CS548 Spring 2015 Clustering II

Showcase by
Michael Barry, Cheng Deng, Junwei Guan, Xing Liu, Robert Van Reenen

Showcasing work by
James Biagioni and Jacob Eriksson from the University of Illinois at Chicago
and Jia Qiu, Ruisheng Wang, and Xin Wang from the University of Calgary

on **Inferring Road Networks from GPS Traces**

References

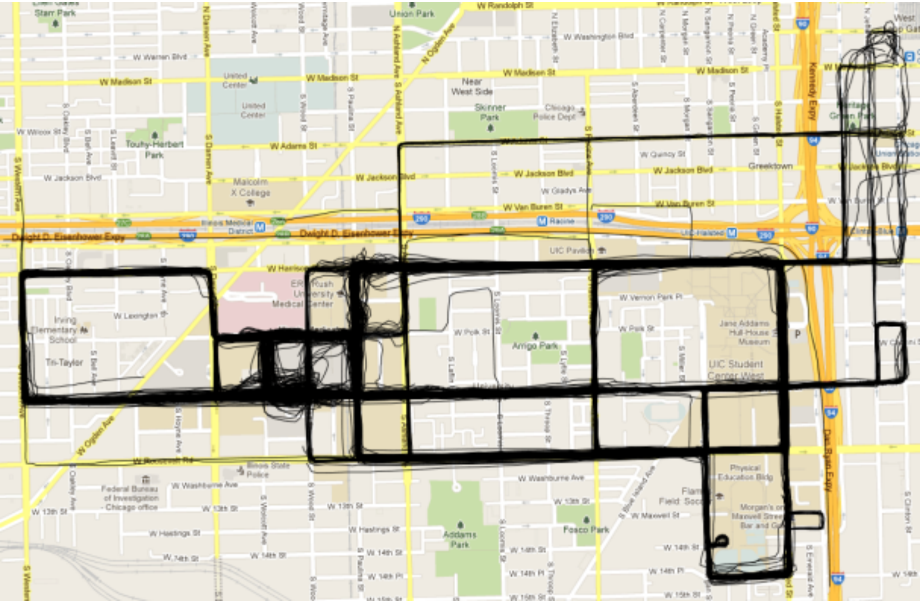
J. Biagioni and J. Eriksson, “Inferring Road Maps from GPS Traces: Survey and Comparative Evaluation,” in *91st Annual Meeting of the Transportation Research Board*, 2012.

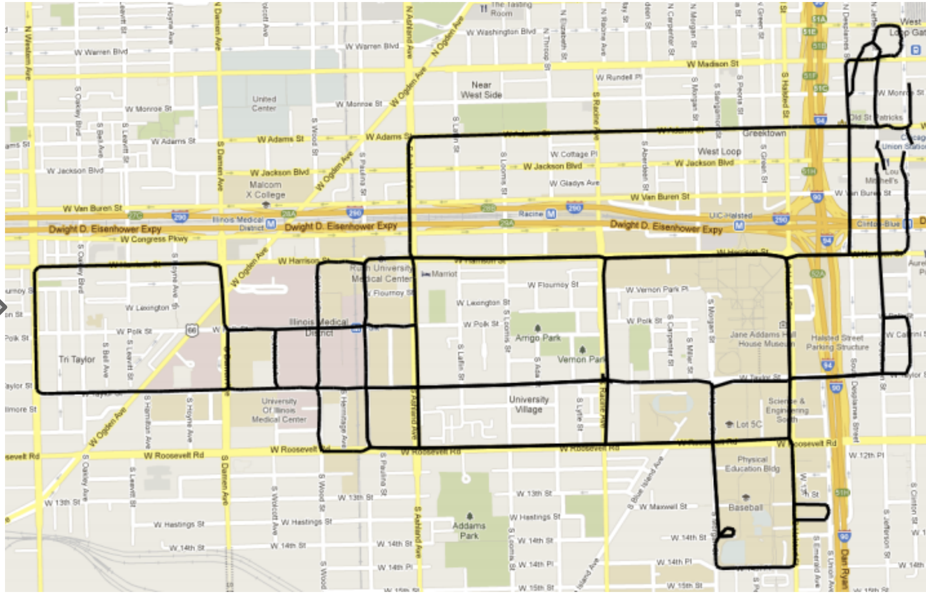
J. Biagioni and J. Eriksson, “Map Inference in the Face of Noise and Disparity,” In *SIGSPATIAL GIS*, *ACM*, 2012, pp. 79-88.

J. Qiu, R. Wang et al., “Inferring Road Maps from Sparsely-Sampled GPS Traces,” In *Canadian Conference on AI*, 2014, pp. 339-344.

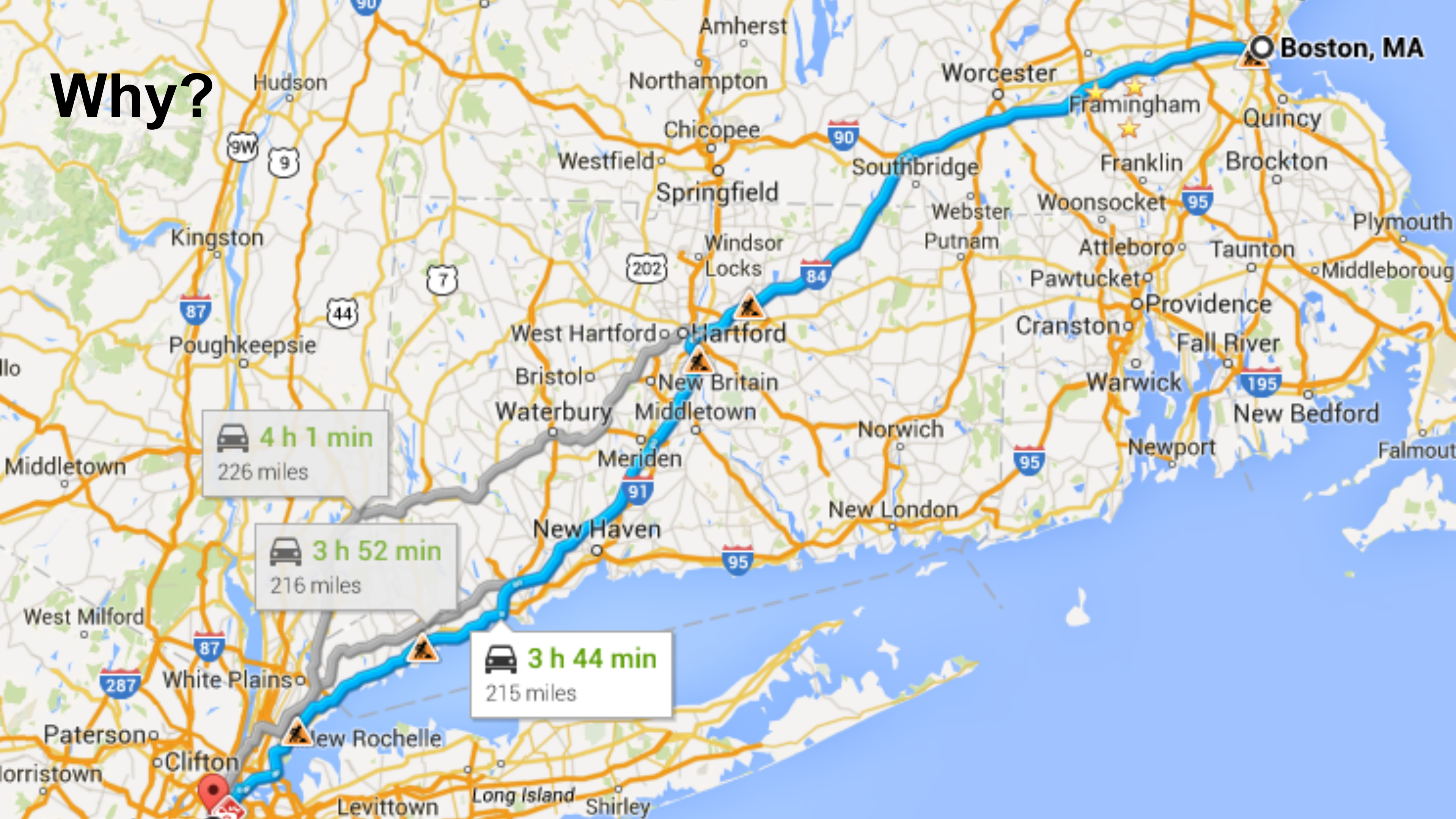
T. Duong, “An introduction to kernel density estimation,” http://www.mvstat.net/tduong/research/seminars/seminar-2001-05/

Inferring Road Networks from GPS Traces





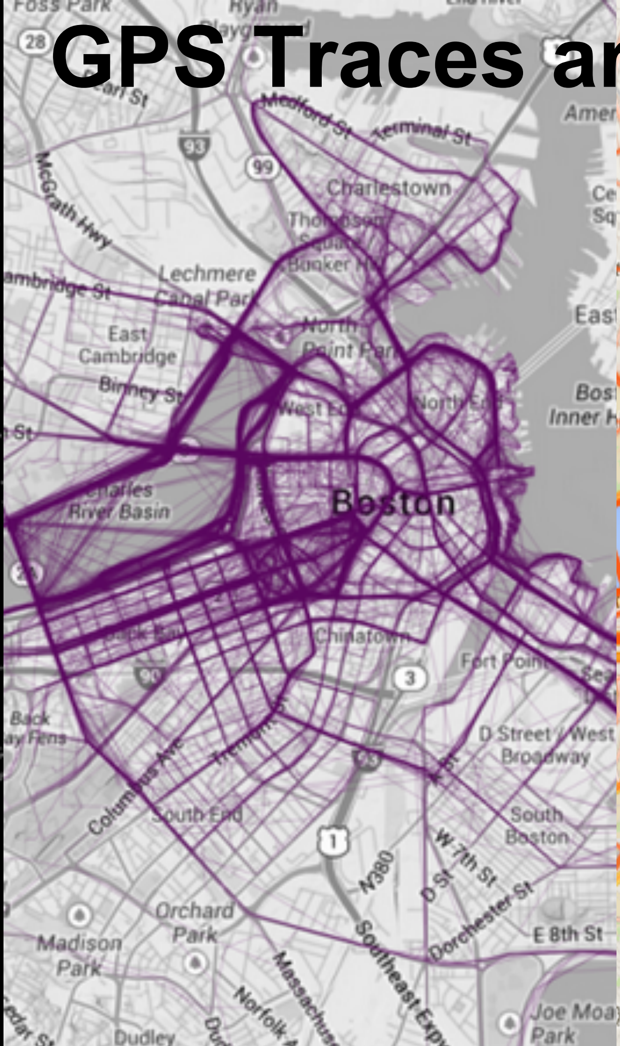
Why?



