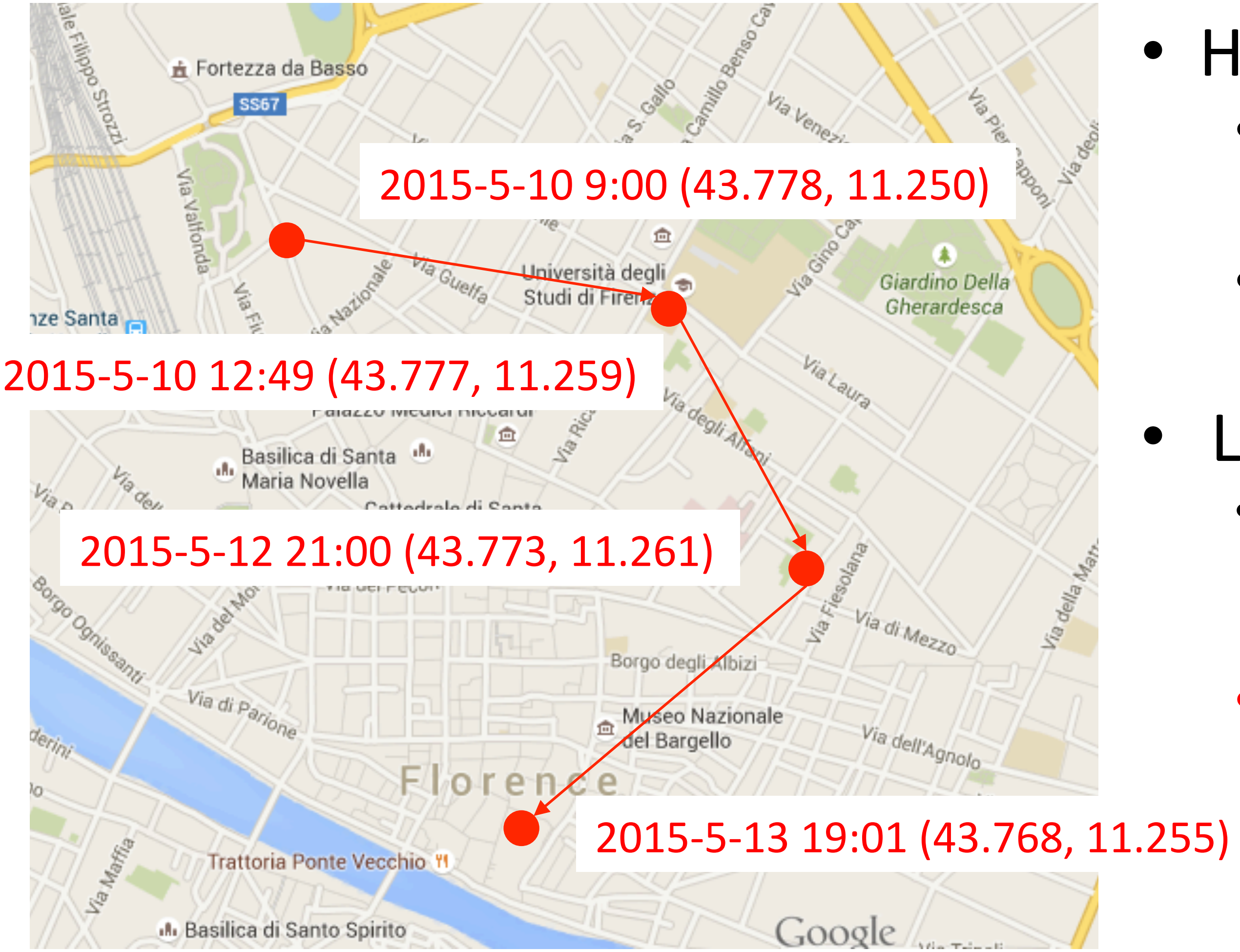
Semantic Annotation of Mobility Data using Social Media

Fei Wu, Zhenhui (Jessie) Li, Wang-Chien Lee, Hongjian Wang, and Zhuojie Huang

The Pennsylvania State University

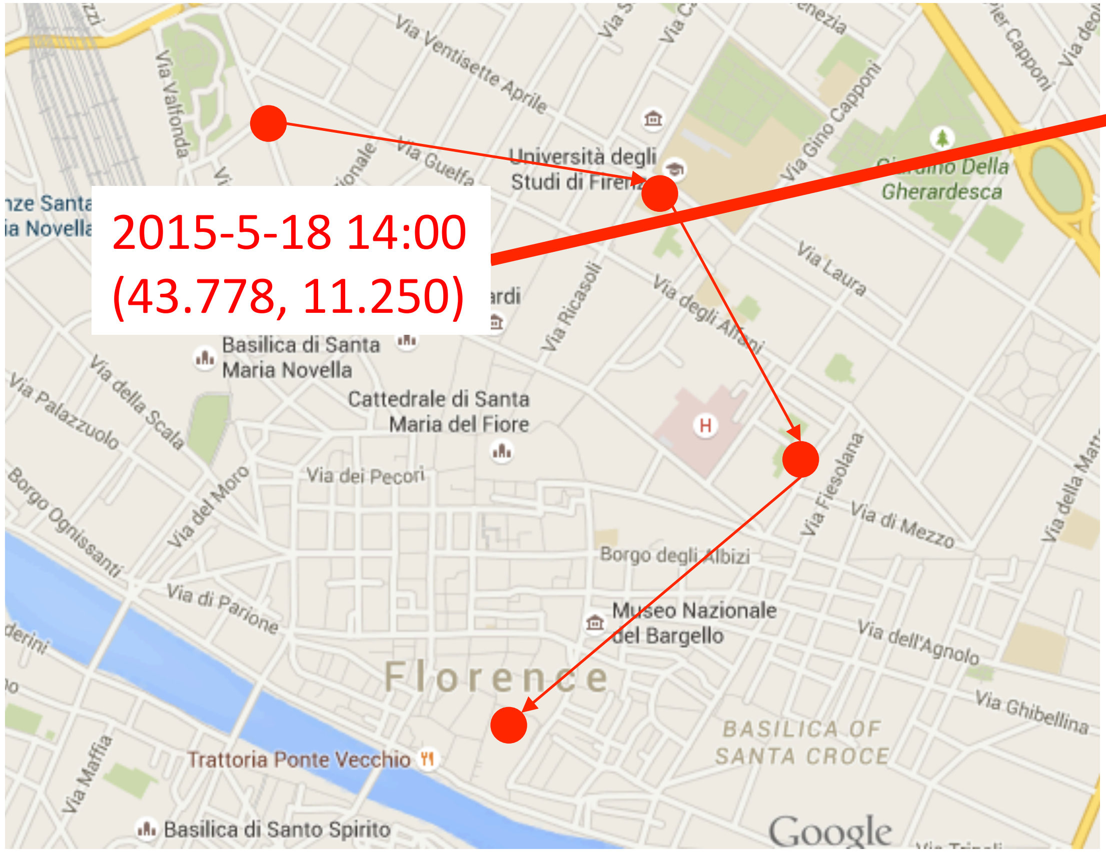
Mining Human Mobility Data



- Human Mobility Data
  - A sequence of time-stamped GPS coordinates
  - Mobile phones, GPS on vehicles, location-based services
- Literature: Mining raw mobility data
  - E.g., Regularity in movements (González et al., 2008; Song et al., 2010; Li et al., 2010), Frequent pattern (Mamoulis et al., 2004; Giannotti et al., 2007)
  - Do not consider the spatial context of locations

