Tutorial in Location-based Services

Anastasios Noulas

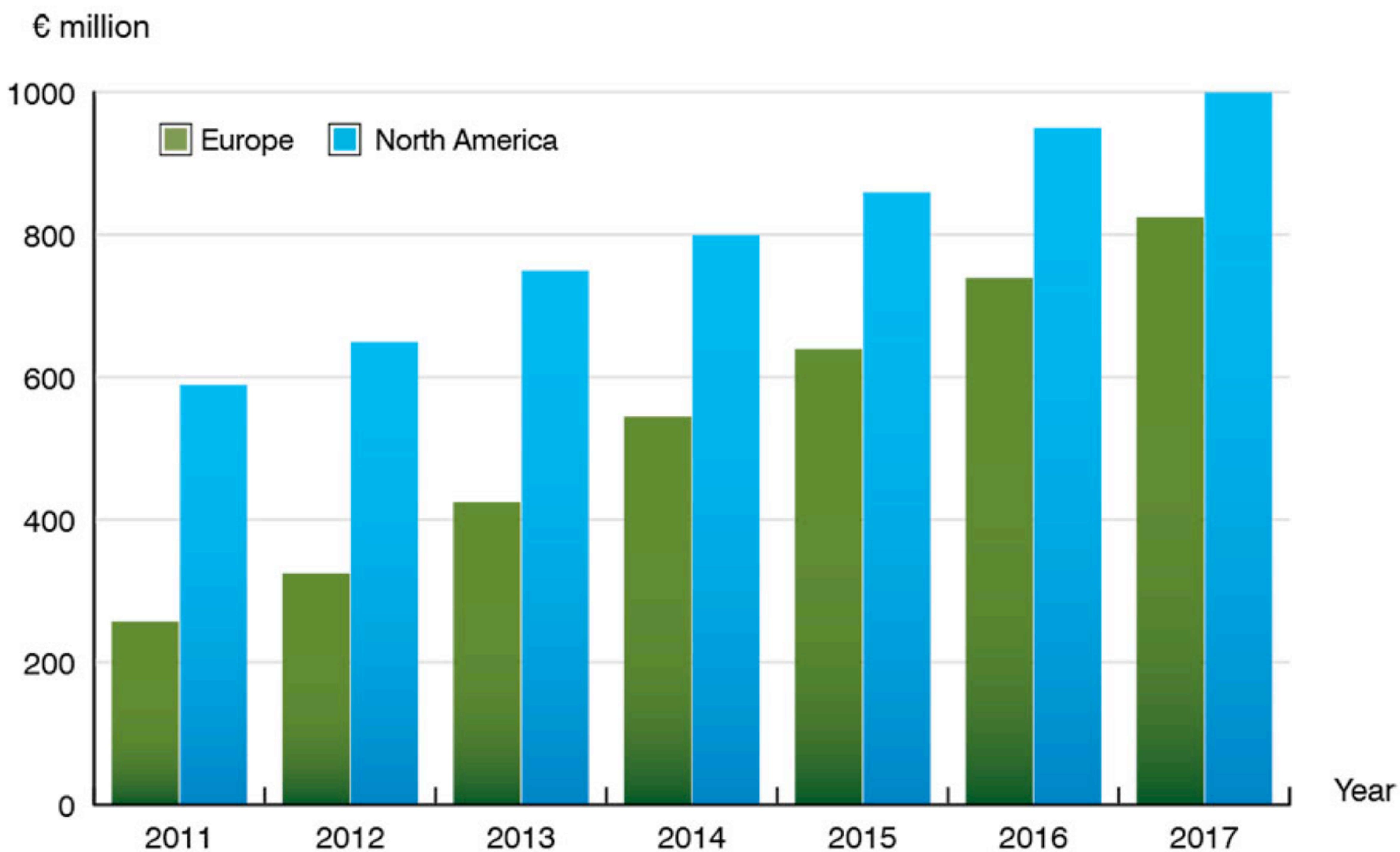
Computer Laboratory, University of Cambridge

@ iSocial Workshop - Online Social Networks: Emerging Trends

University of Cyprus, October, 2014

Location-based social networks

More and more people want to share their geographic position with their friends.



Mobile LBS revenue forecast, € million (2011–2017)

A little a bit of history...

Dodgeball is Foursquare's precursor, founded in 2000 @ New York University.

Not necessarily the first service, but perhaps the most iconic and widespread at the time.