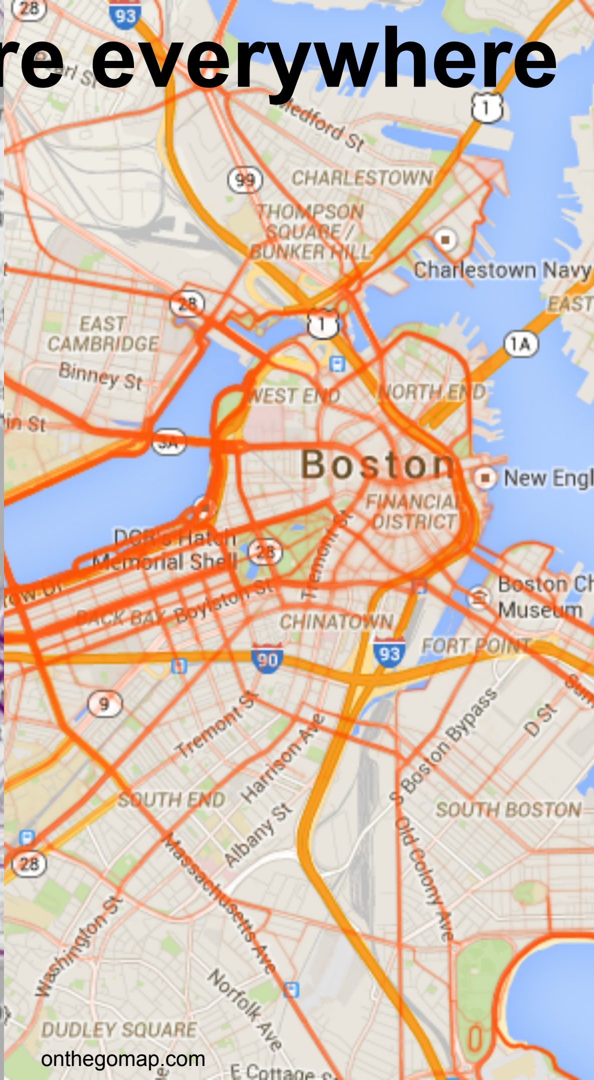
Boston, MA
4 h 1 min
226 miles
3 h 52 min
216 miles
3 h 44 min
215 miles
Amherst
Hudson
Northampton
Worcester
Framingham
Quincy
Chicopee
Westfield
Springfield
Southbridge
Franklin
Brockton
Webster
Woonsocket
Plymouth
Kingston
Windsor Locks
Putnam
Attleboro
Taunton
Pawtucket
Middleborough
Providence
Poughkeepsie
West Hartford
Hartford
Cranston
Fall River
Bristol
New Britain
Warwick
Waterbury
Middletown
New Bedford
Norwich
Meriden
Newport
Falmouth
Middletown
New London
New Haven
West Milford
White Plains
Paterson
Clifton
New Rochelle
Morristown
Long Island
Levittown
Shirley