Ultimate Objective: Semantically Understand Mobility Data

Raw Location Traces



Semantics

What the person is doing at that location

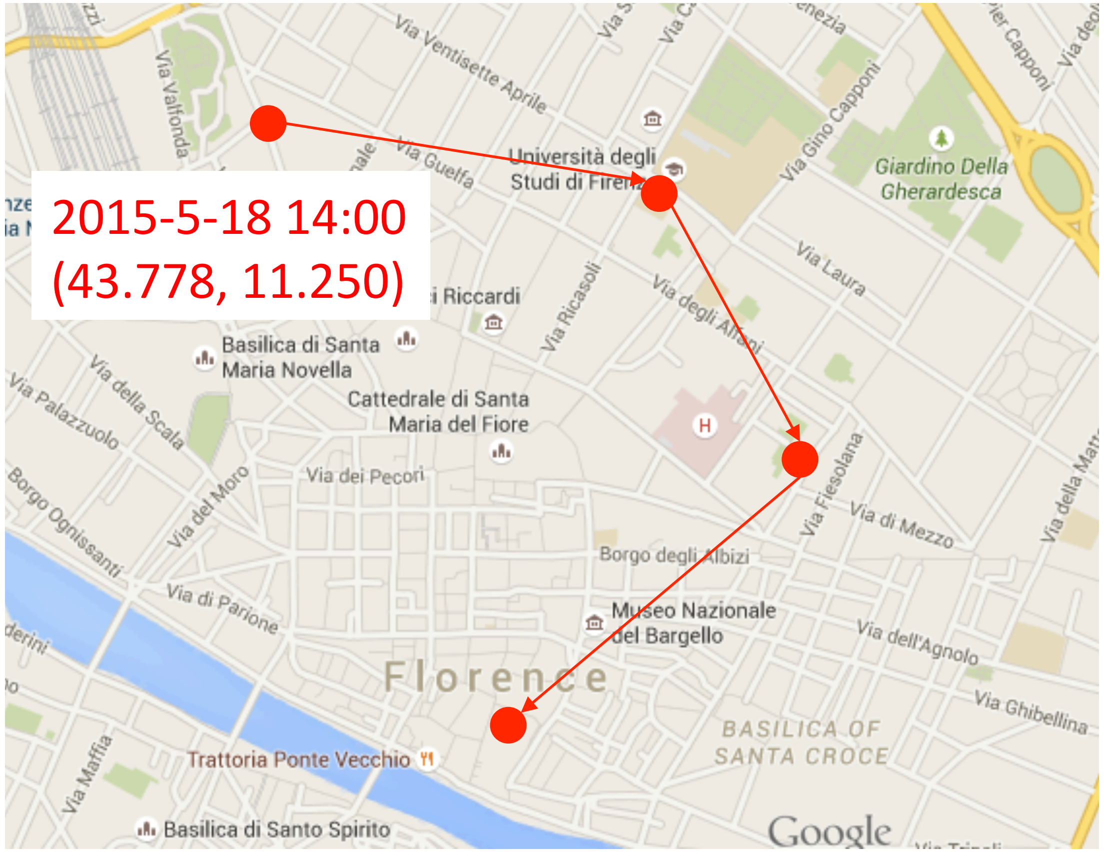
Visit Fortezza da Basso to attend WWW'15

Applications:

1. User targeting and profiling
   - Provide WWW relevant information
   - Conferences & universities → academic person
2. Social science application

Ultimate Objective: Semantically Understand Mobility Data

Raw Location Traces



Semantics

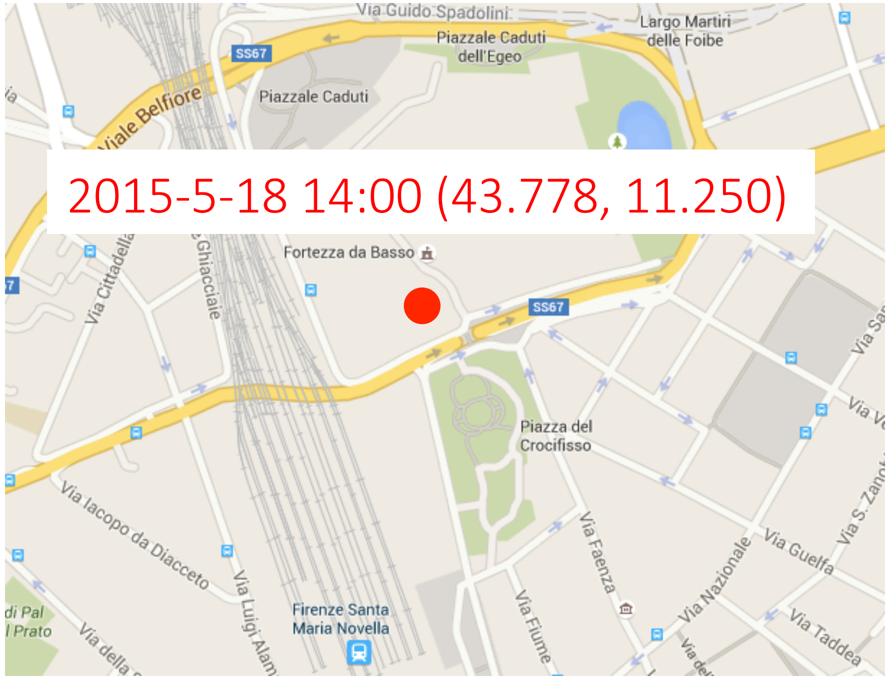
What the person is doing at that location

<u>A very hard problem</u>:

- What is the true destination?
  - conference venue or train station
- Which event is attending?
  - WWW or daily work
- Sparse observations
  - 1-2 points in 3-5 days

A First Step Towards Semantically Understand Mobility Data

Raw Location Traces



Semantics (a sub-problem)

1. Look at a single location record
2. Use some words to describe that record

Questions:

1. How to know what is happening at this place? → geo-tagged social media
2. How to define the relevance of POI/ event to the query location? → looking for the right spatial model

Problem Definition: Semantic Annotation of Location Records

Input: Location history of a mobile user

Record ID	Time	Longitude	Latitude
r_1	2013-1-20	40.75051	-73.99349
r_2	2013-2-10	40.68312	-73.97597
r_3	2013-2-19	40.75051	-73.993499

+

Input: Geo-tagged tweets from the crowd

Time	Longitude	Latitude	Tweets
2013-1-20	40.75051	-73.99349	LETS GO RANGERS
2013-1-20	40.61219	-74.15814	I'm at Buffalo Wild Wings
2013-1-20	40.75050	-73.99350	I'm @ Madison Square Garden for Pittsburgh Penguins vs New York Rangers

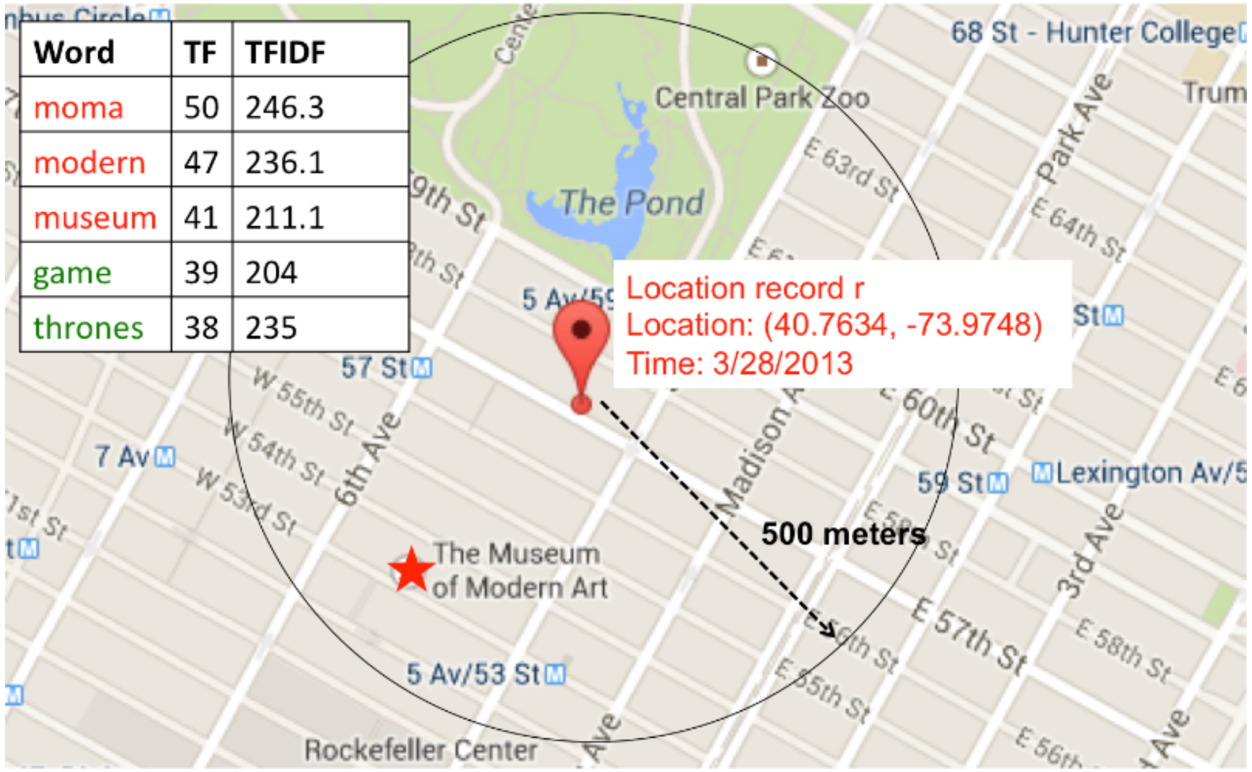
→

Annotation

Record ID	Annotations
r_1	madison, garden, rangers, penguins
r_2	nets, barclays, center, nba
r_3	rangers, madison, square, montreal, canadiens

Baseline Method: Frequency-Based Method

- Count frequency of nearby words
- Did not consider the distance from the user location to the center of the word
- Need to set a threshold to define "nearby"



