Mobility Data Management & Exploration

Epilogue

Yannis Theodoridis & Nikos Pelekis

InfoLab | University of Piraeus | Greece
infolab.cs.unipi.gr

Apr. 2013

Epilogue

- (Privacy-preserving) Mobility Data Acquisition, Querying, and Mining strives for a win-win situation
  - Obtaining the advantages of collective mobility knowledge without disclosing inadvertently any individual mobility knowledge.
- A word of wisdom:
  - ***solutions can only be obtained via an alliance of technology, legal regulations, and social norms*** (Rakesh Agrawal)
- Challenge: **ubiquitous computing**
  - flood of with new complex data (in a decentralized setting)
  - data miners have only begun digging by scratching the surface of this problem

Future trends

- Towards complex social networks of moving interacting objects:
  - **Movement** + **Interactions** + **Context**
- Future trend: **Mobility-aware social networks**
  - Facebook, Twitter, etc.: currently, 1 billion users of social media; what if their movement is added?
  - Second Life: movement / interactions / context in virtual life (= virtual projection of real life)
  - Foursquare, Gowalla, …

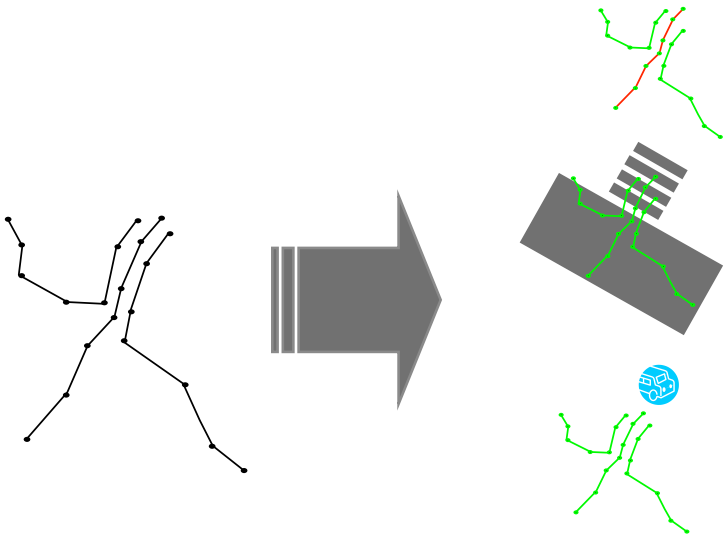
Mobility-aware social networks

- Mobility-aware social networks → discovering interaction patterns

Movement behaviour is a complex system of interactions.

Mobility-aware social networks

- Pedestrian movement = Interaction !!



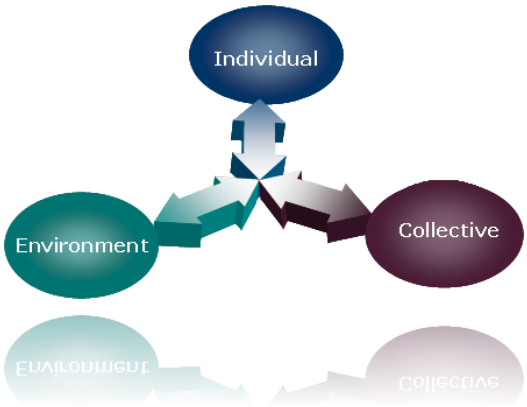
Leadership

Path following

Goal seek

Mobility-aware social networks

- Patterns (= events, behaviours) discovered by pedestrian movement
  - Encounter, Collision avoidance,
  - Attraction, Route choosing,
  - Obstacle avoidance, Suspension,
  - Lane formation, Flocking,
  - Moving cluster



- An interdisciplinary field
  - Computer Scientists + Geographers + Social Scientists + ...

To summarize …

A pot-pourri or research problems

At a glance

I. Mobility data management (storage and querying)
   a. *Generating GPS-alike trajectories by example*
   b. *Reconstructing network-constrained trajectories*
   c. *Reconstructing trajectories from hybrid outdoor/indoor positioning*

II. Mobility data exploration (OLAP analysis and data mining)
   a. *Sampling trajectory datasets*
   b. *Discovering maximal frequent sub-trajectories*
   c. *Detecting trajectory outliers (online)*

III. Mobility-aware applications
   a. *Mobile social networks and mobility data management challenges*
   b. *Predicting future location under network constraints*
   c. *Constrained routing*

IV. Privacy aspects
   a. *Comparing K-anonymity techniques for trajectories – the effect of 'K'*
   b. *Trajectory sampling/voting for anonymization purposes*
   c. *Disclosing trajectories from anonymized MODs*

Problem I.a.