GPS Traces are everywhere

cities.human.co/posts/boston.html

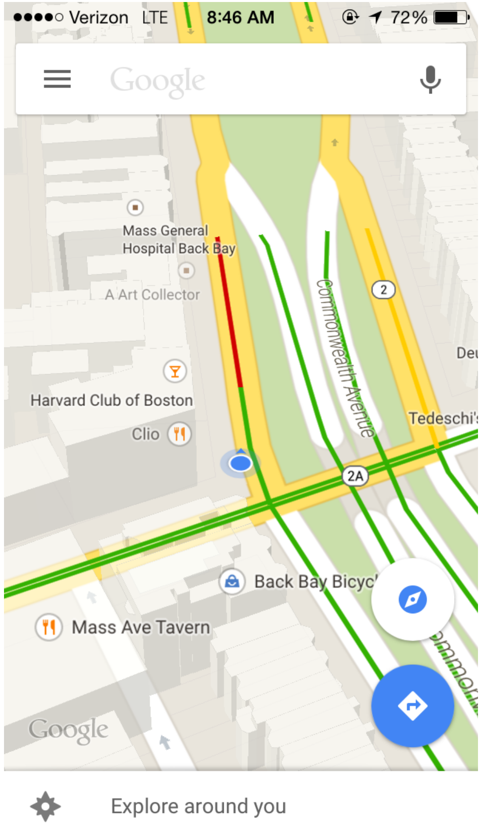


flowingdata.com/2014/02/05/where-people-run/

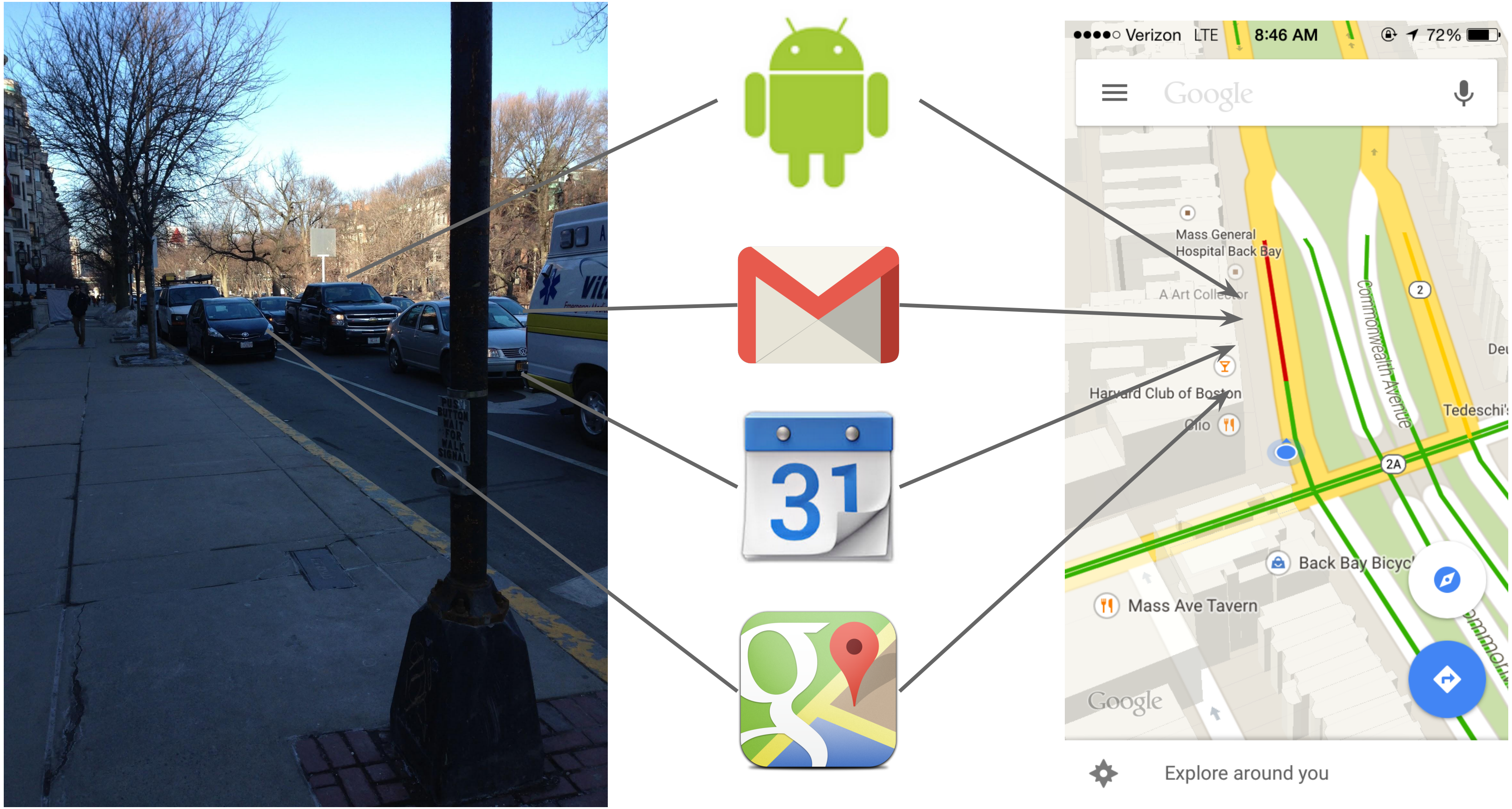


onthegomap.com

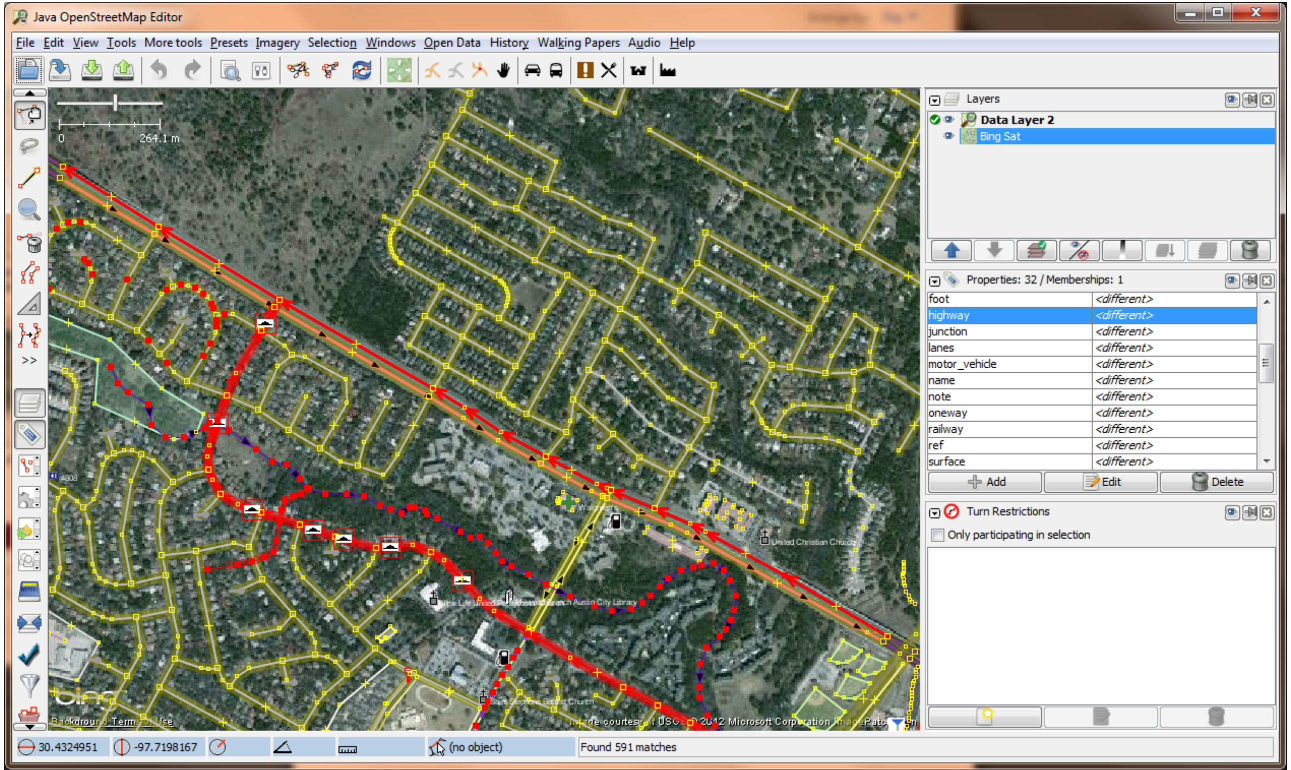
Google uses them for traffic



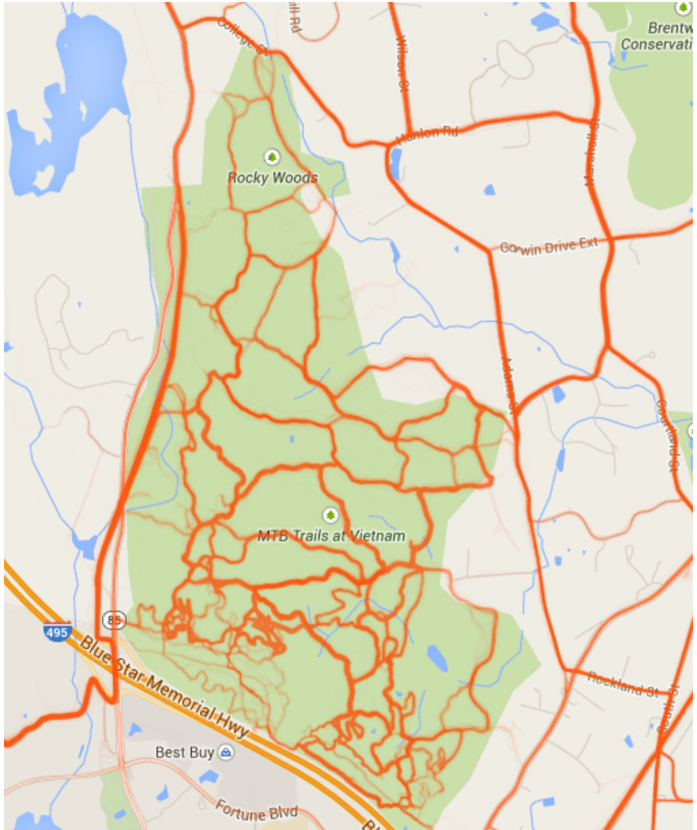
Google uses them for traffic

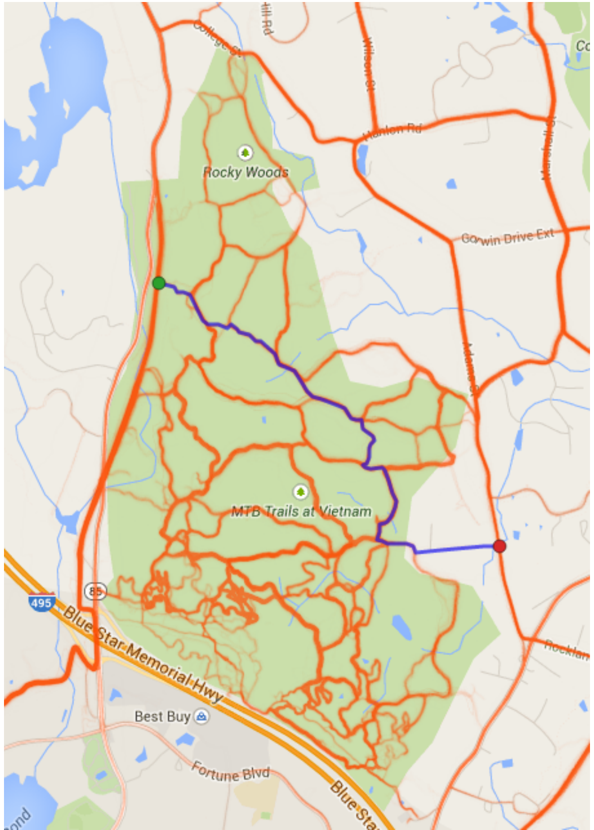


OpenStreetMap lets users make updates



But we can’t do this yet…





Papers Surveyed

2010 2011 2012 2013 2014 2015

2011 Biagioni and Eriksson survey techniques used since 2000

2012 Biagioni and Eriksson devise their own hybrid pipeline to improve upon current state of the art - test on GPS traces from vehicles.

2014 Qui, et al. create their own pipeline, focusing on sparser, opportunistically collected traces from mobile devices

Existing Map Inference Papers

Paper	Year	Class
Edelkamp & Schrodl	2003	**k-means**
Schroedl et.al.	2004	**k-means**
Davies et. al.	2006	**KDE**
Worral & Nebot	2007	**k-means**
Guo, Iwamura & Koga	2007	**k-means**
Chen & Cheng	2008	**KDE**
Niehofer et al.	2009	**trace merge**

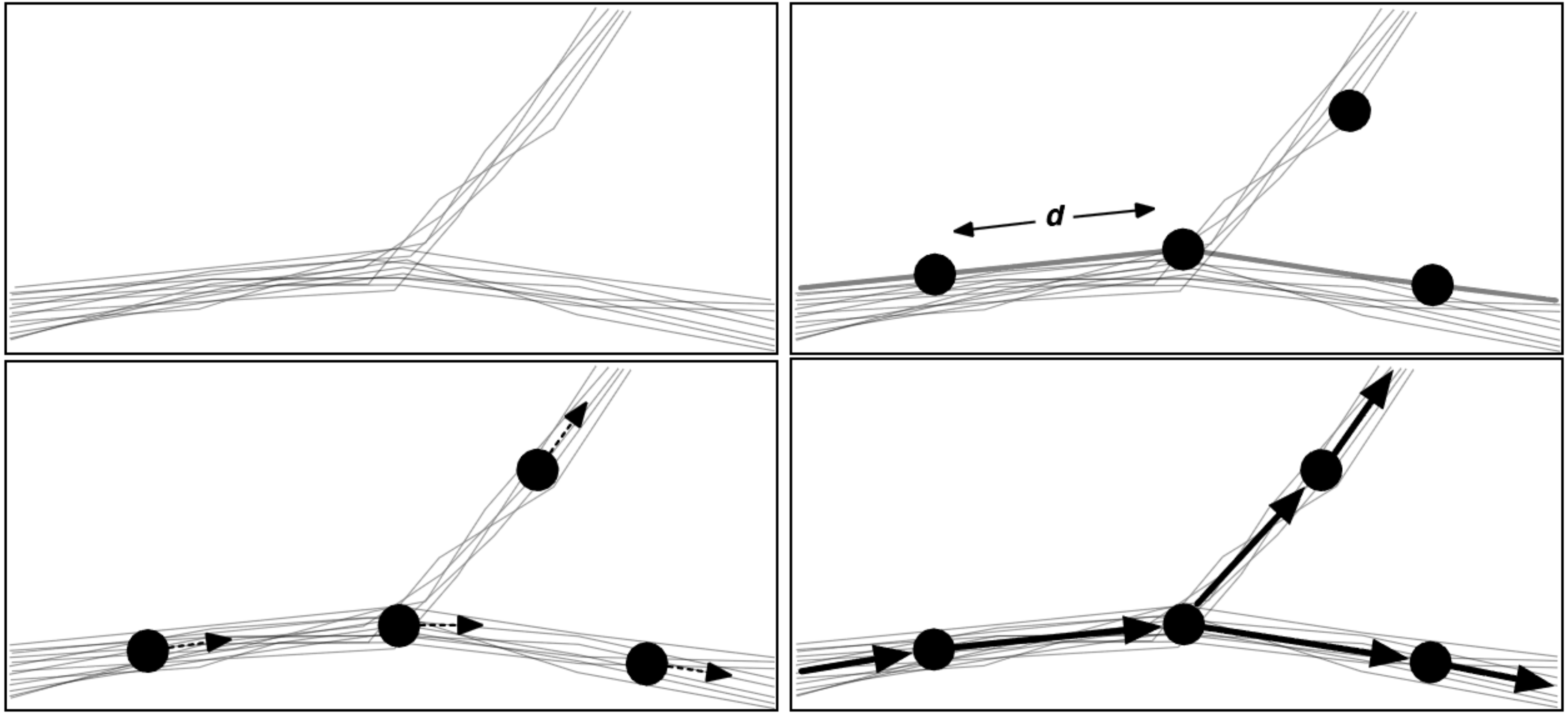
Paper	Year	Class
Cao & Krumm	2009	**trace merge**
Shi, Shen & Liu	2009	**KDE**
Jang, Kim & Lee	2010	**k-means**
Agamennoni et al.	2011	**k-means**
Biagioni & Eriksson	2012	**hybrid**
Qiu et al.	2014	**DBSCAN**