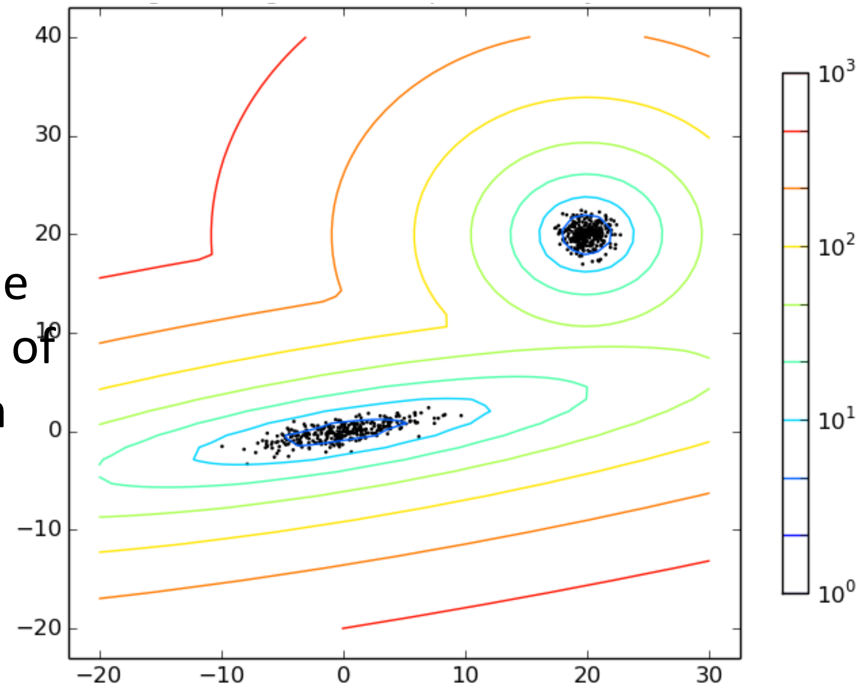
Word	TF	TFIDF
moma	50	246.3
modern	47	236.1
museum	41	211.1
game	39	204
thrones	38	235

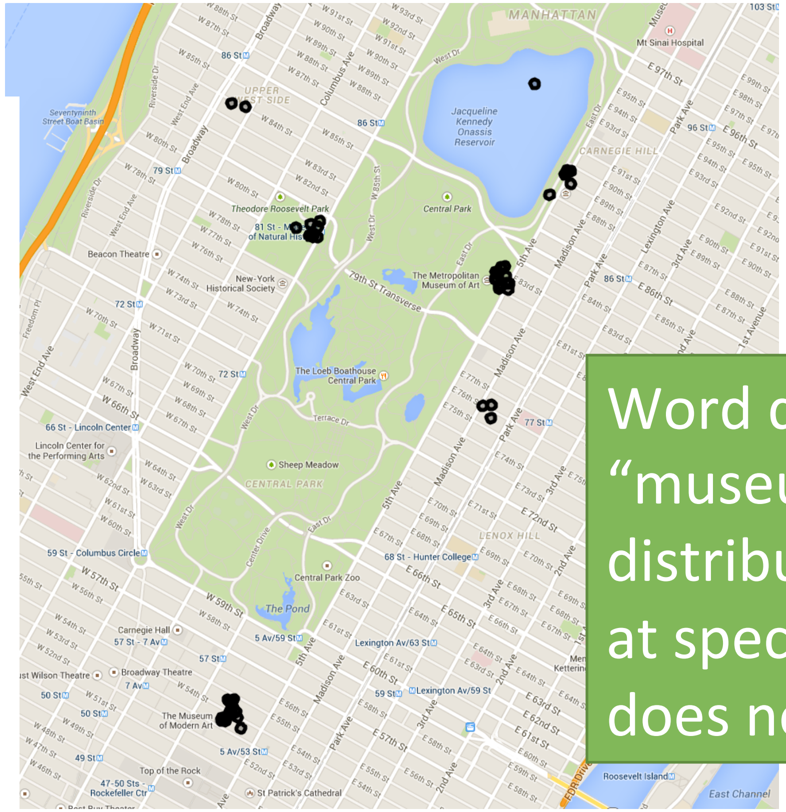
An example illustrating the problem of frequency based methods. The true user's intention of this location record **is to attend the Game of Thrones** event. But since **MOMA is a more popular venue nearby**, frequency-based methods will incorrectly use words "moma" and modern" for annotation.

Baseline Method: Gaussian Mixture Model

- Cons: (1) need to set parameter K; K may vary for different words at different locations. (2) Gaussian distribution may not be the true underlying distributions

An example of mixture of 2 Gaussian





Word distribution for word “museum” in NYC. The distribution is quite skewed at specific locations and does not follow Gaussian

Left figure from: http://scikit-learn.org/stable/modules/mixture.html

A more suitable model – Kernel Density Estimation

- A non-parametric distribution model
  - Check-in distribution (Zhang et al., 2013), (Lichman et al., 2014): demonstrate the effectiveness over Gaussian models
  - Animal's home range (Worton et al., 1989), Epidemiology (Bithell, 1990), Marketing (Donthu and Rust, 1989)

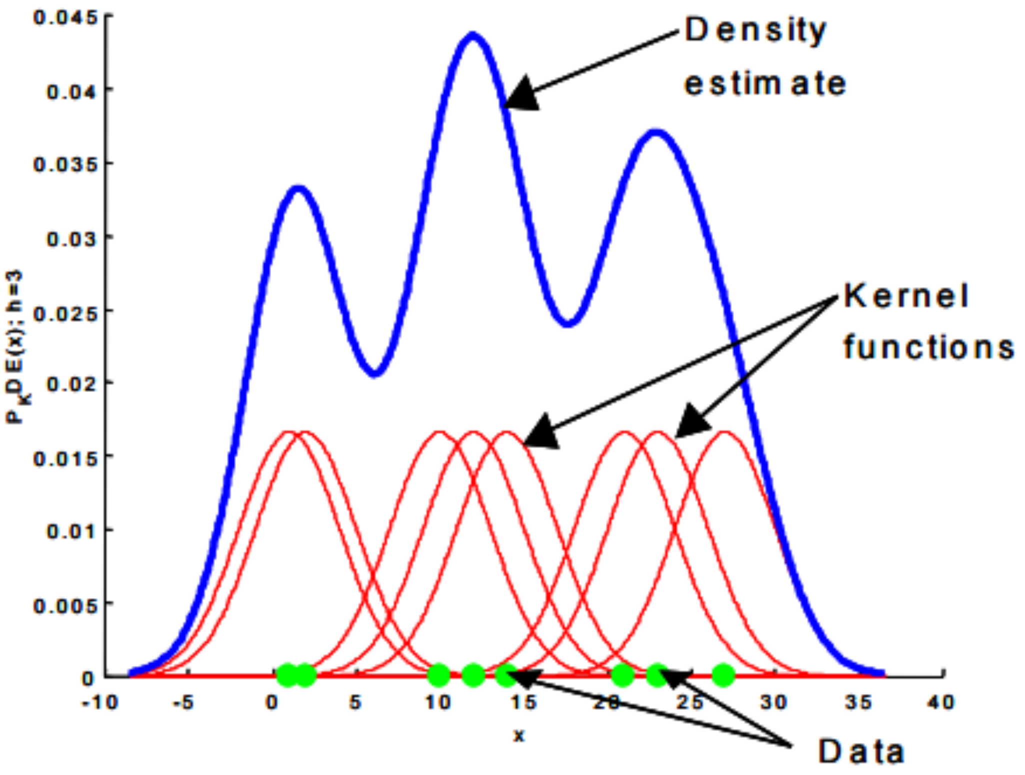
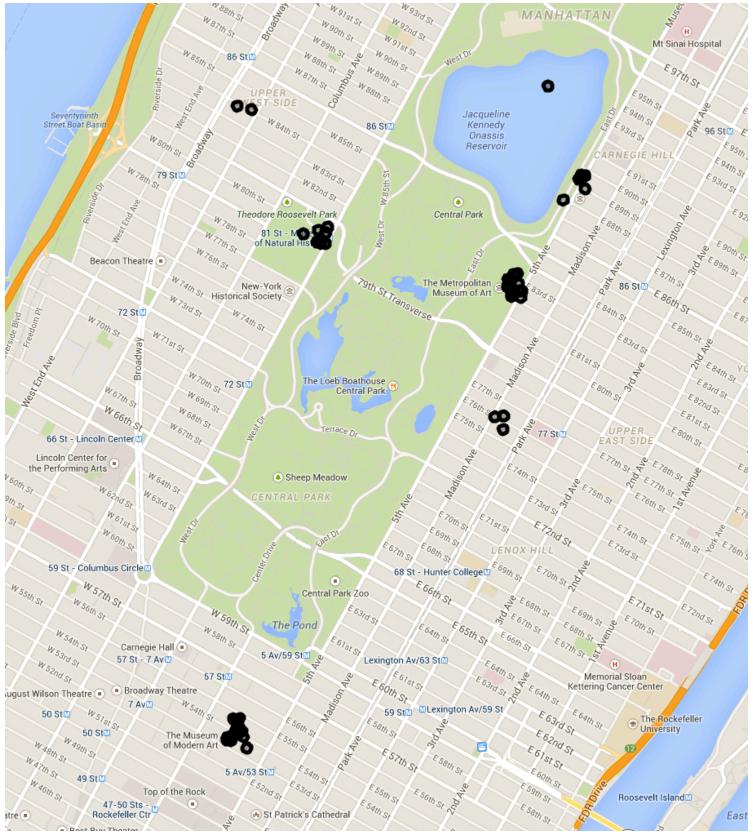


Figure from Wikipedia

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) \quad = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{x - x_i}{h}\Big)$$

