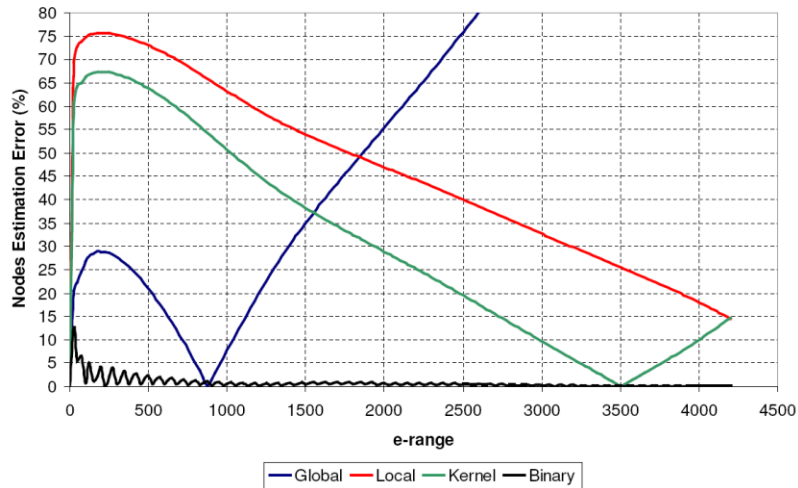


Estimation Accuracy in California (Edges)

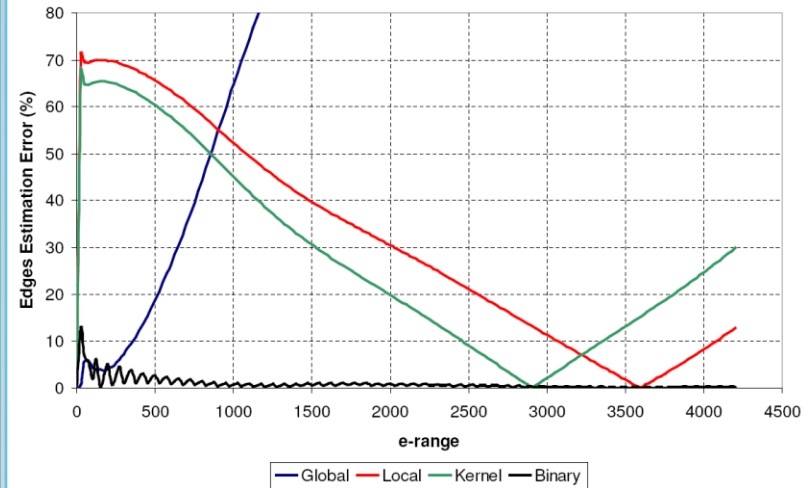


Εκτίμηση της Επιλεξιμότητας ερωτημάτων σε Χωρικά Δίκτυα – Πειραματικά Αποτελέσματα

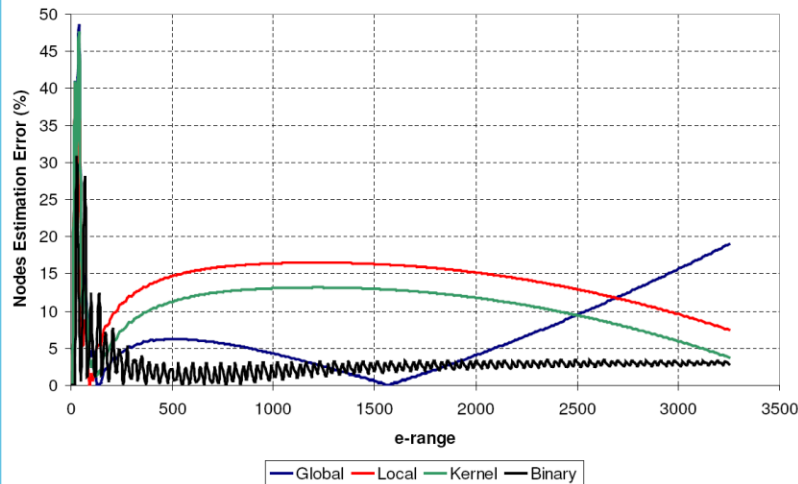
Estimation Accuracy in San Francisco (Nodes)