Summary of Existing Literature

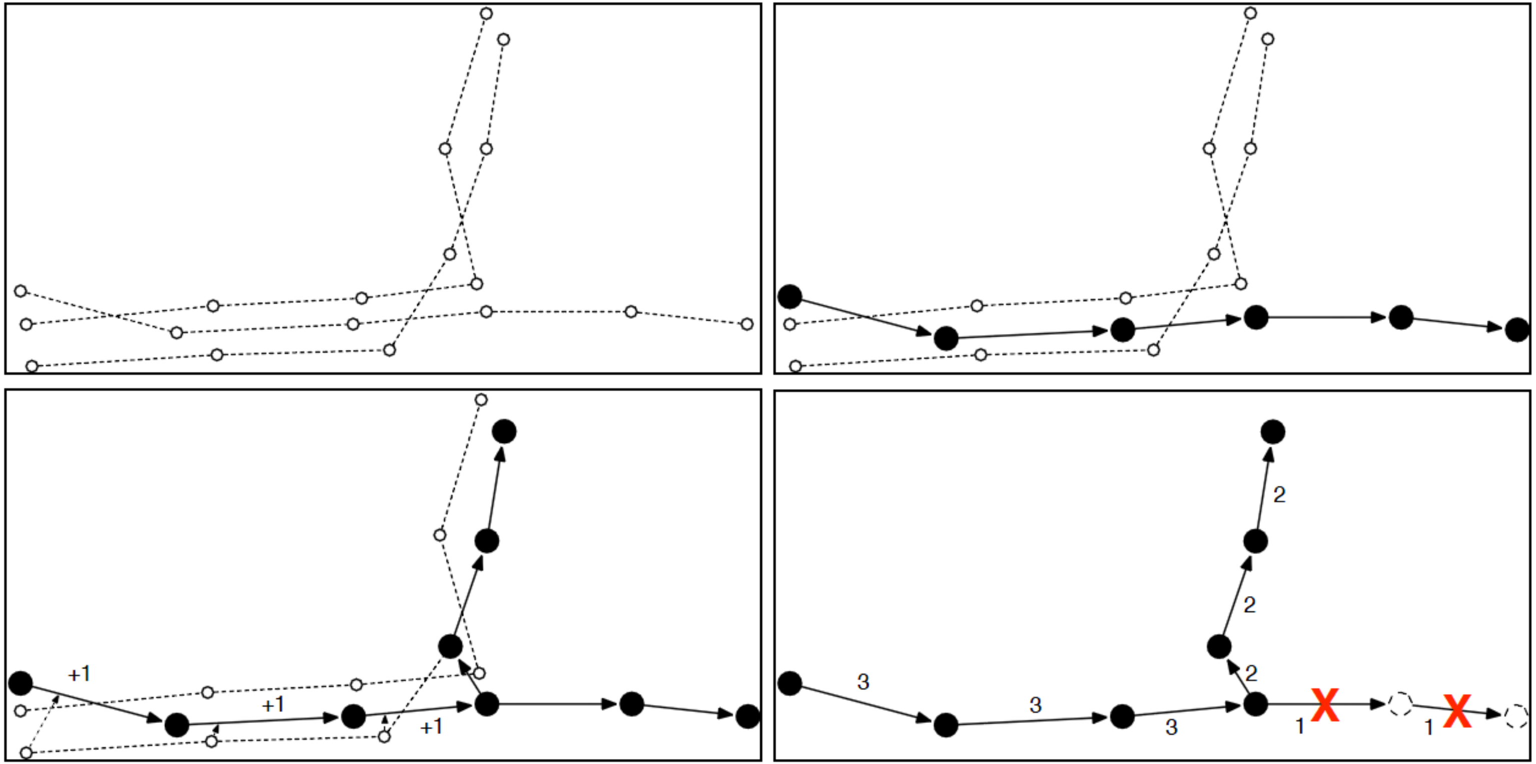
Out of the 13 papers on automatic map generation

- K-means based methods (6)
- Density estimation methods (4)
- Trace merging methods (2)
- Hybrid (1)

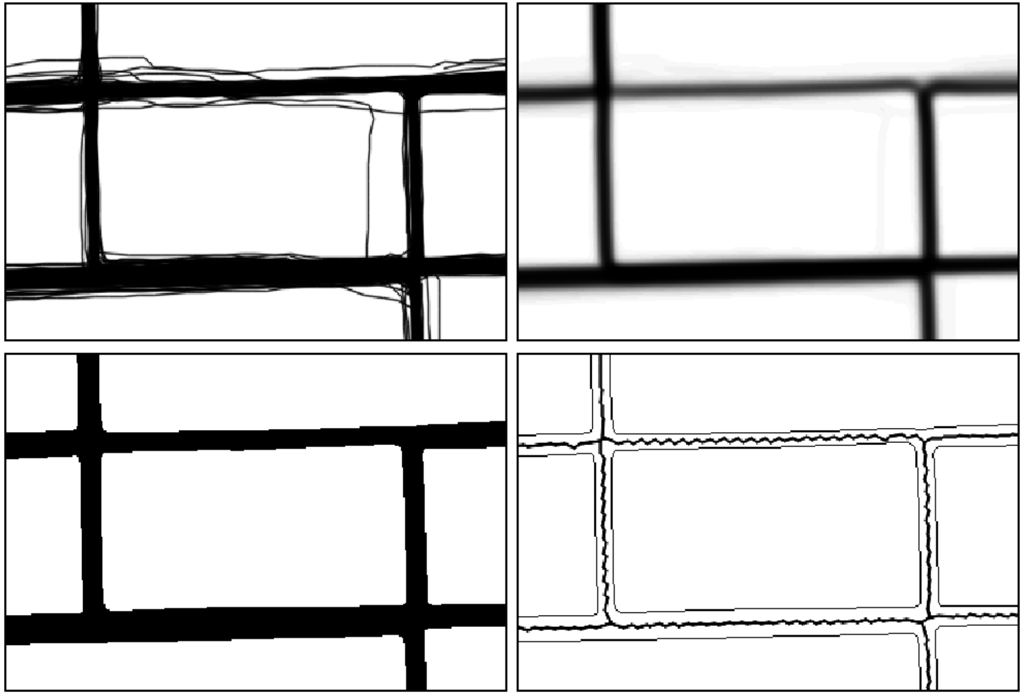
Map Inference (K-means Algorithm)



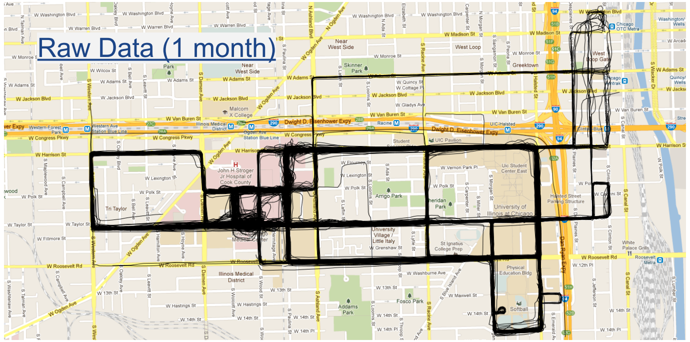
Map Inference (Trace Merging)

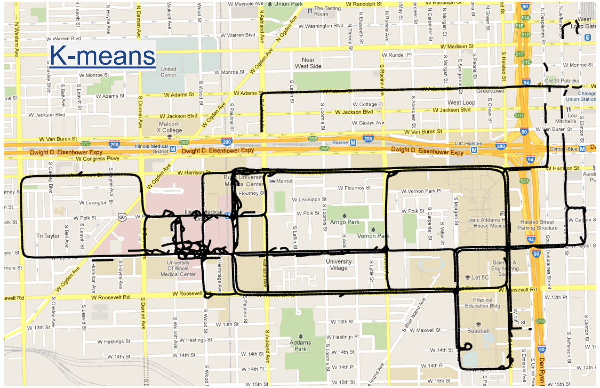


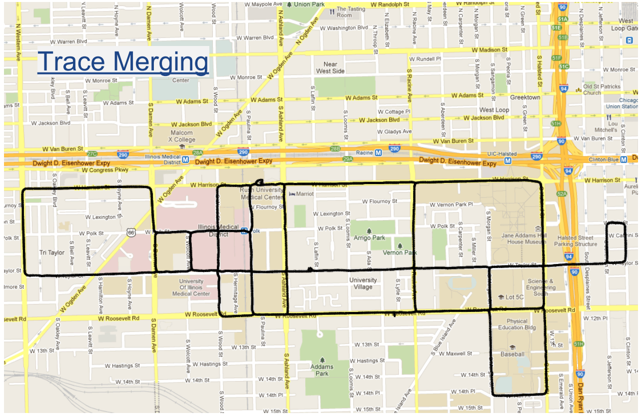
Map Inference (Density Estimation)