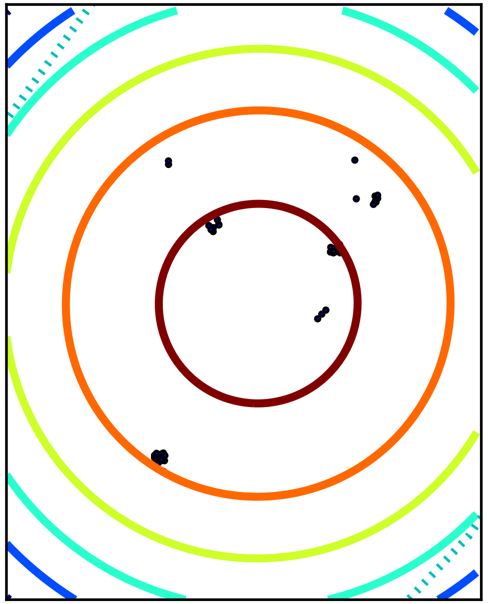
Kernel function (e.g., Gaussian)

KDE: parameter *h* controls the "sharpness" of spatial distribution

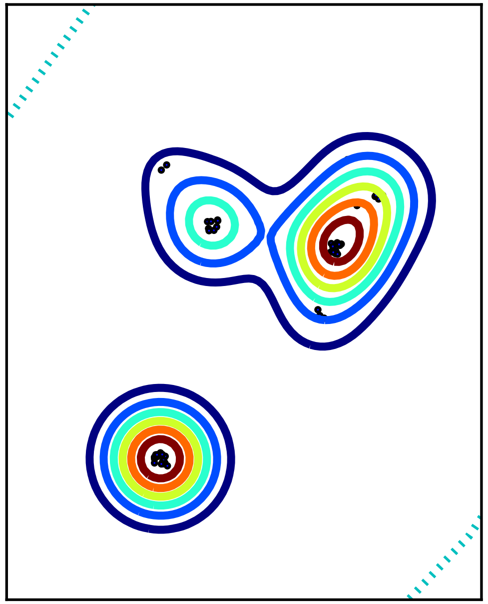


distribution for word "museum"

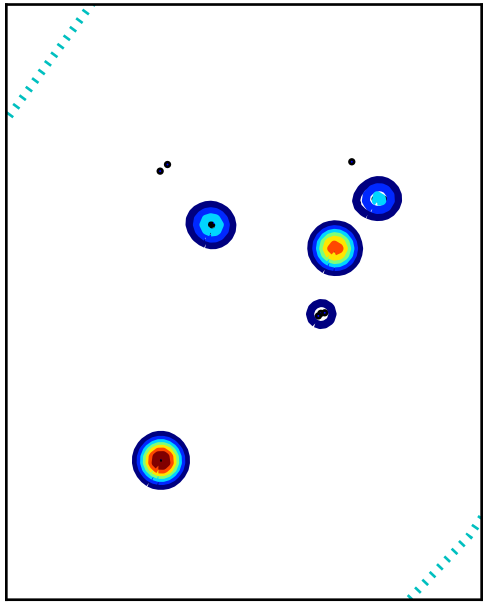
h=0.01



h=0.003



h=0.001



Experiment Setting

- Use geo-tagged tweets from crowd as spatial context
- Select some check-in tweets as the "ground truth"

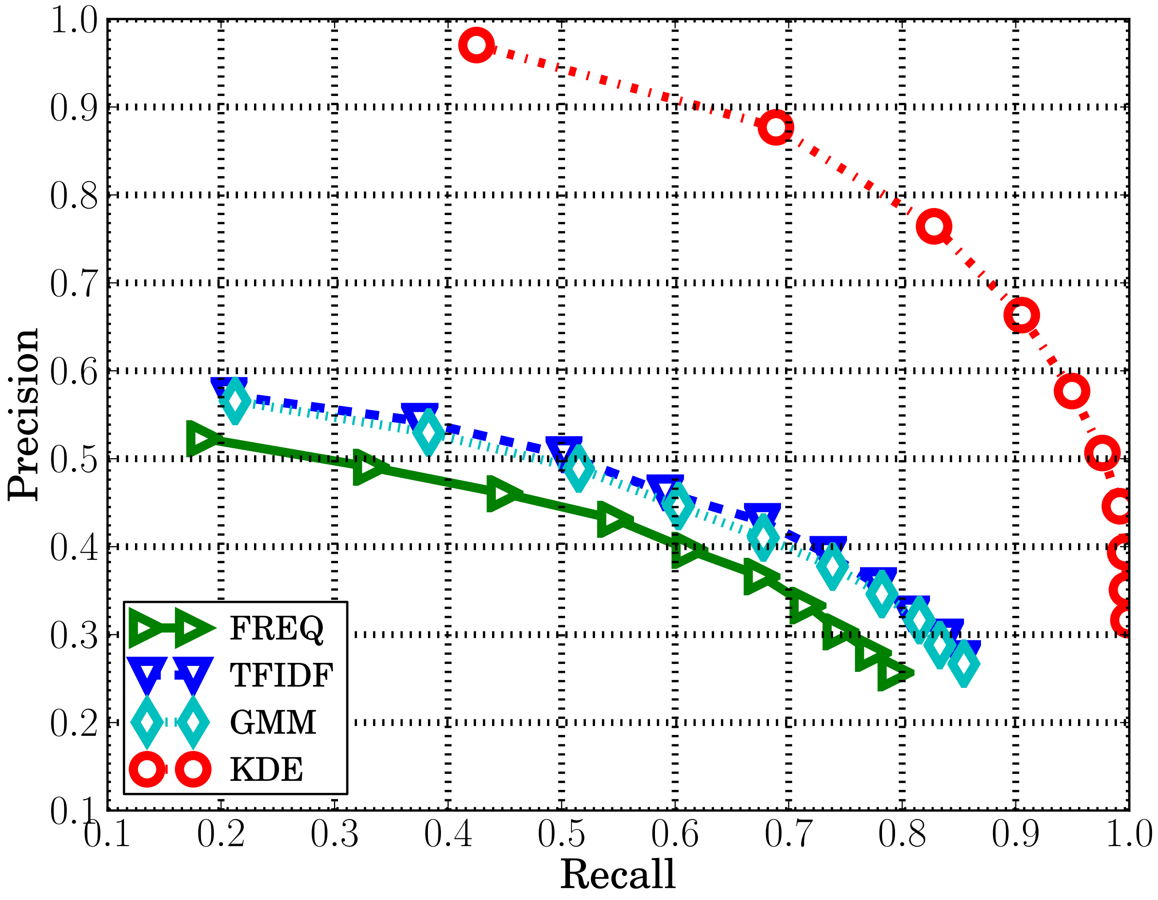
LETS GO RANGERS (@ Madison Square Garden for Pittsburgh Penguins vs New York Rangers w/ 60 others) at time 2013-4-20 19:00

City	#Testing Tweets	#Crowd tweets	Time range
New York City (NYC)	1,540	15,612,712	11/2012-7/2013
Chicago (CHI)	697	11,269,220	10/2011-7/2013
Los Angeles (LA)	623	10,989,333	11/2012-7/2013

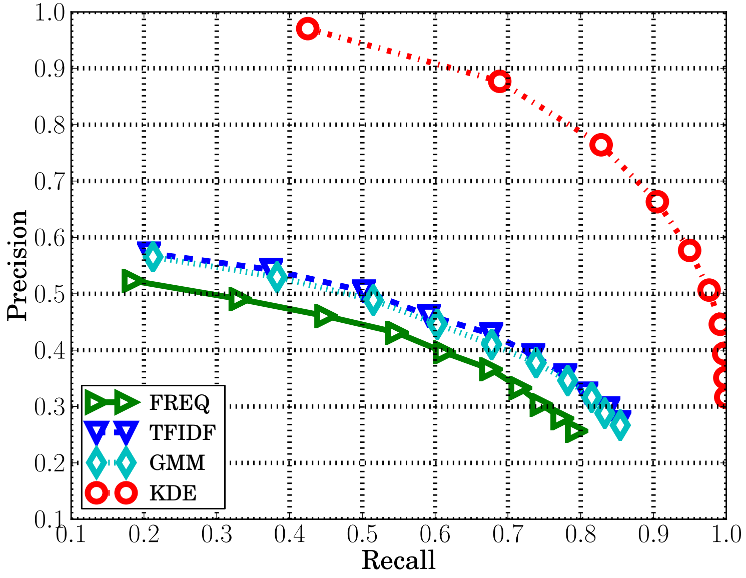
Comparison with Baseline Methods

- FREQ: Rank words based on frequency
- TFIDF: Frequency weighted by IDF
- GMM: Gaussian Mixture Model
- KDE: Kernel Density Estimation (our method)