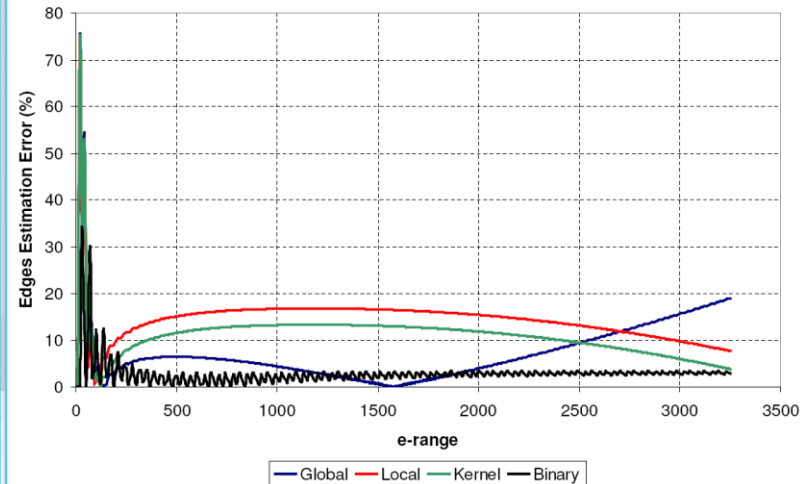
Estimation Accuracy in San Francisco (Edges)



Estimation Accuracy in Synthetic (Nodes)

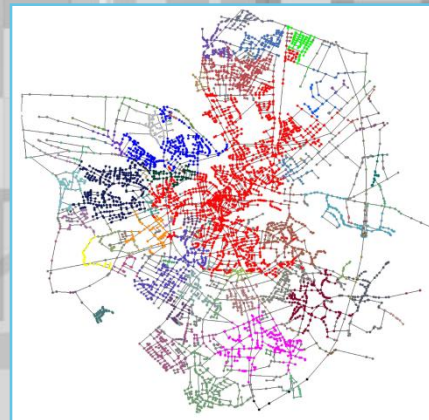
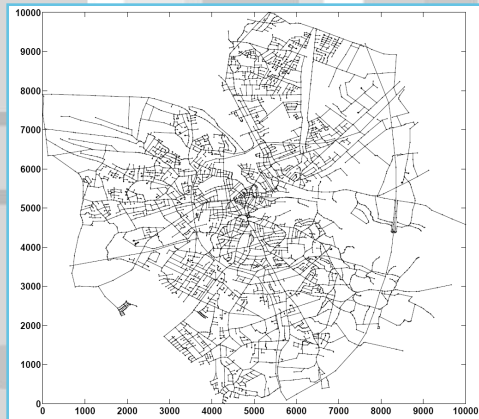


Estimation Accuracy in Synthetic (Edges)



Ομαδοποίηση Κόμβων Γράφου με Μεταβατική Ομοιότητα

Oldenburg

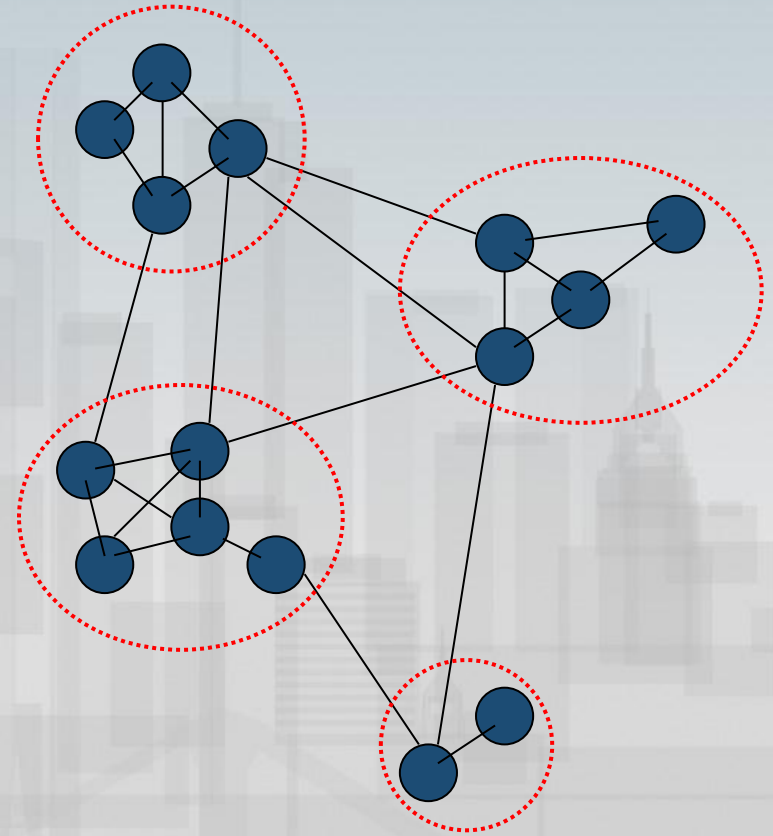


Ομαδοποίηση Κόμβων Γράφου με Μεταβατική Ομοιότητα – Ορισμός - Κίνητρα - Εφαρμογές