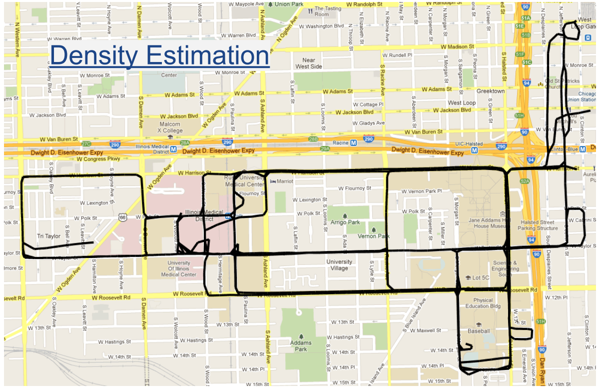


Visual illustration of each algorithm

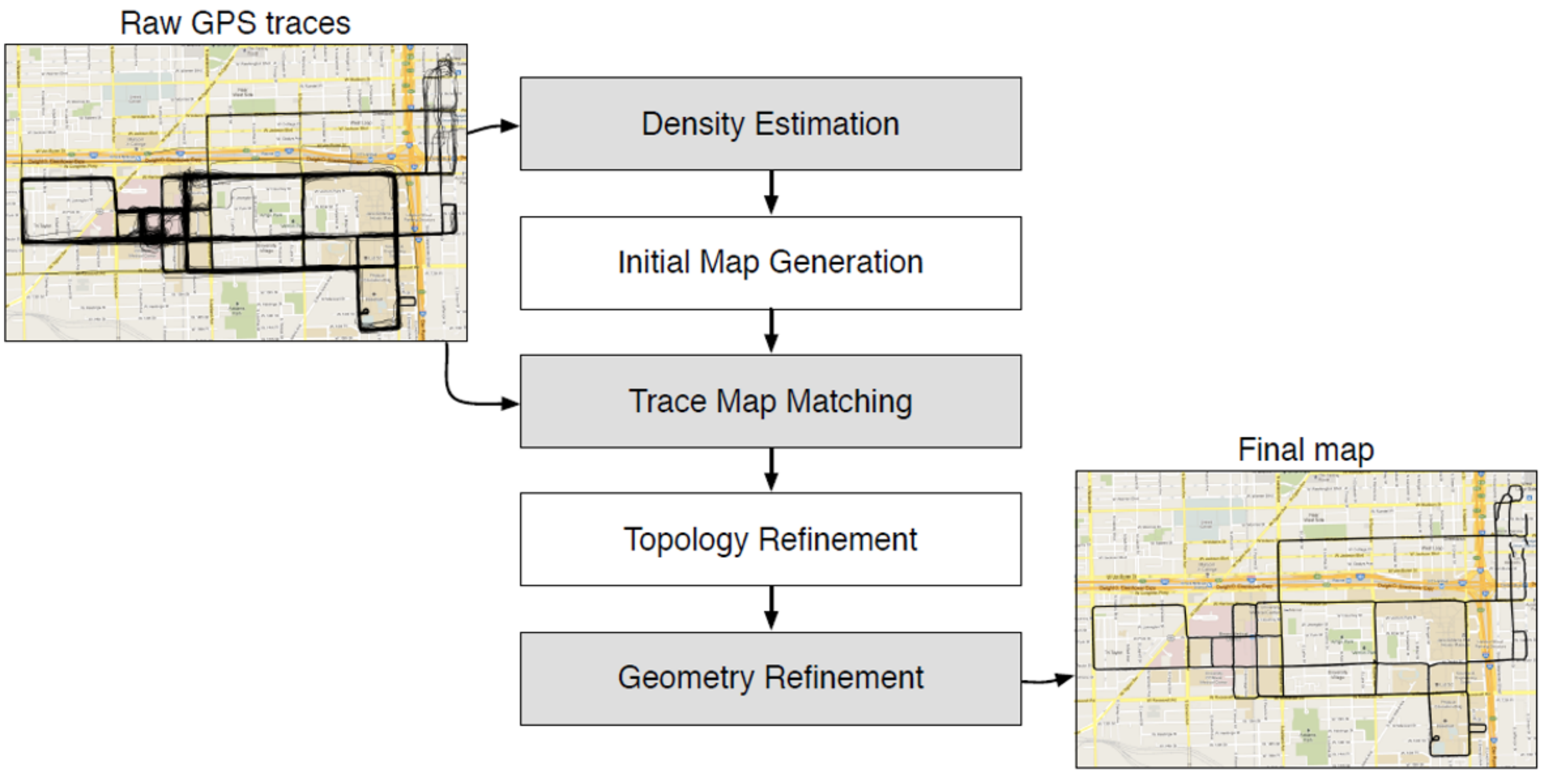








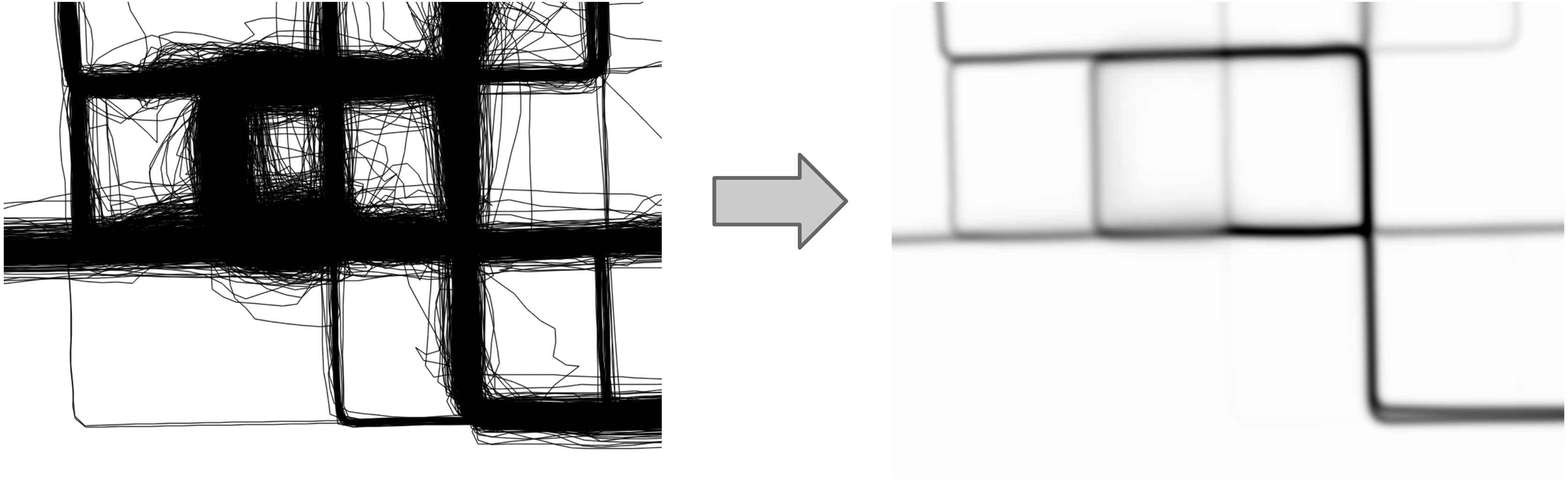
A Hybrid Map Inference Pipeline



Density Estimation

Raw GPS traces

Density estimate

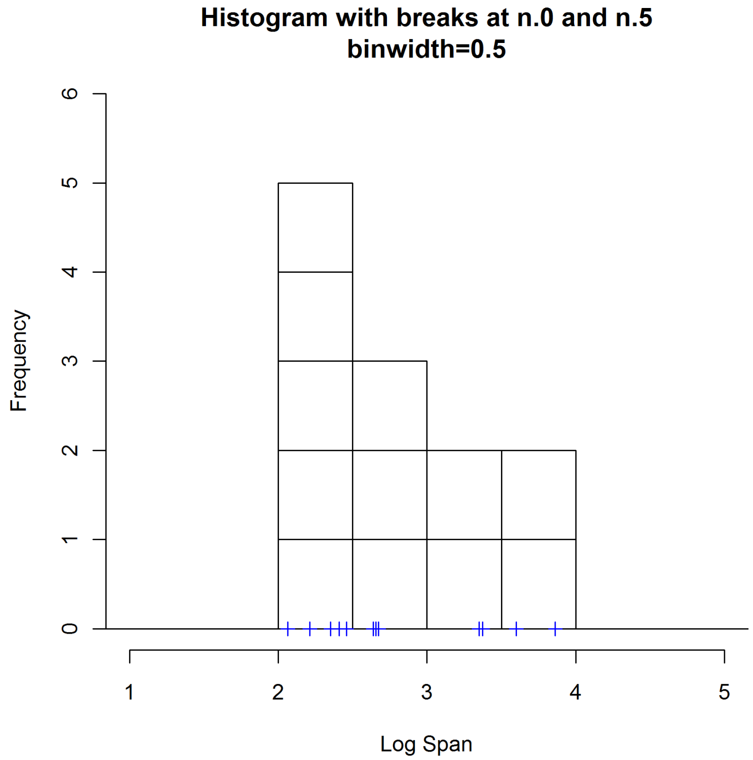


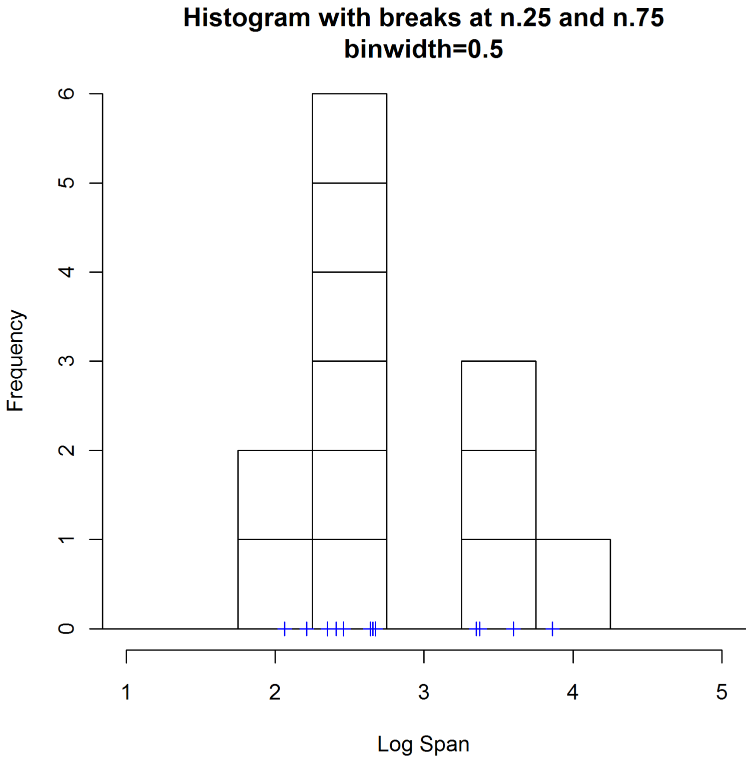
Density Estimation

- Bin trace data into 2D histogram
  - 1x1 m cells

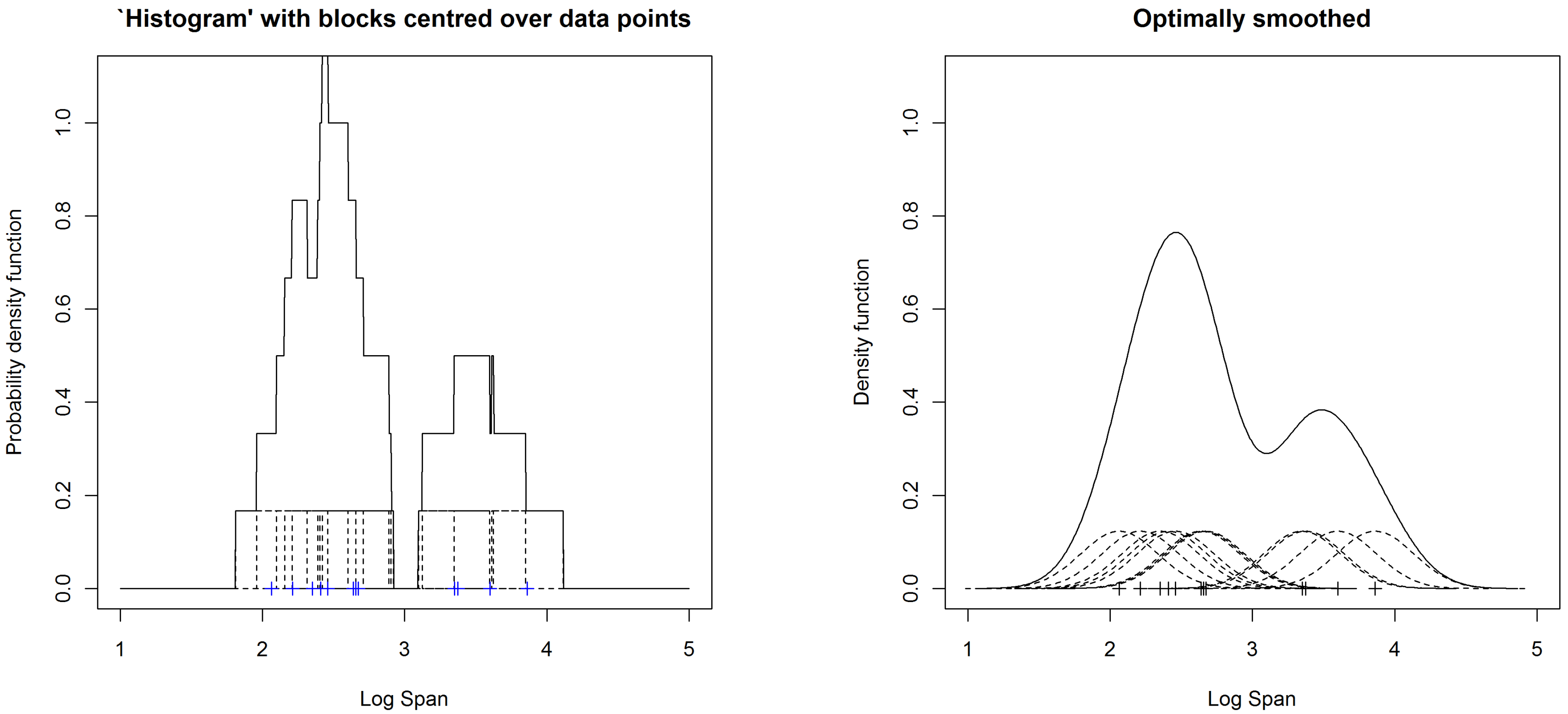
- Calculate Kernel Density Estimate (KDE)
  - Approximated by convolution with Gaussian
  - $N(0, \sigma^2)$, $\sigma = 8.5$ m
    - based on GPS error and road width