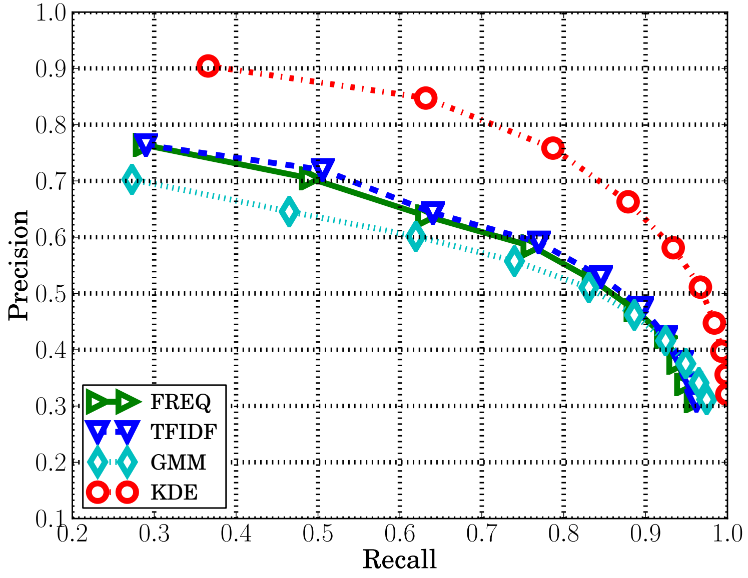
NYC



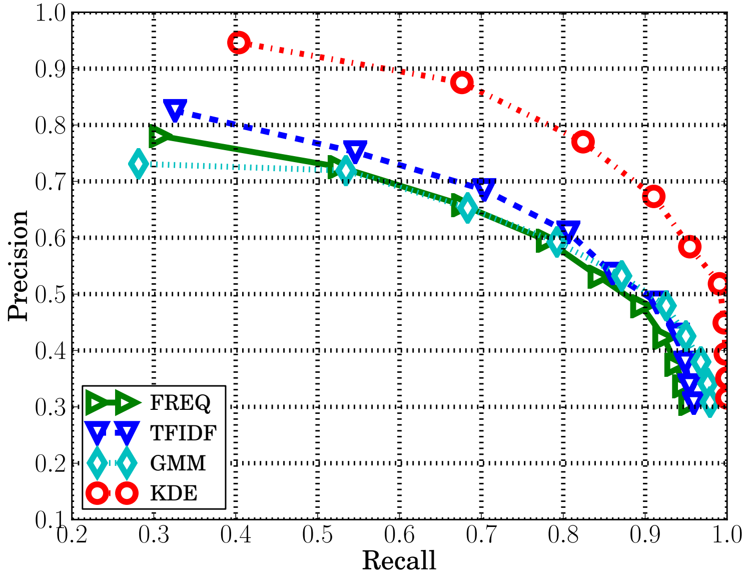
Comparison with Baseline Methods



NYC



Los Angeles

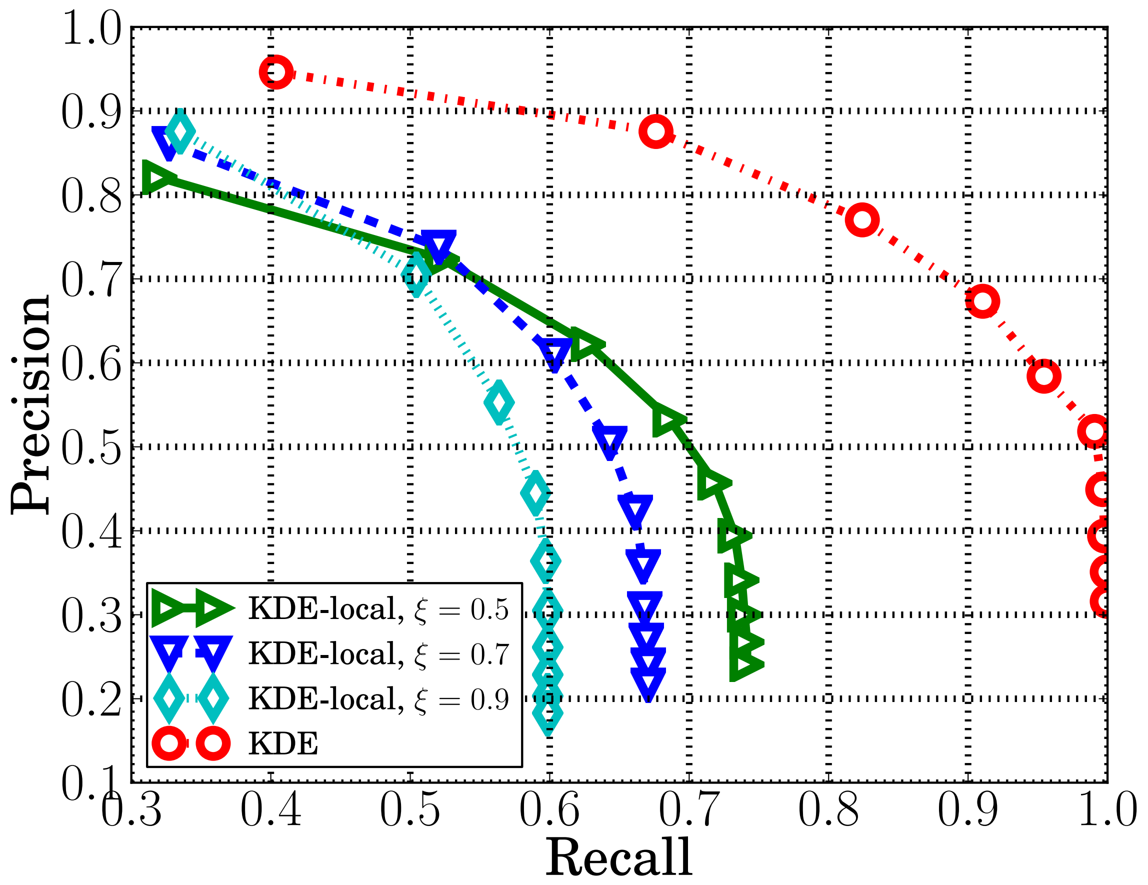


Chicago

- FREQ: Rank words based on frequency
- TFIDF: Frequency weighted by IDF
- GMM: Gaussian Mixture Model
- KDE: Kernel Density Estimation (our method)

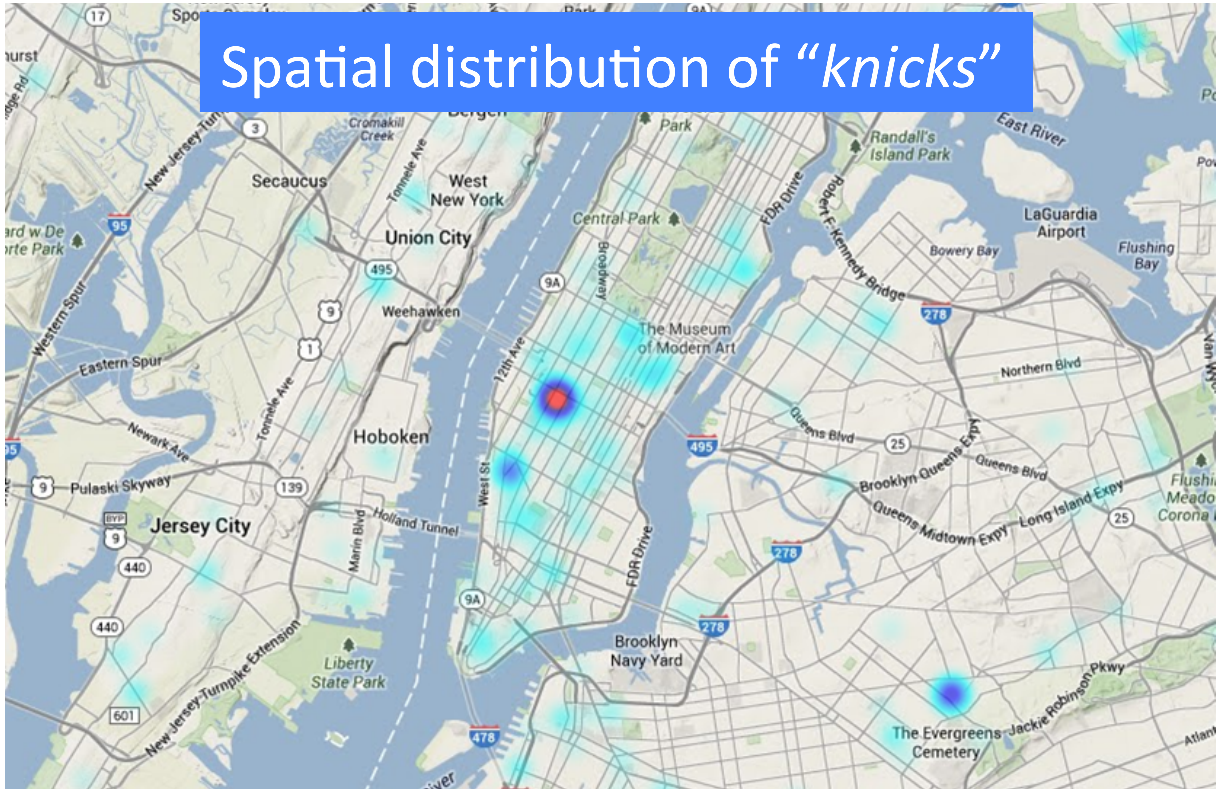
Related Work for Comparison: Local Words

- Local word (highly skewed in spatial distribution)
  1. Measure the locality score of each word using method (Backstrom et al., 2008; Cheng et al., 2010)
  2. Filter non-local words with score lower than ξ