- Στόχος είναι ο εντοπισμός των **πυκνών υπο-γράφων**, με βάση ένα μέτρο **ομοιότητας / απόστασης**.
- Πολλές υπάρχουσες μέθοδοι αλλά με αρκετά μειονεκτήματα:
 - μέθοδοι **top-down, bottom-up**: αυξημένο κόστος υπολογισμών λόγω των αναδρομών και των aggregations.
 - μέθοδοι **spectral clustering**: αυξημένες απαιτήσεις σε χώρο και σε προϋπολογισμούς, μη αποδοτικές σε μεγάλους γράφους.
 - μέθοδοι **kernel based clustering**: αυξημένες απαιτήσεις σε χώρο (χρήση τετραγωνικών πινάκων, πράξεων τους κλπ.).
 - Έτσι δεν μπορούν να υποστηρίξουν αποδοτικά μεγάλους γράφους.
- Σημαντικές εφαρμογές σε **graph data mining, knowledge discovery**.
- Εφαρμογές σε **ανακάλυψη κοινοτήτων** στο διαδίκτυο και στα **κοινωνικά δίκτυα**.

Ομαδοποίηση Κόμβων Γράφου με Μεταβατική Ομοιότητα – Συνεισφορά

- Ο νέος αλγόριθμος ομαδοποίησης έχει τις εξής ιδιότητες:
 - Υπολογισμός ομοιότητας **on-the-fly** (δεν απαιτείται χώρος για τις τιμές ομοιότητας).
 - **Προοδευτικός** εντοπισμός ομάδων (1 προς 1).
 - Χαμηλή πολυπλοκότητα (**scalability support**).
 - Εφαρμογή σε γράφους με **βάρη** ή **χωρίς**.
 - Εφαρμογή σε **τμήματα** γράφων.
 - Εφαρμογή και σε **μη συνδεδεμένους** γράφους.



Ομαδοποίηση Κόμβων Γράφου με Μεταβατική Ομοιότητα – Μέτρα Ομοιότητας

- Για δύο γειτονικούς κόμβους $v_i \sim v_j$:

$$sim(v_i, v_j) = \frac{w_{ij}}{\sum_{k=1}^{deg(v_i)} w_{ik} + \sum_{k=1}^{deg(v_j)} w_{jk} - w_{ij}}$$

$$sim(v_1, v_2) = \frac{2}{(2+3+1)+(2+1)-2} = \frac{2}{7} = 0.28571$$

- Για δύο μη γειτονικούς κόμβους v_a, v_b :

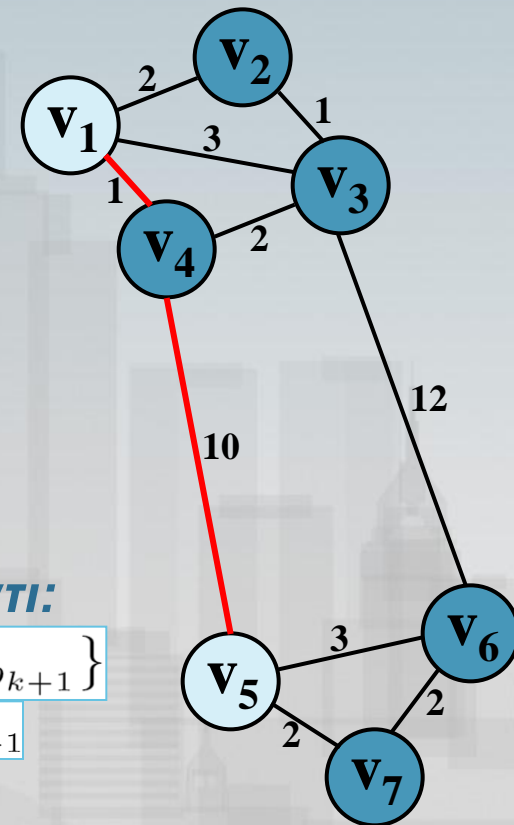
$$sim(v_a, v_b) = \prod_{i=1}^k sim(v_{p_i}, v_{p_{i+1}})$$