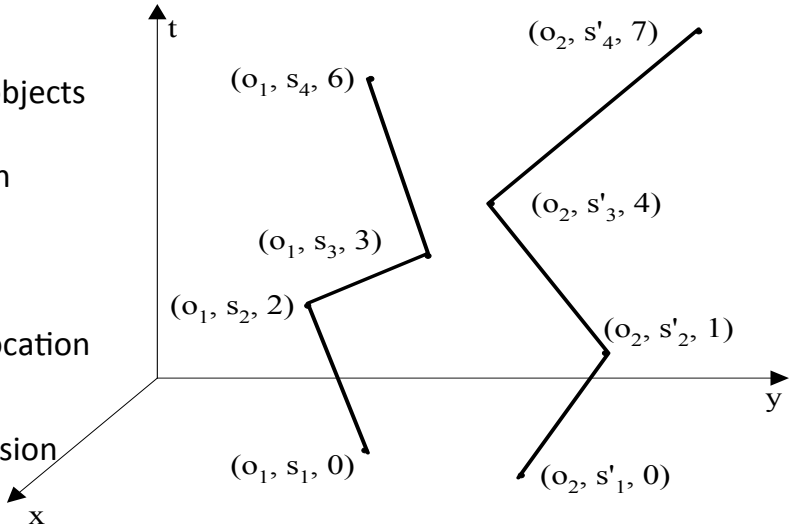
Generating GPS-alike trajectories by Example

- Goal
  - Assume a 'small' database D of real movements (trajectories) and (optional) a road network G(V,E). Our goal is to generate a 'large' database D' of synthetic movements following the 'behavior' or 'trend' of D.

- Reading list:
  - Trajectory generators: GSTD (Theodoridis & Nascimento, 2000), Brinkhoff's Network-constrained trajectory generator (Brinkhoff, 2002)
  - Pseudo-trajectories (Yin & Wolfson, 2003)

GSTD

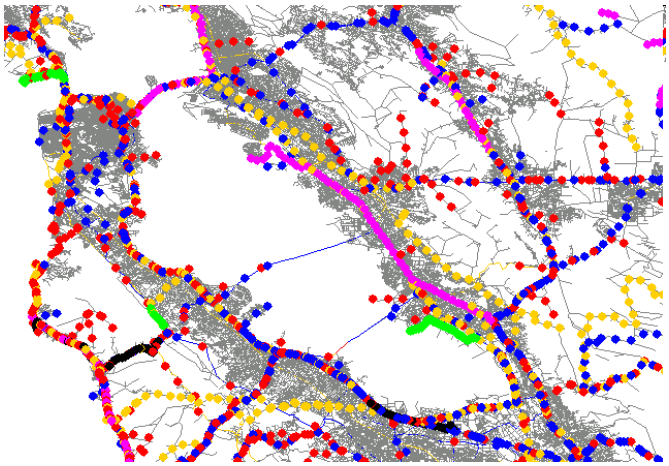
- Methodology
  - define a set of objects with a starting position for each
  - compute new timestamp
  - compute new location
  - compute new location's extension



Example: Two moving objects

Brinkhoff's generator

- Methodology:
  - generate starting points
  - generate length of route (depending on object class)
  - generate destination for each object
  - compute the route
  - compute the trajectory by generating a random speed every time unit
    - based on capacity, weather, edge class, etc.



source: www.fh-oow.de/institute/iapg/personen/brinkhoff/generator

Problem I.b.