Why (not) histograms



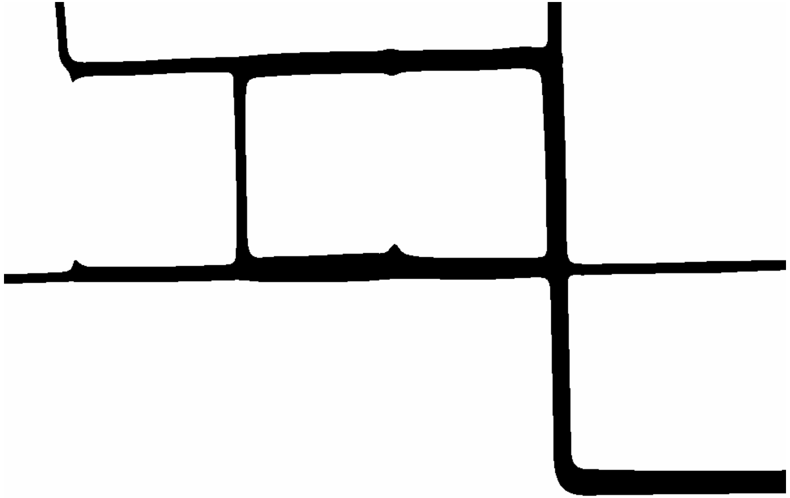


Kernel Density Estimate (KDE)

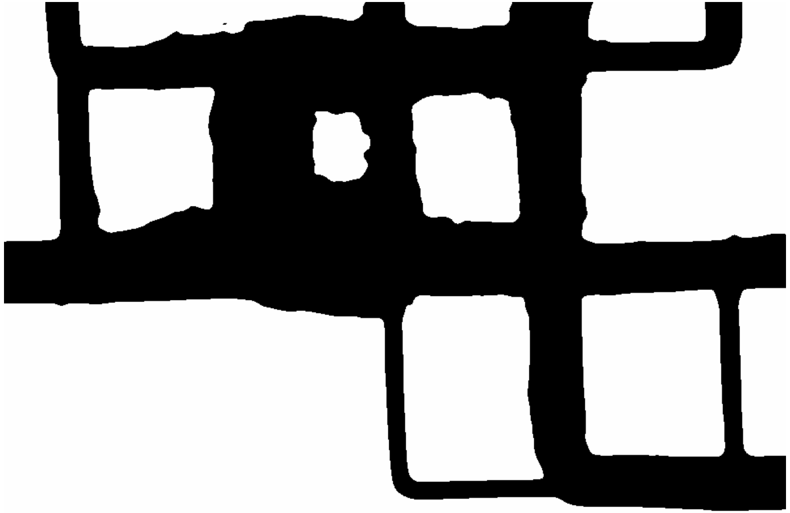


KDE Thresholds

[Applying high threshold to KDE]



[Applying low threshold to KDE]



Neither high nor low thresholds produce good results.
Biagioni and Eriksson used skeletonization instead.

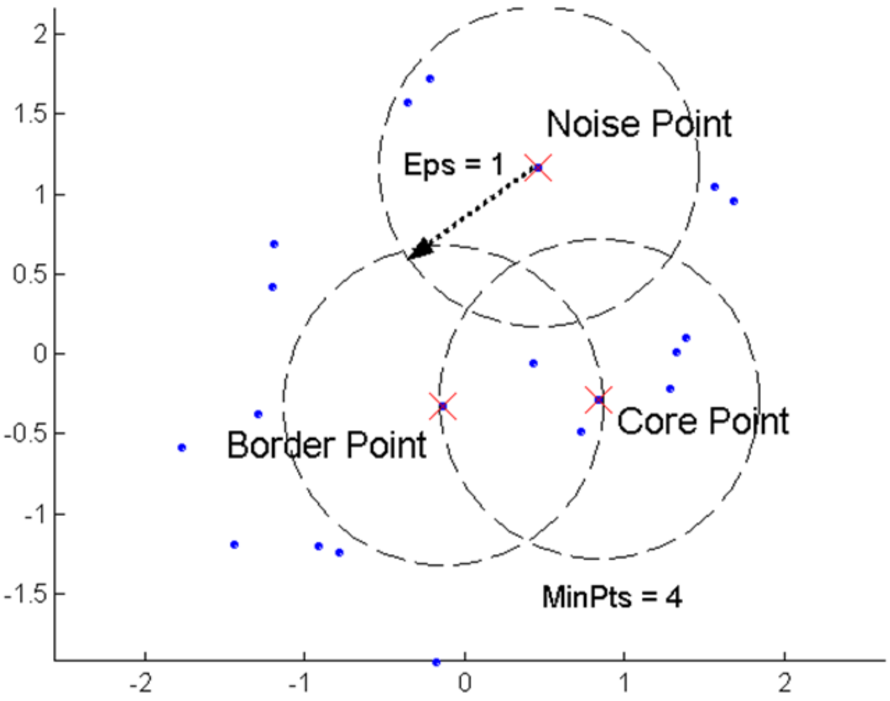
DBSCAN with Orientation Constraint

- Define neighborhood with 2 components:
  - ε - distance from center
  - α - difference of orientation (angle)

Definition 1: The $\varepsilon-$*neighborhood with orientation constraint* of a point p, denote by N_p, is defined by $N_p = \{q \in P | dis(p,q) \leq \varepsilon \wedge diffAngle(p,q) \leq \alpha\}$, where P is the whole set of points, $dis(p,q)$ is the Euclidian distance between point p and q, and $diffAngle(p,q)$ is the orientation difference of point p and q.