Related Work for Comparison: Local Words

- Local word detection as a filter
- Conclusion:
  1. Local word may filter the true local words w.r.t a query location
  2. There is no need to apply local word filter because KDE already captures the locality at a given location



Word "knicks" is considered as a non local word because people watching games at home/bar are also talking about "knicks". But "knicks" should be the annotation words for people at the arena

Related Work for Comparison: Semantic Trajectory

- Semantic trajectory annotation
  - Map the location to the nearest Point-Of-Interest (Yan et al., 2011; Bogorny et al., 2014)
- Cons:
  1. does not capture the dynamic event information from the social media
  2. the nearest POI might not be the true destination

Case study for a user A

(a) Annotation documents using tweets (our method)

r_1: (40.7505, -73.9934), 1/20/13

Word	Relevance Score
garden	152
madison	146
square	144
rangers	99
penguins	58

r_2: (40.6831, -73.9760), 2/10/13

Word	Relevance Score
barclays	123
center	120
nets	83
spurs	73
san	40

r_3: (40.7505, -73.9934), 3/24/13

Word	Relevance Score
garden	79
madison	77
square	76
nyr	51
rangers	44

r_4: (40.7634, -73.9748), 3/28/13

Word	Relevance Score
game	33
thrones	32
apple	4
store	3
plaza	3

r_5: (40.8295, -73.9270), 4/18/13

Word	Relevance Score
yankee	152
stadium	142
yankees	94
others	38
bronx	36

r_6: (40.8295, -73.9269), 6/5/13

Word	Relevance Score
yankee	155
stadium	148
yankees	108
indians	48
game	43

(b) Locations on the map

(c) Annotation by static POI information (method for comparison)

r_1, r_3: (40.7505, -73.9934)

POI	Distance
Madison Square Garden	3
Auntie Anne's Pretzels	9
Forest Electric Corporation	16
Auntie Anne's Pretzels	21
Event Bar	31

r_2: (40.6831, -73.9760)

POI	Distance
Barclays Center	72
Burlington Coat Factory	79
Jean-Baptiste Gary E MD	84
National Vision	98
Wellness Center	99

r_5, r_6: (40.8259, -73.9269)

POI	Distance
New York Yankees	34
Yankees Audi Club	37
Johnny Rockets	64
Yankee Stadium	78
Apple Bank For Saving	129

r_4: (40.7634, -73.9748)

POI	Distance
Manhattan Dental Health	9
Och-Ziff Capital Management Group	23
Oz Management Group	23
Summit Rock Advisors	23
Ingres Corporation	23

(d) Tweets as ground truth:

r_1: Daddy's home!!!! LETS GO RANGERS (@ Madison Square Garden for Pittsburgh Penguins vs New York Rangers) [pic]: http://t.co/cUYIVcRx
r_2: I'm at Barclays Center for San Antonio Spurs vs Brooklyn Nets (Brooklyn, NY) w/ 69 others http://t.co/UgJxCvlX
r_3: Let's do it. (@ Madison Square Garden - @thegarden for @washcaps vs @NYRangers w/ 44 others) http://t.co/oaaE4f1t9n
r_4: with @jedafrank (@ Game Of Thrones Exhibition w/ 14 others) http://t.co/U5ztm1RWRu
r_5: I'm at Yankee Stadium - @mlb for Arizona Diamondbacks vs New York Yankees (Bronx, NY) w/ 129 others [pic]: http://t.co/MnHVcn170P
r_6: I'm at Yankee Stadium - @mlb for Cleveland Indians vs New York Yankees (Bronx, NY) w/ 143 others http://t.co/KTidl9wtEv

Case study for a user A

(a) Annotation documents using tweets (our method)

Annotated words(our method)

r_1: (40.7505, -73.9934), 1/20/13

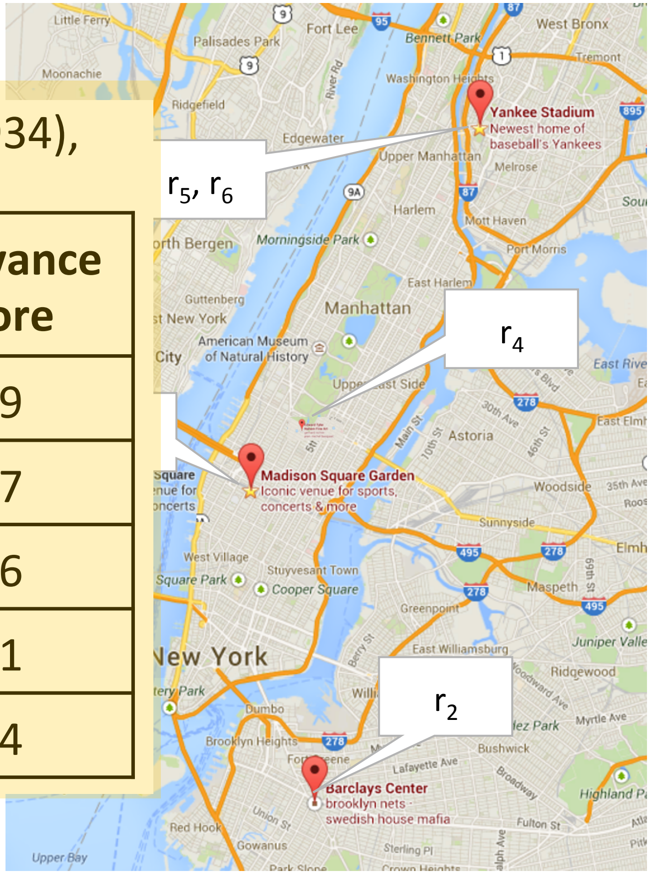
Word	Relevance Score
garden	152
madison	146
square	144
rangers	99
penguins	58

r_3: (40.7505, -73.9934), 3/24/13

Word	Relevance Score
garden	79
madison	77
square	76
nyr	51
rangers	44

plaza	3
bronx	36
game	43

(b) Locations on the map



(c) Annotat… comparison)

r_2: (40.6831, -73.9760)

Annotated by closest POI

r_1, r_3: (40.7505, -73.9934)

POI	Distance
Madison Square Garden	3
Auntie Anne's Pretzels	9
Forest Electric Corporation	16
Auntie Anne's Pretzels	21
Event Bar	31

Apple Bank For Saving	129

Summit Rock Advisors	23
Ingres Corporation	23

(d) Tweets as ground truth:

r_1: Daddy's home!!!! LETS GO RANGERS (@ Madison Square Garden for Pittsburgh Penguins vs New York Rangers) [pic]: http://t.co/cUYIVcRx
r_2: I'm at Barclays Center for San Antonio Spurs vs Brooklyn Nets (Brooklyn, NY) w/ 69 others http://t.co/UgJxCvlX
r_3: Let's do it. (@ Madison Square Garden - @thegarden for @washcaps vs @NYRangers w/ 44 others) http://t.co/oaaE4f1t9n
r_4: with @jedafrank (@ Game Of Thrones Exhibition w/ 14 others) http://t.co/U5ztm1RWRu
r_5: I'm at Yankee Stadium - @mlb for Arizona Diamondbacks vs New York Yankees (Bronx, NY) w/ 129 others [pic]: http://t.co/MnHVcn170P
r_6: I'm at Yankee Stadium - @mlb for Cleveland Indians vs New York Yankees (Bronx, NY) w/ 143 others http://t.co/KTidl9wtEv

Case study for a user A

(a) Annotation documents using tweets (our method)

r_1: (40.7505, -73.9934), 1/20/13

Word	Relevance Score
garden	152
madison	146
square	144
rangers	99
penguins	58

r_2: (40.6831, -73.9760), 2/10/13

Word	Relevance Score
barclays	123
center	120
nets	83
spurs	73
san	40

r_3: (40.7505, -73.9934), 3/24/13

Word	Relevance Score
garden	79
madison	77
square	76
nyr	51
rangers	44

r_4: (40.7634, -73.9748), 3/28/13

Word	Relevance Score
game	33
thrones	32
apple	4
store	3
plaza	3

r_5: (40.8295, -73.9270), 4/18/13

Word	Relevance Score
yankee	152
stadium	142
yankees	94
others	38
bronx	36

r_6: (40.8295, -73.9269), 6/5/13

Word	Relevance Score
yankee	155
stadium	148
yankees	108
indians	48
game	43

(b) Locations on the map

(c) Annotation by static POI information (method for comparison)

r_1 , r_3: (40.7505, -73.9934)

POI	Distance
Madison Square Garden	3
Auntie Anne's Pretzels	9
Forest Electric Corporation	16
Auntie Anne's Pretzels	21
Event Bar	31

r_2: (40.6831, -73.9760)

POI	Distance
Barclays Center	72
Burlington Coat Factory	79
Jean-Baptiste Gary E MD	84
National Vision	98
Wellness Center	99

r_5 , r_6: (40.8259, -73.9269)

POI	Distance
New York Yankees	34
Yankees Audi Club	37
Johnny Rockets	64
Yankee Stadium	78
Apple Bank For Saving	129

r_4: (40.7634, -73.9748)