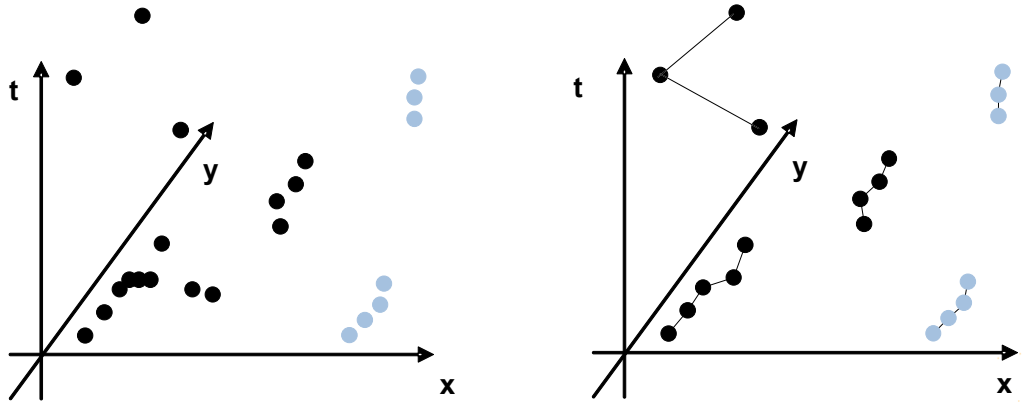
Reconstructing Network-Constrained Trajectories

- Goal
  - Assume a stream S of GPS recordings, a road network G(V,E), and a set C of constraints regarding movement (maximum speed, turn angle, etc.). Our goal is to reconstruct a database D of network-constrained trajectories from S under the constraints enforced by G and C.
    - c.f. the "trajectory refresh" operation discussed in the "04-apps" lecture

- Reading list:
  - Trajectory reconstruction (Marketos et al. 2008)
  - Trajectory map-matching (Brakatsoulas et al. 2005) (Kellaris et al. 2009)
  - Network-constrained trajectory generator (Brinkhoff, 2002)

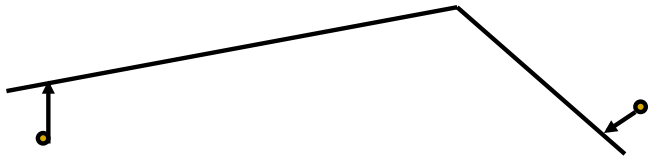
Trajectory reconstruction

- [Marketos et al. 2008] A number of filters / thresholds decide whether the new series of data is to
  - be appended to an existing trajectory, or
  - initiate a new trajectory, or
  - be considered as noise

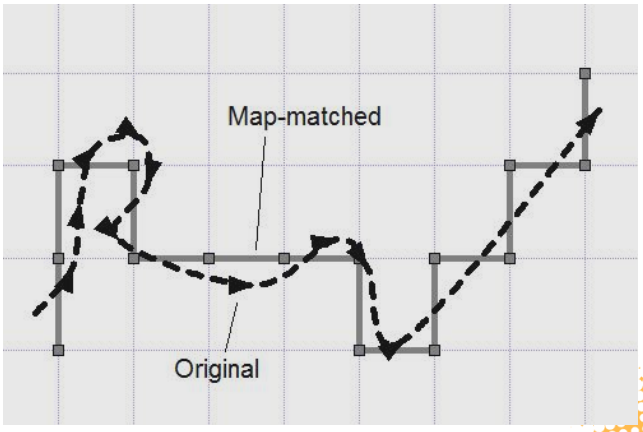


Trajectory map-matching

- The main idea:
  - "Snap" GPS points on road network
  - Find shortest path on map between consecutive GPS points



- State-of-the-art algorithm
  - Map-Matching (Brakatsoulas et al. 2005)



Problem I.c.

Reconstructing Trajectories from hybrid Outdoor/Indoor Positioning

- Goal
  - Assume an object moving outdoor (e.g. home-to-campus) and indoor (e.g. in the campus), and traced by GPS and indoor positioning devices (WiFi access points and RFID readers), respectively. Our goal is to generate reconstruct the entire trajectory of the object, based on the above recordings.

- Reading list:
  - Indoor positioning trajectory management (Jensen et al. 2009) (Yang et al. 2009) (Jensen et al. 2010)
  - MIT GeoBLOG project (Nabian et al. 2009)
  - Trajectory reconstruction (Marketos et al. 2008)

MIT GeoBLOG project

- MIT campus, 4000 wifi access points (Nabian et al. 2009)

source: http://senseable.mit.edu/geoblog/

Problem II.a.