Acquired by Google in 2005...

Dodgeball forever!

Dodgeball founders Dennis Crawley and Alex Rainert leave Google in 2007 ...

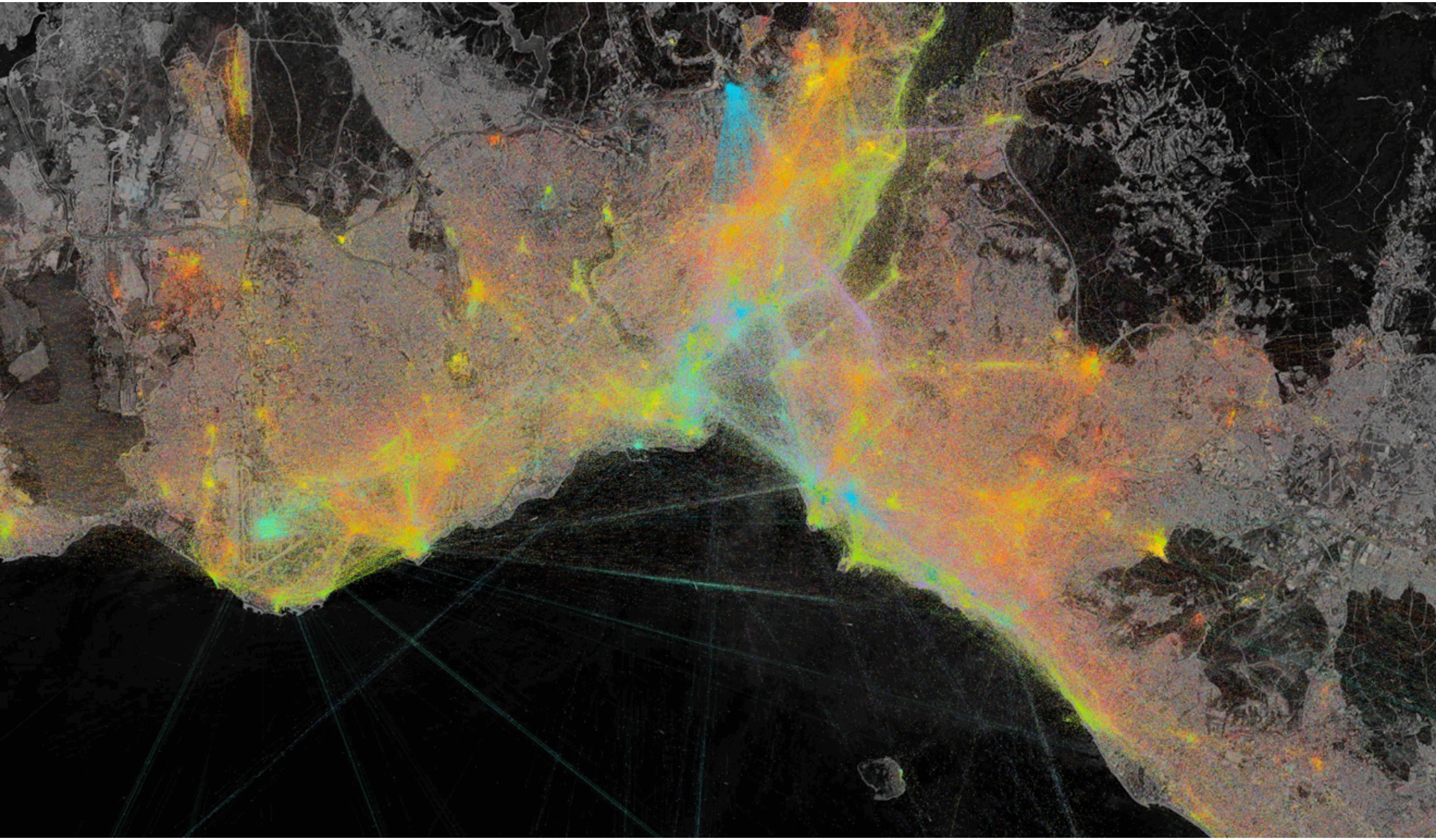
Foursquare was founded by Crawley and Naveen Selvadurai in 2009.

After a short race against Gowalla that ended in 2011 when the latter was acquired by Facebook, Foursquare is today the main player in the area.

Foursquare’s vision

After five years of intense crowdsourcing that generated billions of check-ins, Foursquare is evolving to become the search engine of the city.