POI	Distance
Manhattan Dental Health	9
Och-Ziff Capital Management Group	23
Oz Management Group	23
Summit Rock Advisors	23
Ingres Corporation	23

(d) Tweets as ground truth:

r_1: Daddy's home!!!! LETS GO RANGERS (@ Madison Square Garden for Pittsburgh Penguins vs New York Rangers) [pic]: http://t.co/cUYIVcRx
r_2: I'm at Barclays Center for San Antonio Spurs vs Brooklyn Nets (Brooklyn, NY) w/ 69 others http://t.co/UgJxCvlX
r_3: Let's do it. (@ Madison Square Garden - @thegarden for @washcaps vs @NYRangers w/ 44 others) http://t.co/oaaE4f1t9n
r_4: with @jedafrank (@ Game Of Thrones Exhibition w/ 14 others) http://t.co/U5ztm1RWRu
r_5: I'm at Yankee Stadium - @mlb for Arizona Diamondbacks vs New York Yankees (Bronx, NY) w/ 129 others [pic]: http://t.co/MnHVcn170P
r_6: I'm at Yankee Stadium - @mlb for Cleveland Indians vs New York Yankees (Bronx, NY) w/ 143 others http://t.co/KTidl9wtEv

Case study for a user B

(a) Annotation documents using tweets (our method)

r_1: (40.7583, -73.9854), 5/5/13

Word	Relevance Score
museum	25
art	24
publicar	2
much	2
creative	2

r_2: (40.778, -73.962), 5/6/13

Word	Relevance Score
art	68
metropolitan	66
museum	65
metgala	29
met	24

r_3: (40.7684, -73.9745), 5/13/13

Word	Relevance Score
ice	13
fox	13
upfront	6
central	5
centralpark	2

r_4: (40.8294, -73.9269), 5/15/13

Word	Relevance Score
yankee	223
stadium	208
yankees	88
bronx	52
america	32

r_5: (40.8295, -73.9269), 4/18/13

Word	Relevance Score
yankee	183
stadium	169
yankees	31
bronx	30
baseball	8

r_6: (40.7564, -73.8460), 6/5/13

Word	Relevance Score
field	106
citi	102
edcny	68
edc	67
edm	31

(b) Locations on the map

(c) Annotation by static POI information (method for comparison)

r_1 : (40.7583, -73.9854)

POI	Distance
Forever 21	26
Disney Store	26
Hodgson Russ LLP	30
00 Commercial Blinds	33
New York Society of Security analysts	35

r_2: (40.778, -73.962)

POI	Distance
995 Fifth Avenue	56
Frank E. Campbell – The Funeral Chapel	56
Thomas W. Loeb, MD	59
Olive & Bettes	70
Universal Funeral Chapel	70

r_3: (40.7684, -73.9745)

POI	Distance
Victorian Garden Amusement Park	66
Wollman Ice Skating Park	70
Central Park Carousel	174
Central Park Zoo	174
The Arsenal	282

r_6: (40.7564, -73.8460)

POI	Distance
Citi Field	9
Shea Stadium Home Run Apple	138
Mama's of Corona	151
Mets Plaza	170
NYC Parks & Recreation	183

*r_4,r_5: (40.8295, -73.9269): Annotation by POI is the same as r_5, r_6 in Figure 7.

(d) Tweets as ground truth:

r_1: katyperry @ Museum of Modern Art (MoMA) http://t.co/j9ql2faeRw
r_2: .emmyrossum #MetGala #PunkFashion @ The Metropolitan Museum of Art http://t.co/T11YQfm6KG
r_3: @minkakelly #foxupfront @ Wollman Park http://t.co/56dcTzx5YJA
r_4: @bearpascoee @4stillrunning @ruebenrandle #bleachercreatures #rollcall yankees #204 @ Yankee Staduim http://t.co/PIGxyatsF9
r_5: #photoday yankees @ Yankee Stadium http://t.co/ySy2uHTrDh
r_6: @djafrojack Kinetic Field @ edc_lasvegas @insomniac #edcny @ Citi Field http://t.co/dGADzm1911

Case study for a user B

(a) Ann Annotated words(our method) **(b) Locations on the map** **(c) An** Annotated by closest POI **r comparison)**

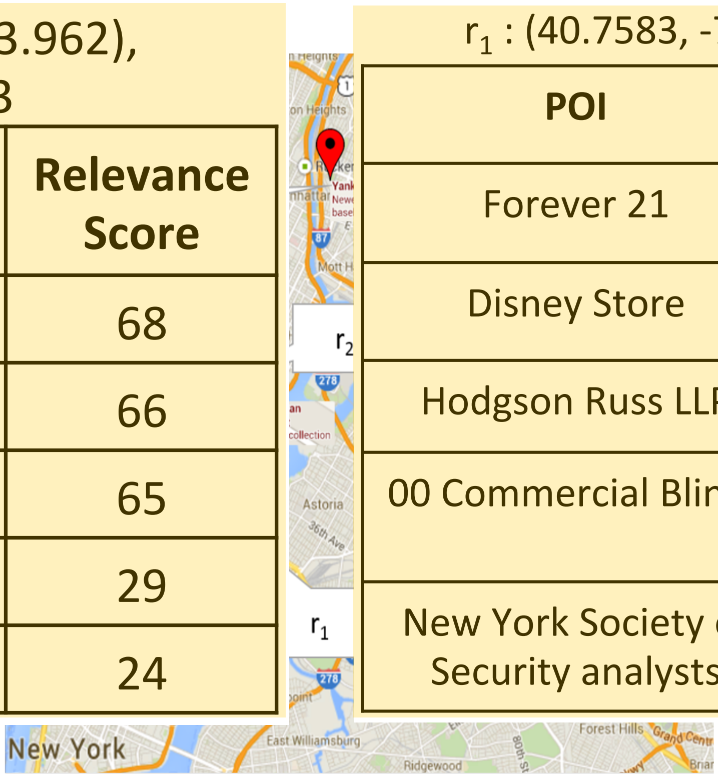
r_1: (40.7583, -73.9854), 5/5/13

Word	Relevance Score
museum	25
art	24
publicar	2
much	2
creative	2

r_2: (40.778, -73.962), 5/6/13

Word	Relevance Score
art	68
metropolitan	66
museum	65
metgala	29
met	24

america	32	baseball	8	edm	31



r_1 : (40.7583, -73.9854)

POI	Distance
Forever 21	26
Disney Store	26
Hodgson Russ LLP	30
00 Commercial Blinds	33
New York Society of Security analysts	35
Central Park Carousel	174
Central Park Zoo	174
The Arsenal	282

r_2: (40.778, -73.962)

POI	Distance
995 Fifth Avenue	56
Frank E. Campbell – The Funeral Chapel	56
Thomas W. Loeb, MD	59
Olive & Bettes	70
Universal Funeral Chapel	70
Mama's of Corona	151
Mets Plaza	170
NYC Parks & Recreation	183

(d) Tweets as ground truth:

r_1: katyperry @ Museum of Modern Art (MoMA) http://t.co/j9ql2faeRw
r_2: .emmyrossum #MetGala #PunkFashion @ The Metropolitan Museum of Art http://t.co/T11YQfm6KG
r_3: @minkakelly #foxupfront @ Wollman Park http://t.co/56dcTzx5YJA
r_4: @bearpascoee @4stillrunning @ruebenrandle #bleachercreatures #rollcall yankees #204 @ Yankee Staduim http://t.co/PIGxyatsF9
r_5: #photoday yankees @ Yankee Stadium http://t.co/ySy2uHTrDh
r_6: @djafrojack Kinetic Field @ edc_lasvegas @insomniac #edcny @ Citi Field http://t.co/dGADzm1911

*r_4,r_5: (40.8295, -73.9269): Annotation by POI is the same as r_5, r_6 in Figure 7.

Case study for a user B

(a) Annotation documents using tweets (our method)

r_1: (40.7583, -73.9854), 5/5/13

Word	Relevance Score
museum	25
art	24
publicar	2
much	2
creative	2

r_2: (40.778, -73.962), 5/6/13

Word	Relevance Score
art	68
metropolitan	66
museum	65
metgala	29
met	24

r_3: (40.7684, -73.9745), 5/13/13

Word	Relevanc e Score
ice	13
fox	13
upfront	6
central	5
centralpark	2

r_4: (40.8294, -73.9269), 5/15/13

Word	Relevance Score
yankee	223
stadium	208
yankees	88
bronx	52
america	32

r_5: (40.8295, -73.9269), 4/18/13

Word	Relevance Score
yankee	183
stadium	169
yankees	31
bronx	30
baseball	8

r_6: (40.7564, -73.8460), 6/5/13

Word	Relevance Score
field	106
citi	102
edcny	68
edc	67
edm	31

(b) Locations on the map

(c) Annotation by static POI information (method for comparison)

r_1 : (40.7583, -73.9854)

POI	Distance
Forever 21	26
Disney Store	26
Hodgson Russ LLP	30
00 Commercial Blinds	33
New York Society of Security analysts	35

r_2: (40.778, -73.962)

POI	Distance
995 Fifth Avenue	56
Frank E. Campbell – The Funeral Chapel	56
Thomas W. Loeb, MD	59
Olive & Bettes	70
Universal Funeral Chapel	70

r_3: (40.7684, -73.9745)

POI	Distance
Victorian Garden Amusement Park	66
Wollman Ice Skating Park	70
Central Park Carousel	174
Central Park Zoo	174
The Arsenal	282

r_6: (40.7564, -73.8460)

POI	Distance
Citi Field	9
Shea Stadium Home Run Apple	138
Mama's of Corona	151
Mets Plaza	170
NYC Parks & Recreation	183

*r_4,r_5: (40.8295, -73.9269): Annotation by POI is the same as r_5, r_6 in Figure 7.