ελάχιστο μονοπάτι:

$$p = \{v_{p_1}, v_{p_2}, \dots, v_{p_{k+1}}\}$$

$$v_a = v_{p_1} \quad v_b = v_{p_{k+1}}$$

$$sim(v_1, v_5) = sim(v_1, v_4) \cdot sim(v_4, v_5) = \frac{1}{18} \cdot \frac{10}{18} = 0.03086$$



Ομαδοποίηση Κόμβων Γράφου με Μεταβατική Ομοιότητα – Ο Αλγόριθμος Ομαδοποίησης

- Προαιρετική προβαθμολόγηση των κόμβων: $rank[]$
- Οριακή τιμή ομοιότητας για την κατασκευή των ομάδων: r
- Όταν $r=0$ κάθε κόμβος αποτελεί και μία ομάδα.
- Όταν $r=1$ κάθε συνεκτικό κομμάτι του γράφου γίνεται ομάδα.
- Όταν $0 < r < 1$ το πλήθος των ομάδων c είναι αντιστρόφως ανάλογο του r .
- Πολυπλοκότητα: $O(c(M+N \log N))$.
- Χώρος: $O(N+M)$.

Graph-Node-Clustering($G, r, \text{optional } rank[]$)

01. **if** $rank$ not defined **then** $rank[i] = i, \forall i = 1, \dots, N$

02. $cluster[i] = 0, \forall i = 1, \dots, N$

03. $v_s = rank[1], curCluster = 1, H = \emptyset$

04. **while** $v_s \neq 0$ **do**

Similarity Computations

05. $s[i] = 0, parent[i] = 0, dist[i] = \infty, \forall i = 1, \dots, N$

06. empty the heap $H, v = v_s, dist[v] = 0, s[v] = 1, \text{insert } v \text{ into } H$

07. **while** H is not empty **do**

08. **for** all adjacent nodes u of v ($u \sim v$)

09. **if** $dist[u] > dist[v] + w_{uv}$ **then**

10. $dist[u] = dist[v] + w_{uv}, parent[u] = v$

11. $s[u] = s[v] * sim(v, u)$ ←

12. **end-if**

13. **if** $u \notin H \wedge cluster[u] = 0$ **then** insert u into H ←

14. **end-for**

15. $v = \text{top item of } H$ (if any)

16. **end-while**

17. $cluster[v_s] = curCluster$

18. **for** $i = 1$ to N

19. **if** $s[i] \geq r \wedge cluster[i] = 0$ **then** $cluster[i] = curCluster$ ←