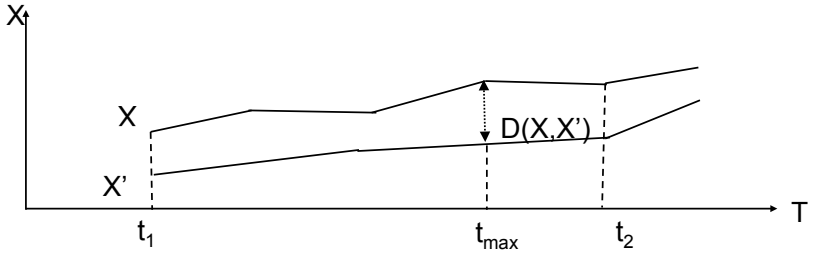
Sampling Trajectory Datasets

- Goal
  - Assume a historical trajectory database D and a target percentage p (0 < p < 1). Our goal is to find an 'appropriate' subset D' of D of cardinality |D'|=p|D| to be used as a sample dataset.
    - Issue: what is 'appropriateness' in our case?

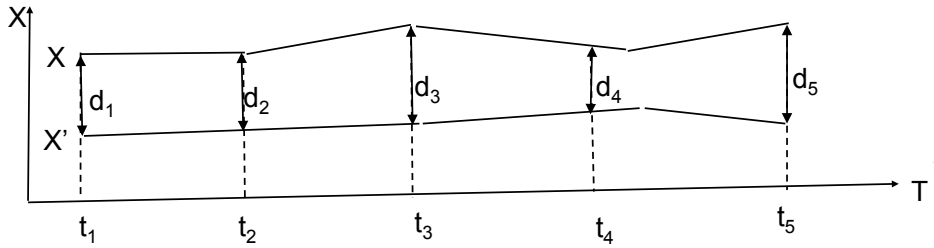
- Reading list:
  - Trajectory voting (Panagiotakis et al. 2009, 2011) and sampling (Pelekis et al. 2010)
  - Trajectory similarity (Vlachos et al. 2002) (Tiakas et al. 2006) (Frentzos et al. 2007) (Pelekis et al. 2007, 2011)

Trajectory similarity

- Maximum (or Minimum) Euclidean distance

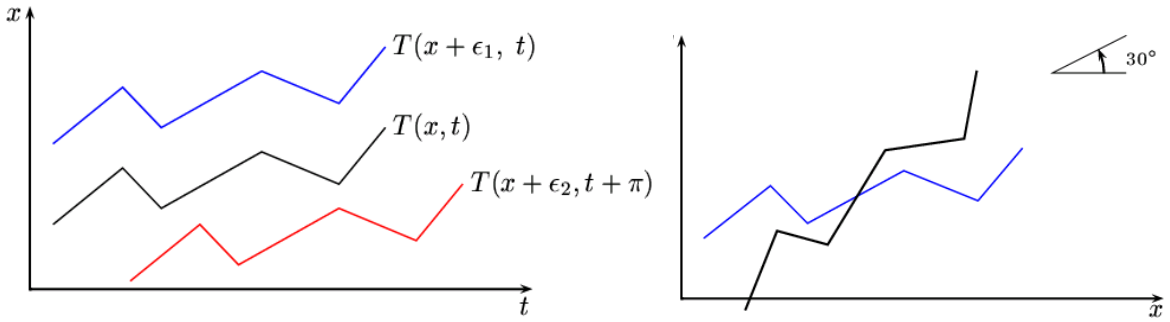


- Root Mean Error Euclidean Distance
  - $D(\mathbf{X},\mathbf{X'}) = (1/5*(d_1^{2+} d_2^{2+} d_3^{2+} d_4^{2+} d_5^{2}))^{1/2}$



Trajectory similarity

- Geometric transformations: translation, rotation



- Translation, rotation are rigid transformations.
  - Example: similar motion patterns in different places / time periods

Problem II.b.

Discovering Maximal Frequent Sub-trajectories in MODs

- Goal
  - Assume a road network G(V,E), a historical trajectory database D, and a support threshold s. Our goal is to find the most frequent paths in terms of maximal sub-trajectories in D.

- Reading list:
  - Mining maximal frequent itemsets (Burdick et al. 2001)
  - Mining frequent trajectories (Morzy 2007) and T-patterns (Giannotti et al. 2007)

Problem II.c.