(d) Tweets as ground truth:

r_1: katyperry @ Museum of Modern Art (MoMA) http://t.co/j9ql2faeRw
r_2: .emmyrossum #MetGala #PunkFashion @ The Metropolitan Museum of Art http://t.co/T11YQfm6KG
r_3: @minkakelly #foxupfront @ Wollman Park http://t.co/56dcTzx5YJA
r_4: @bearpascoee @4stillrunning @ruebenrandle #bleachercreatures #rollcall yankees #204 @ Yankee Staduim http://t.co/PIGxyatsF9
r_5: #photoday yankees @ Yankee Stadium http://t.co/ySy2uHTrDh
r_6: @djafrojack Kinetic Field @ edc_lasvegas @insomniac #edcny @ Citi Field http://t.co/dGADzm1911

Extension: User Profiling by Clustering the Annotated Location Records

33 location records from user A in NYC

Sports fan

Cluster	Top-5 Words
1	yankee, stadium, bronx
2	garden, madison, square, rangers, penguins
3	barclays, center, brooklyn, nets, spurs
4	yankee, stadium, oldtimersday
5	public, island, plaza, staten, drinking

Top words from this person's tweets:
staten, stadium, rpx, hylan, island, drinking, rangers, yankee, bronx, plaza, madison, garden, photo, ale

250 location records from user B in NYC

Fashion and Sports fan

Cluster	Top-5 Words
1	art, metropolitan, museum, metgala, punkfashion
2	Yankee, mets, sox, stadium, orioles
3	Citi, filed, mets, edcny, subwayseries

Top words from this person's tweets:
Metgala, punkfashion, yankee, stadium, begatelle, citi, field, metropolitan, marquee, bleacher, creatures, superstudiu, art, museum

Summary and Future work

- Summary
  - A novel problem of annotating semantics to mobility data
  - KDE to capture the word locality and distance to a locational point
- Futur
  - Ap
  - More precise annotation by considering the mobility history

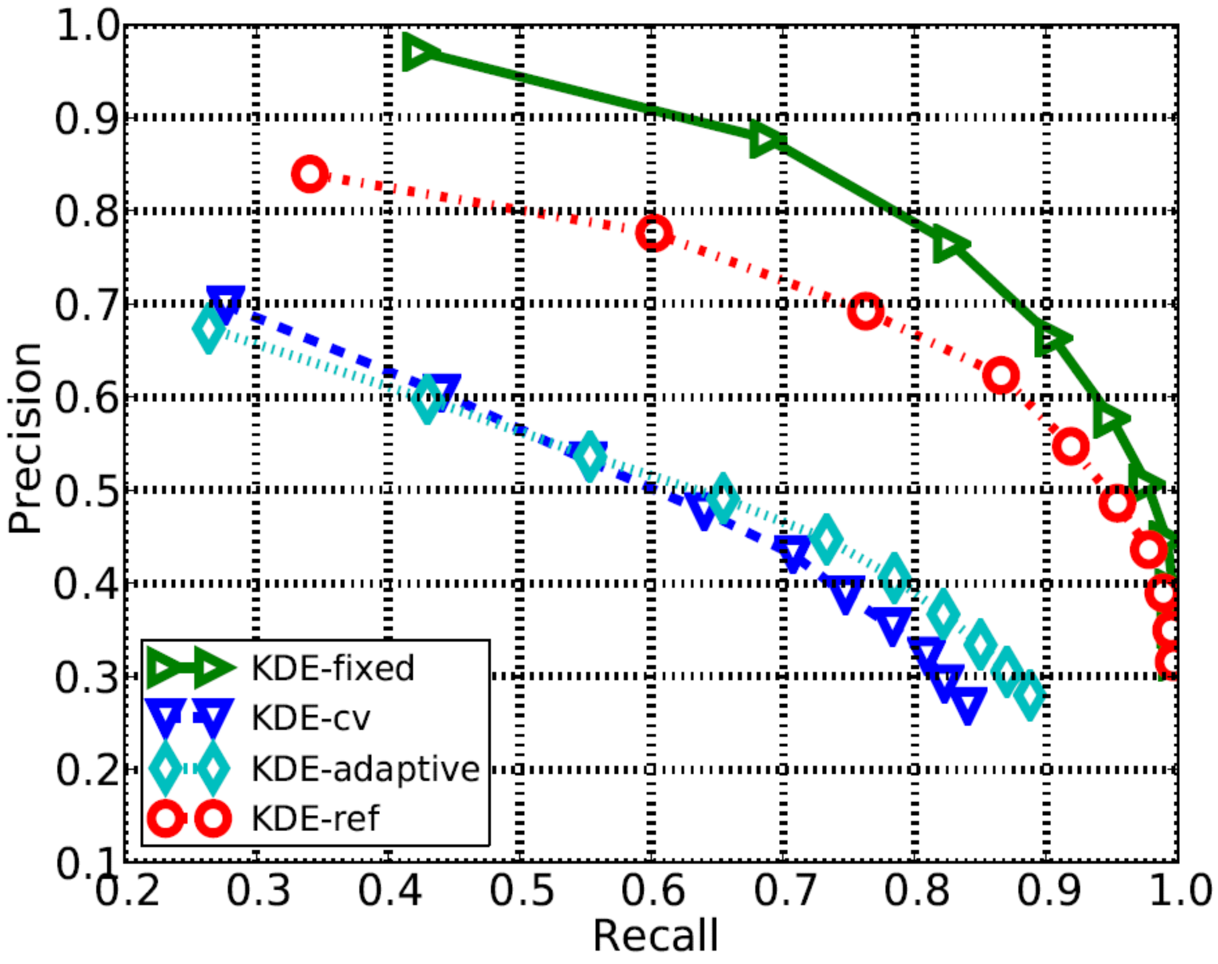
Thanks! Questions?

Determining bandwidth parameter h of KDE

- Fixed values
- Reference rule (heuristic)

$$H_w = n^{-2/(d+4)} \Sigma_w$$

- Cross validation
  - Split data as training and testing
  - Maximize likelihood
- Adaptive bandwidth
  - Choose h inversely proportional to the number of points nearby



Reference

- Gonzalez, Marta C., Cesar A. Hidalgo, and Albert-Laszlo Barabasi. "Understanding individual human mobility patterns." *Nature* 453.7196 (2008): 779-782.
- Song, Chaoming, et al. "Limits of predictability in human mobility." *Science* 327.5968 (2010): 1018-1021.
- Li, Zhenhui, et al. "Mining periodic behaviors for moving objects." *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010.
- Mamoulis, Nikos, et al. "Mining, indexing, and querying historical spatiotemporal data." *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004.
- Giannotti, Fosca, et al. "Trajectory pattern mining." *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007.
- Ye, Mao, et al. "Exploiting geographical influence for collaborative point-of-interest recommendation." *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 2011.
- Ye, Mao, Peifeng Yin, and Wang-Chien Lee. "Location recommendation for location-based social networks." *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2010.
- Yuan Zhong, Nicholas Jing Yuan, Wen Zhong, Fuzheng Zhang, Xing Xie, You Are Where You Go: Inferring Demographic Attributes from Location Check-ins, 8th ACM International Conference on Web Search and Data Mining (WSDM 2015), Shanghai, China, Feb. 2015.

Reference (cont.)

- Zhang, Jia-Dong, and Chi-Yin Chow. "iGSLR: personalized geo-social location recommendation: a kernel density estimation approach." *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2013.
- Lichman, Moshe, and Padhraic Smyth. "Modeling human location data with mixtures of kernel densities." *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014.
- Worton, Brian J. "Kernel methods for estimating the utilization distribution in home-range studies." *Ecology* 70.1 (1989): 164-168.
- Bithell, J. F. "An application of density estimation to geographical epidemiology." *Statistics in medicine* 9.6 (1990): 691-701.
- Donthu, Naveen, and Roland T. Rust. "Note-Estimating Geographic Customer Densities Using Kernel Density Estimation." *Marketing Science* 8.2 (1989): 191-203.
- Backstrom, Lars, et al. "Spatial variation in search engine queries." *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008.
- Cheng, Zhiyuan, James Caverlee, and Kyumin Lee. "You are where you tweet: a content-based approach to geo-locating twitter users." *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010.
- Yan, Zhixian, et al. "SeMiTri: a framework for semantic annotation of heterogeneous trajectories." *Proceedings of the 14th international conference on extending database technology*. ACM, 2011.
- Bogorny, Vania, Bart Kuijpers, and Luis Otavio Alvares. "ST-DMQL: a semantic trajectory data mining query language." *International Journal of Geographical Information Science* 23.10 (2009): 1245-1276.