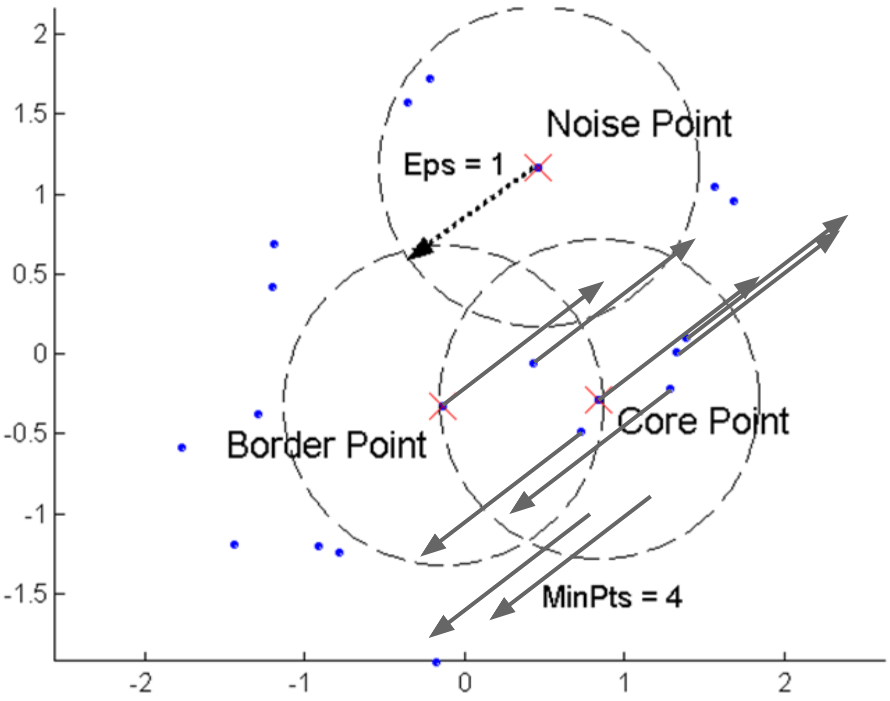
- MinPts = 2, ε = 15 m, α = 1°

DBSCAN



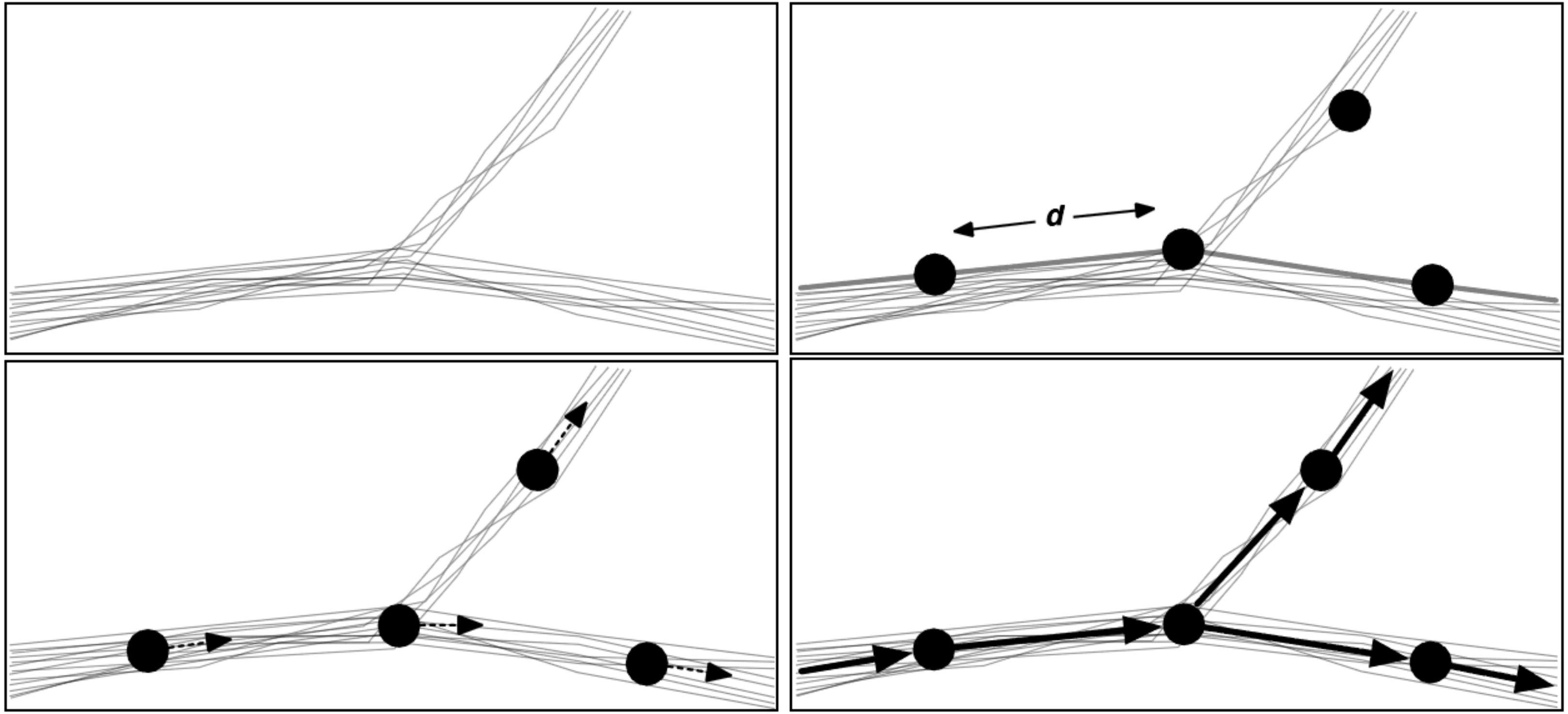
Noise Point
Eps = 1
Border Point
Core Point
MinPts = 4
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-1
0
1
2

DBSCAN with Orientation Constraint



With $\alpha = 1°$, each cluster represents a nearly straight road segment!

Map Inference (K-means Algorithm)



Map Inference (K-means Algorithm)

Nearly Straight Curve Reconstruction

1. Assign seeds of the point cluster

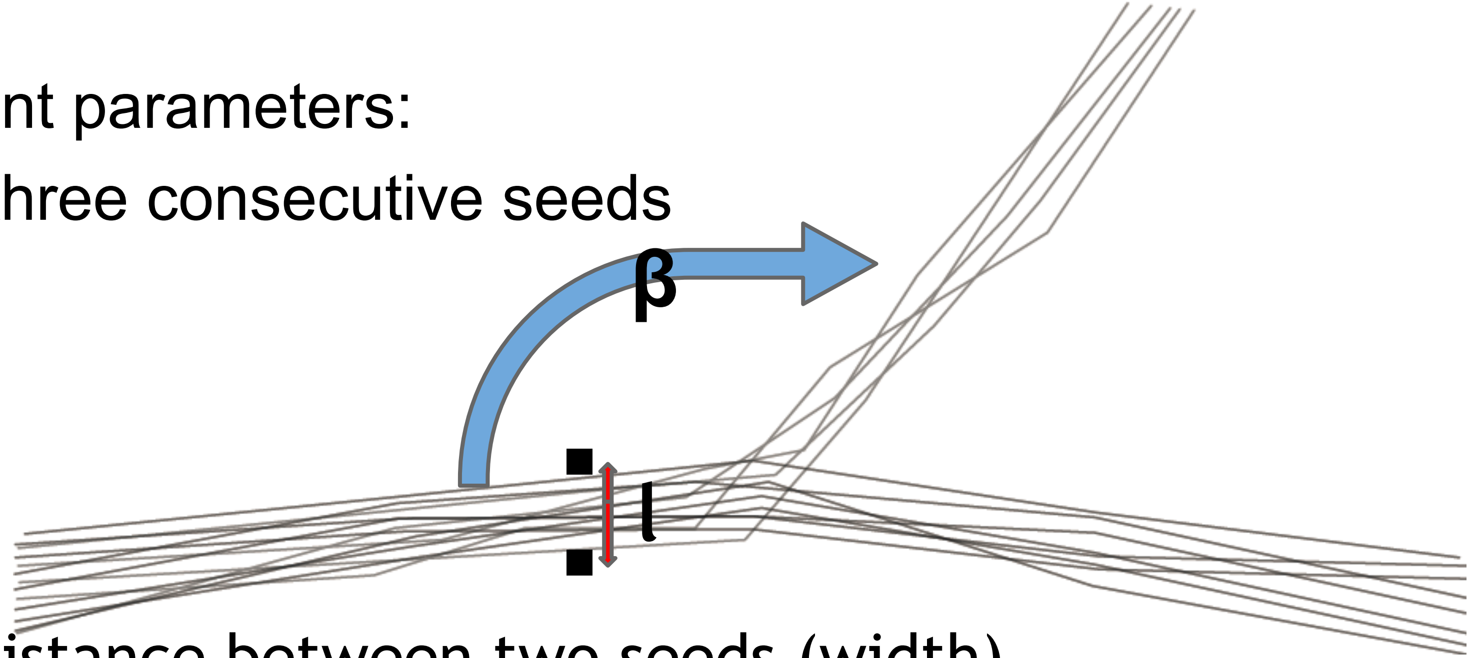
2. Adjust the seeds by K-means algorithms

3. Judge the measurement and repeat step 2 until it fulfill the threshold in the end

Map Inference (K-means Algorithm)

Measurement parameters:

β: angle of three consecutive seeds



l: interval distance between two seeds (width)

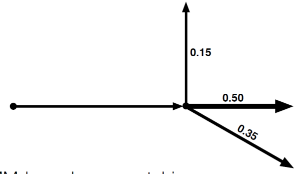
Map Inference (K-means Algorithm)

prefered range or value for the parameters:

β: [105°,150°] (In general β = 120°)
l: 5 meters is desirable

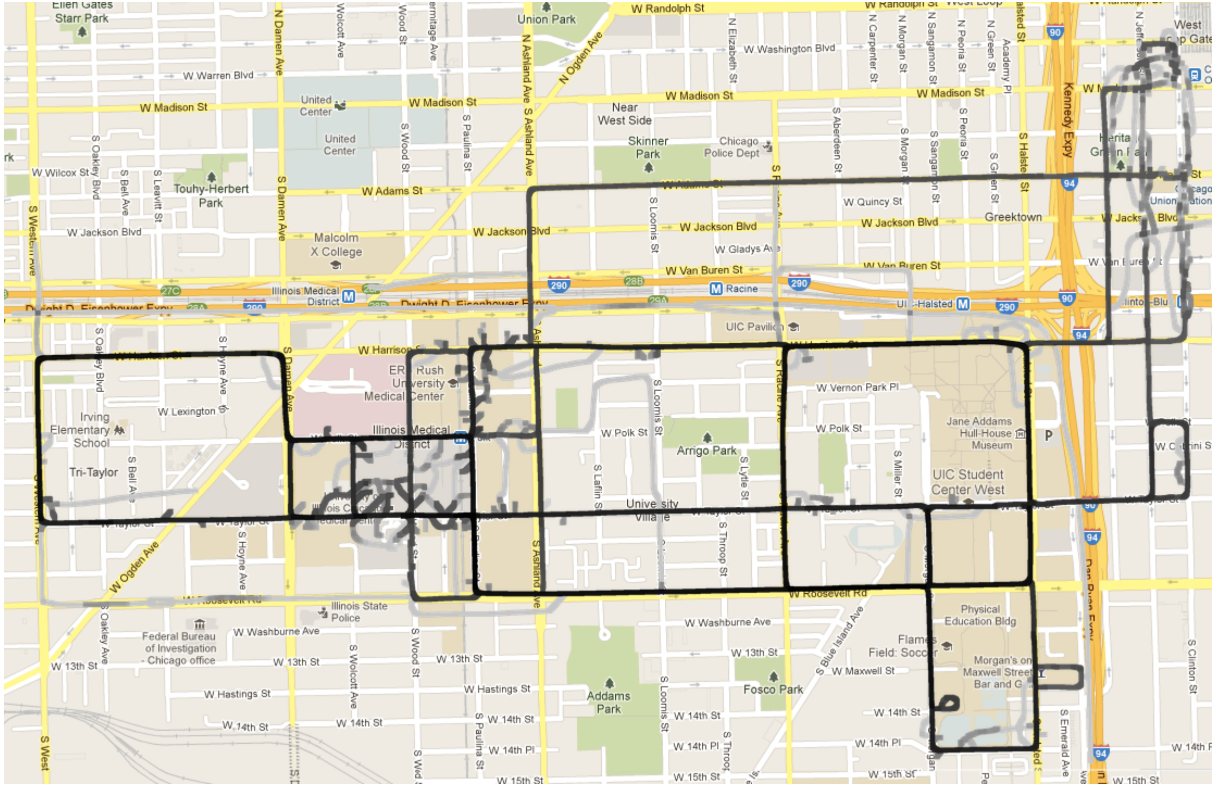
In the end:
set the sequenced seeds as “**centerline**”

Weighted transition probabilities



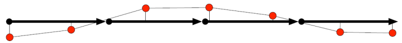
- HMM-based map-matching
  - Based on VTrack (Thiagarajan et al., 2009)