Detecting Trajectory Outliers (online)

- Goal
  - Assume a historical trajectory database D and a set L of current locations of moving objects. Our goal is to find the sets L' and D' of current anomalies and their respective maximal "noisy" sub-trajectories in D.
  - E.g. application in vessel AIS databases

- Reading list:
  - Trajectory outlier detection (Lee et al. 2008) (Ying et al. 2009)
  - Online discovery of hot motion paths (Sacharidis et al. 2008)

Problem III.a.

Mobile Social Networks and Data Management Challenges

- Goal
  - To study in depth the behaviour of popular mobile social networking apps (Google Latitude, Facebook Places, Foursquare, etc.) and
    - Outline the data management techniques that are incorporated there
    - Suggest variations of these apps towards the "moving reference – moving data objects" concept

- Reading list:
  - Material from respective URLs

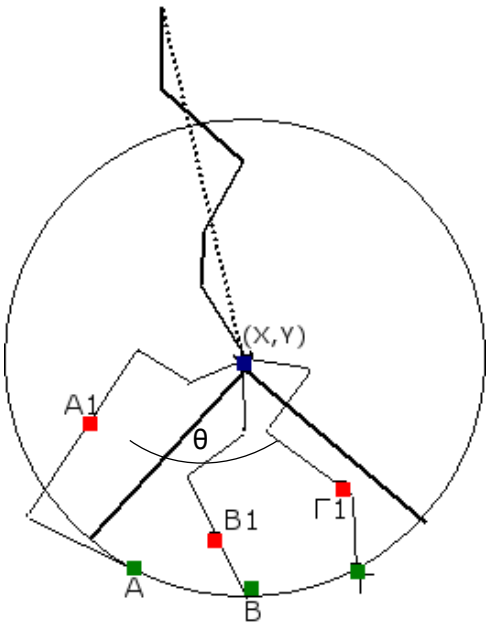
Problem III.b.

Predicting Future Location under network constraints

- Goal
  - Assume a road network G(V,E), the recent motion history (trajectory) T and current instant velocity v of a user, a temporal duration dt, and (optional) a historical trajectory database D of other users' movements. Our goal is to estimate user's location after dt.
- Variations / applications: "Get-together", "Find-me", etc.
  - Study an efficient behaviour of the 'target' (goal: to avoid its hunters) vs. of the hunters (goal: to catch the target; collaborative work)
- Reading list:
  - Location prediction for LBS (Gowrisankar & Nittel, 2002), (Hassan & Xiong, 2003)
  - Pattern-based Location prediction (Elnekave et al. 2007) (Monreale et al. 2009)
  - Network-constrained trajectory generator (Brinkhoff, 2002)

Predicting future location

- A naïve approach:
  - Current location (X, Y), Visibility angle θ, current speed v, imply:
  - a set S (A, B, Γ) in Euclidean terms
  - A set S1 (B1, Γ1) in Network-distance terms
  - We choose the average or the median in S1
- WhereNext (Monreale et al. 2009):
  - finds T-patterns from a trajectory dataset
  - evaluates T-patterns and select the best
  - builds the prediction model



$$R_0 \xrightarrow{\alpha_1} R_1 \xrightarrow{\alpha_2} R_2 \xrightarrow{\alpha_3} R_3$$

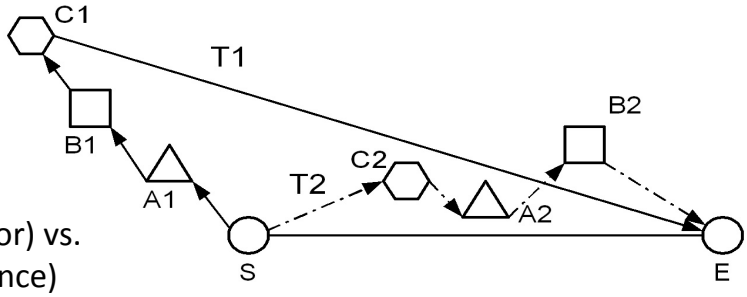
Problem III.c.