20. **end-for**

21. $v_s = 0, i = 1$

22. **while** $i \leq N \wedge cluster[rank[i]] \neq 0$ **do** $i++$

23. **if** $i \leq N \wedge cluster[rank[i]] \neq 0$ **then**

24. $v_s = rank[i], curCluster++$

25. **end-if**

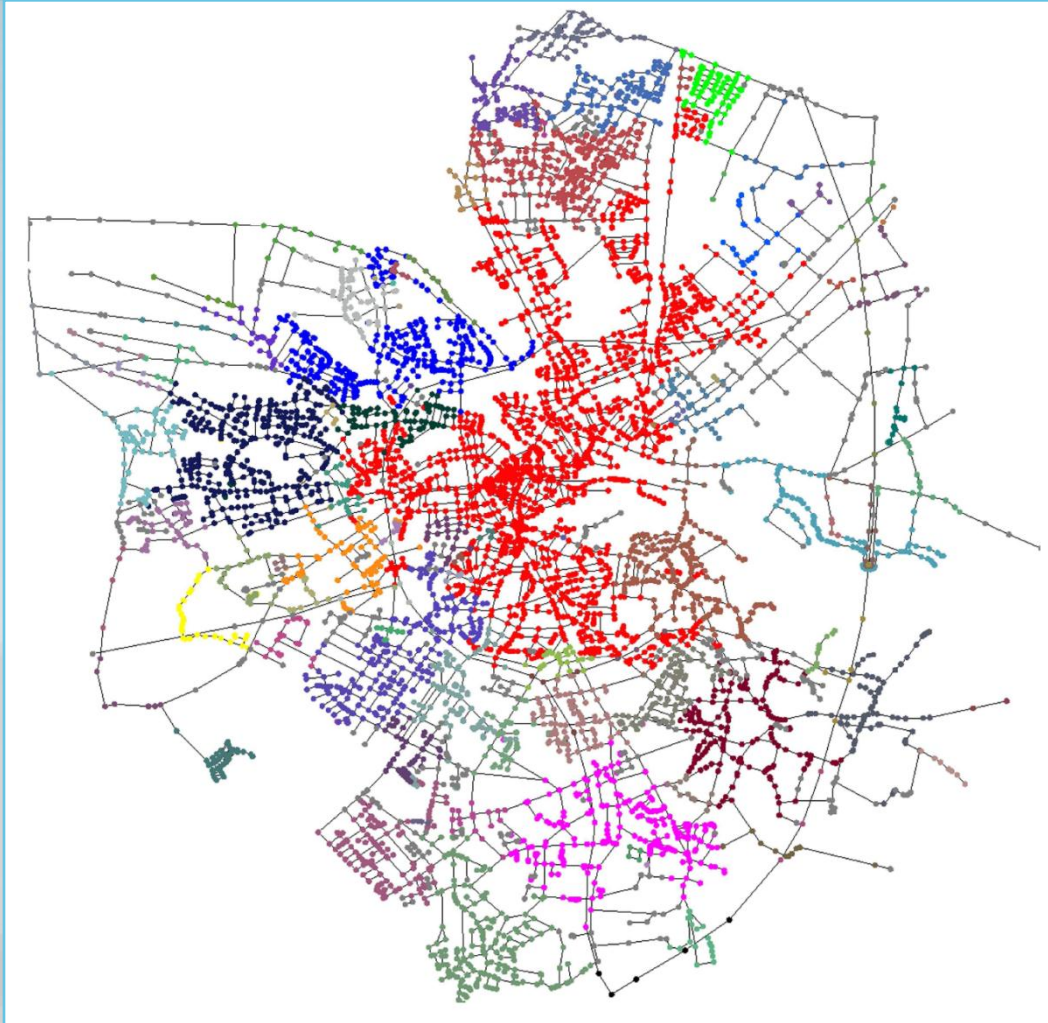
26. **end-while**

Cluster Construction

27. **return** $cluster[]$

Ομαδοποίηση Κόμβων Γράφου με Μεταβατική Ομοιότητα – Η ομαδοποίηση κόμβων του Oldenburg

- Δημιουργία 10.000 τροχιών με μήκη 10-100.
- Προβαθμολόγηση των κόμβων με βάση τη συχνότητα εμφάνισής τους στις τροχιές.
- Οριακή τιμή ομοιότητας $r = 0,005$.
- Εκτέλεση του αλγορίθμου.
- Ακρίβεια ομαδοποίησης: 80,54%.



Ιδέες για Μελλοντική Έρευνα στα θέματα της Ομοιότητας

- Μελέτη αλγορίθμων εύρεσης βέλτιστης τροχιάς όταν εμφανίζονται αστάθμητοι παράγοντες και αβεβαιότητα.
- Μελέτη αλγορίθμων εύρεσης βέλτιστης τροχιάς πάνω σε δυναμικά χωρικά δίκτυα, όπου δεν μπορούν να προϋπολογιστούν shortest path distances.
- Μελέτη προβλημάτων που αφορούν το συνδυασμό χωρικών και κοινωνικών δικτύων.
- Μελέτη της επεξεργασίας ερωτημάτων ομοιότητας με προσεγγίσεις.
- Μελέτη της ομαδοποίησης τροχιών από τα προτεινόμενα μέτρα (ανακάλυψη και επεξεργασία ομάδων).
- Η μελέτη των μεθόδων εκτίμησης του πλήθους των κόμβων και των ακμών σε σύνθετους τύπους ερωτημάτων.
- Η προσαρμογή του αλγορίθμου ομαδοποίησης κόμβων με στόχο την ανακάλυψη και επικαλυπτόμενων ομάδων.

A stylized, monochromatic illustration of a city skyline in shades of blue and grey. The buildings are simplified geometric shapes. In the center, the Greek word 'Τέλος' (The End) is written in a white, bold, sans-serif font with a slight drop shadow. The background is a gradient from light blue at the bottom to a darker blue at the top.

Τέλος