Initial Generated Map



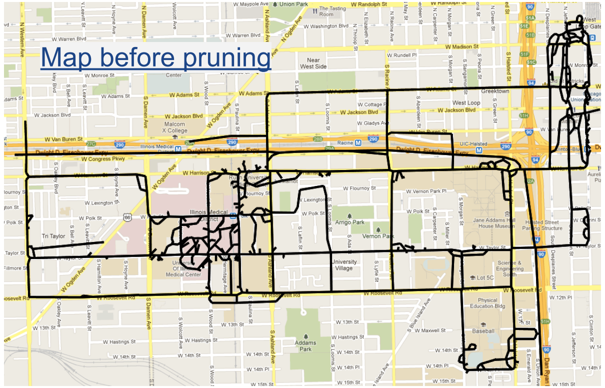
Trace goodness of fit

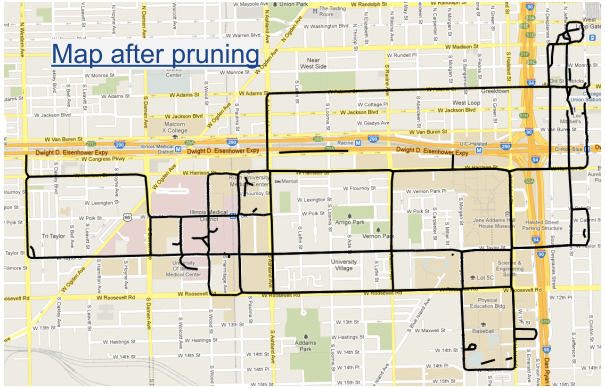
Topology Refinement

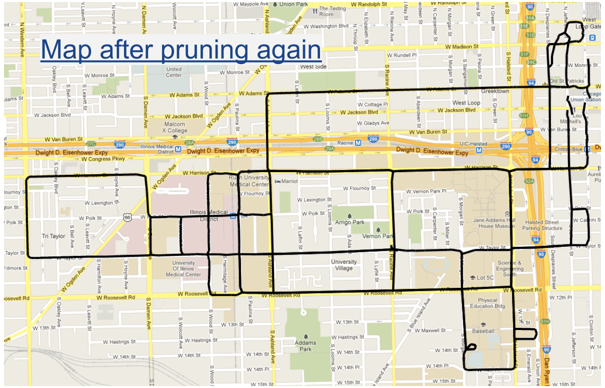


$$RMSD(\tau, e) = \sqrt{\frac{1}{|\tau|}\sum_{p \in \tau} dist(p, e)^2}$$

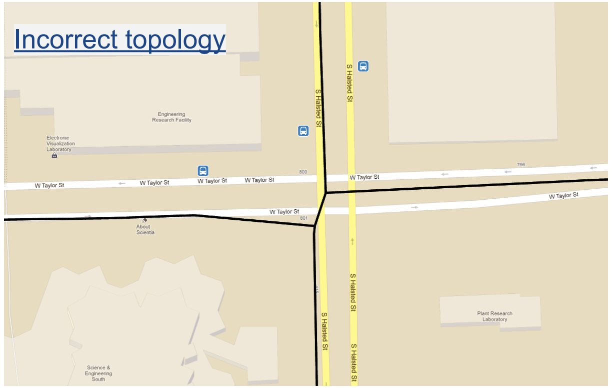
$$RMSD(\tau, e) < RMSD_{max}$$

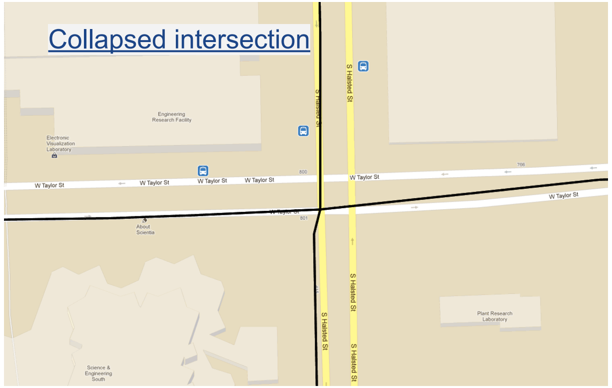


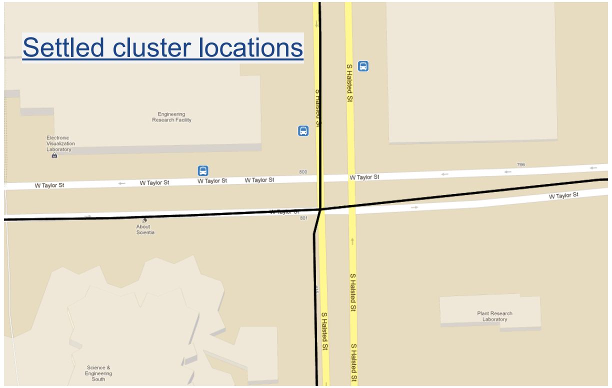


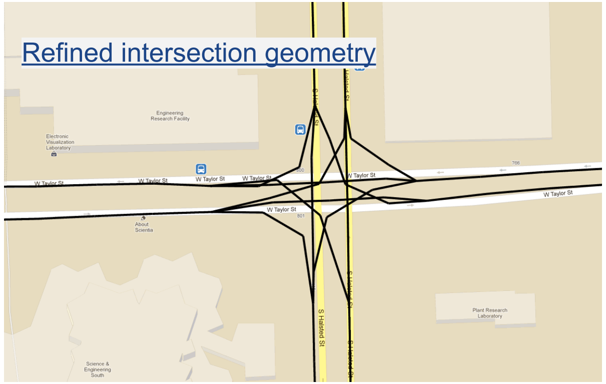


Topology & Geometry Refinement

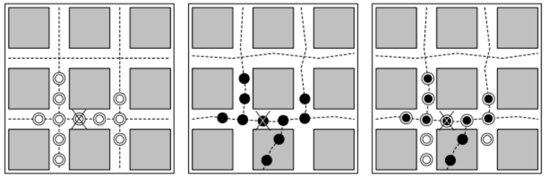








Quantitative Evaluation

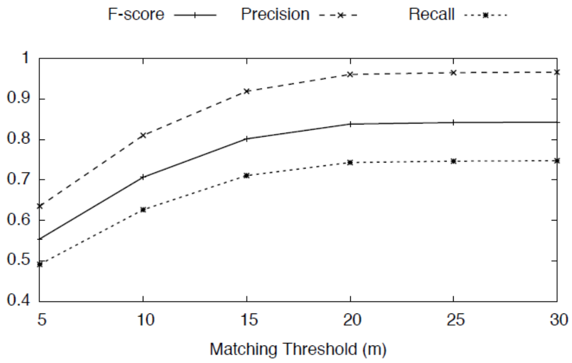


$$F = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

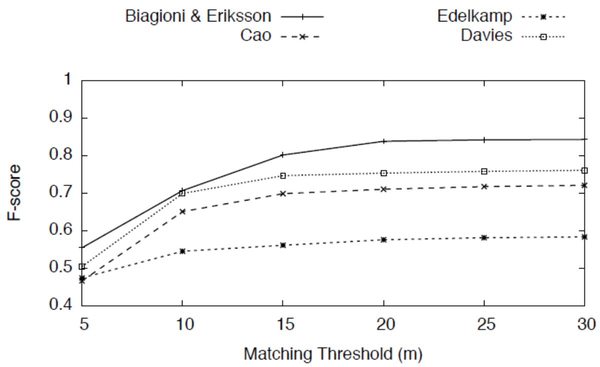
$$precision = 1 - \frac{|spurious\ samples|}{|inferred\ samples|}$$

$$recall = 1 - \frac{|missing\ samples|}{|ground\ truth\ samples|}$$

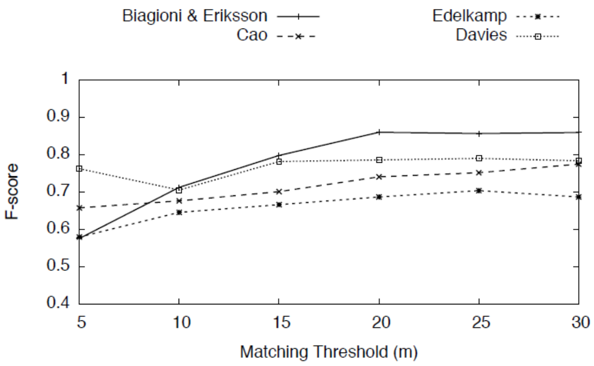
Overall performance



Precision/recall

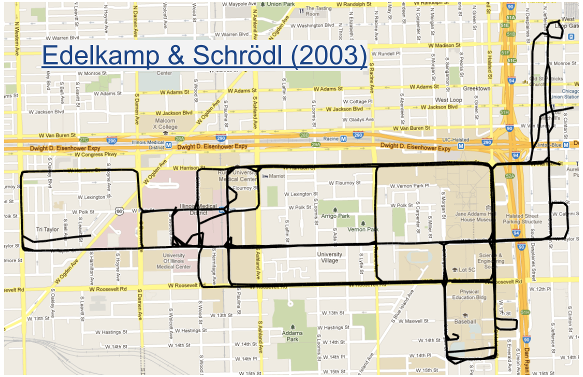


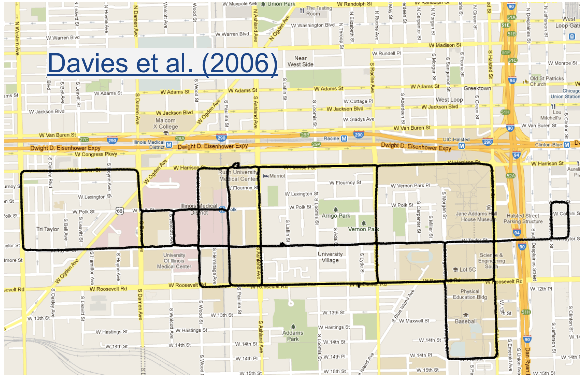
Geometric results

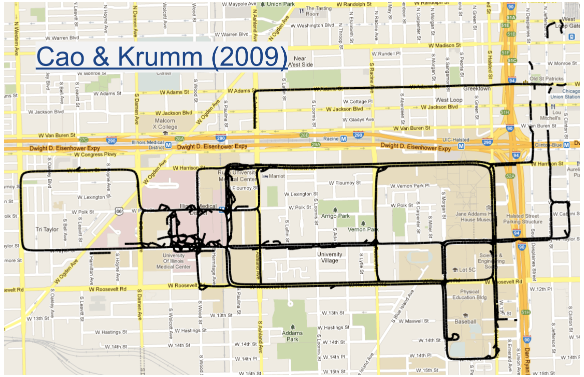


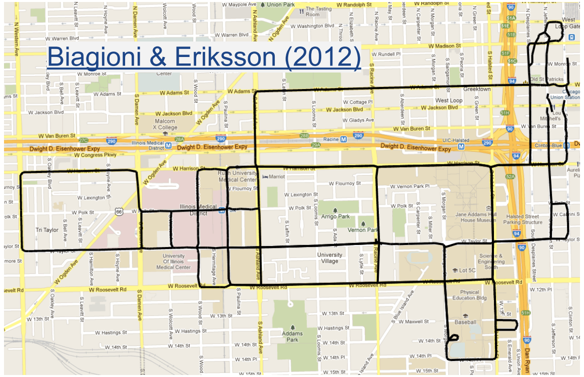
Topological results

Qualitative Evaluation









Limitation


Dense sampled
Chicago shuttle data, 1 second interval


Sparsely Sampled
Shanghai taxi data, 16/61 second interval

Sparsely Sampled Road Maps