Constrained Routing

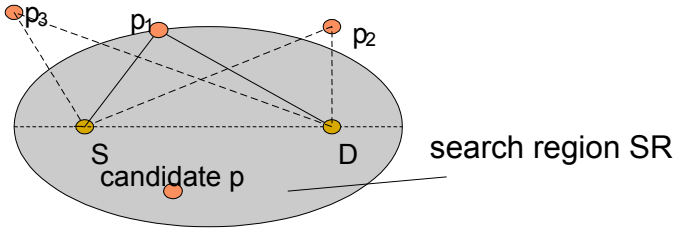
- Goal
  - Assume a road network G(V,E), a pair of start and end locations, A and B, a set of Points of Interest (POI), and a number k > 0. Our goal is to find the optimal route from A to B constrained to pass from k POIs.
- Variation:
  - temporal/ordering constraints between passes from k POIs.
- Reading list:
  - Trip planning queries (Li et al. 2005)
  - Network-constrained trajectory generator (Brinkhoff, 2002)

Trip-planning query

- According to (Li et al. 2005):
  - TPQ is harder than TSP !!
  - NN (Nearest Neighbor) vs. MD (Minimum Distance) algorithms
  - MD restricts search space as an ellipse



source: http://www.cs.fsu.edu/~lifeifei/tpq.html



Problem IV.a.

Comparing K- Anonymity Techniques for Trajectories – the effect of 'K'

- Goal
  - Assume a historical trajectory database D and an anonymization threshold k. Our goal is to compare the performance of different trajectory anonymization techniques (NWA, AWO, W4M) in terms of efficiency and effectiveness wrt k.

- Reading list:
  - Trajectory K- Anonymization (Abul et al. 2008, 2010) (Nergiz et al. 2008)
  - Network-constrained trajectory generator (Brinkhoff, 2002)

Problem IV.b.

Trajectory Sampling/Voting for Anonymization Purposes

- Goal
  - Assume a historical trajectory database D and an anonymization threshold k. Our goal is to introduce trajectory sampling techniques in order to provide a sanitized (anonymized) version of D.

- Reading list:
  - Trajectory K- Anonymization (Abul et al. 2008, 2010) (Nergiz et al. 2008)
  - Trajectory voting (Panagiotakis et al. 2009)
  - Network-constrained trajectory generator (Brinkhoff, 2002)

Problem IV.c.

Disclosing Trajectories from Anonymized MODs

- Goal
  - Assume a sanitized (anonymized) version D' of a historical trajectory database D. Our goal is to re-identify trajectories based on probabilistic techniques and approach as much as possible the original D.

- Reading list:
  - Trajectory K- Anonymization (Abul et al. 2008) (Nergiz et al. 2008)
  - Trajectory reconstruction (Marketos et al. 2008)
  - Pattern-based Location prediction (Monreale et al. 2009)

A few more problems (1/2)

- trajectory simplification under constraints (the vessels case study)
  - reduce the size of a trajectory taking its parameters (heading, speed, etc.) as well as constraints (e.g. obstacles) into consideration. evaluation criteria: quality vs. size reduction.

- optimizing motion pattern queries:
  - implement related work and search for extensions/variations, experiment with synthetic/real datasets*

* Available @ www.chorochronos.org

A few more problems (2/2)

- Network-constrained TDW:
  - Define a star (?) schema. Think of possible dimensions, measures and OLAP scenarios.
- Semantic-aware TDW
  - As above …
- Semantic-aware MOD
  - What about storage, indexing and query processing
- Moving clusters vs. flocks vs. convoys:
  - Evaluate their interrelationships. Compare their efficiency.

Questions

Acknowledgments

- The content of this lecture series has been inspired by collaborative work done in the following EU projects:
  - FP7/SEEK (http://www.seek-project.eu), 2012-15
  - FP7/DATASIM (http://www.datasim-fp7.eu), 2011-14
  - ESF/COST-MOVE (http://move-cost.info), 2009-13
  - FP7/MODAP (http://www.modap.org), 2009-12
  - FP6/GeoPKDD (http://www.geopkdd.eu), 2005-09
- Also, special thanks to InfoLab members

Thank you !!!

*When you set out on your **journey** to Ithaca,*

pray that the road is long,

*full of adventure, full of **knowledge** ...*

C. Cavafy, "Ithaca" (1911)