Sensing the city



https://www.youtube.com/watch?v=pnkD7OnvCgY

Crowdsourcing never ends...

Data collected from the activity of Foursquare users is not about who goes where ...

It is about ''dressing'' the urban landscape and its elements (places, neighborhoods etc.) with digital information that is accessible by today's communication tools.

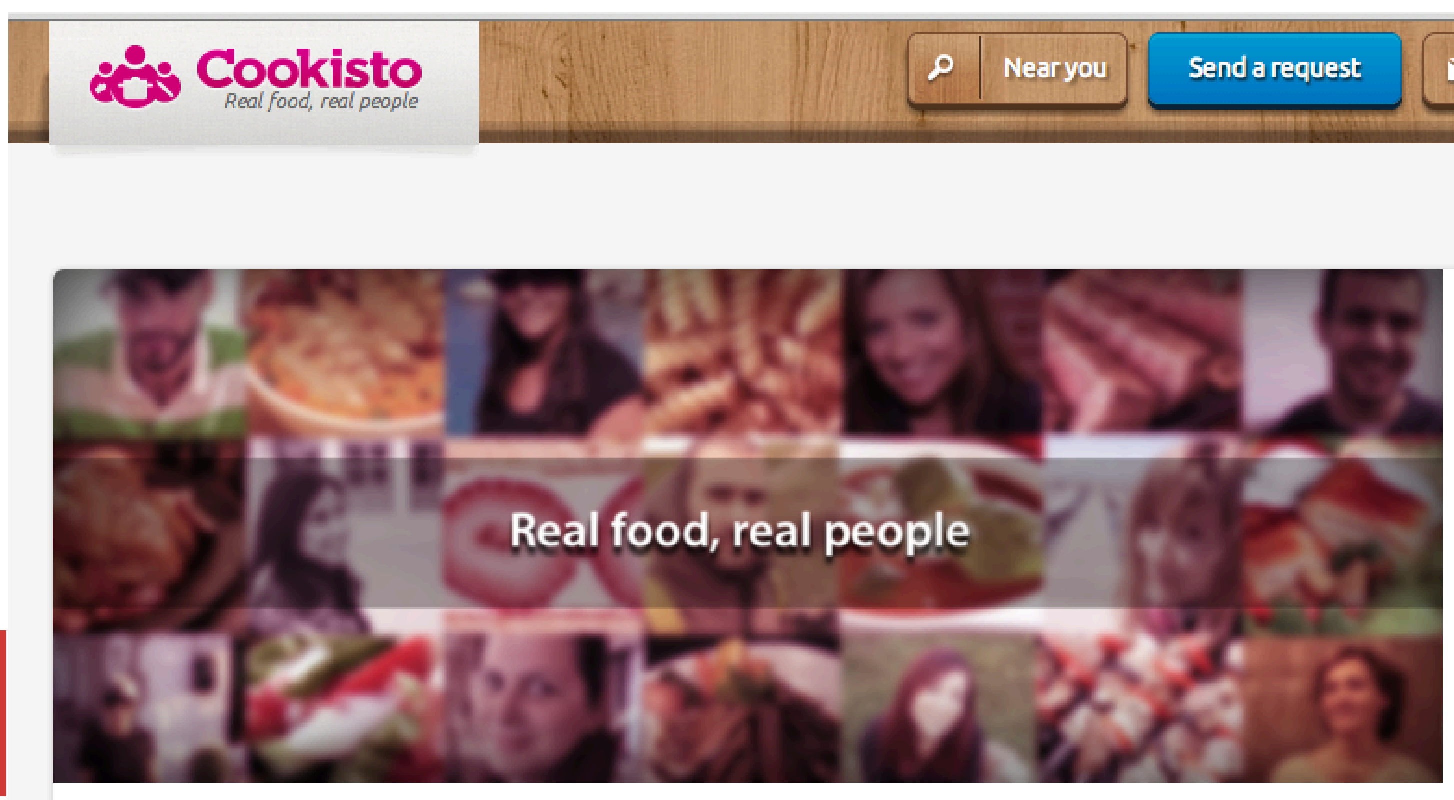
Not only we learn about the location of places, their temporal activity patterns or their categories ... but whatever one can imagine about describing them.

The sharing economy

Mobile web and digital mapping technologies have brought a revolution on private resource utilization....

UBER

Rent out your extra space to travelers or cook for your neighbors … this is just the beginning of a big revolution that brings together the physical and digital space.



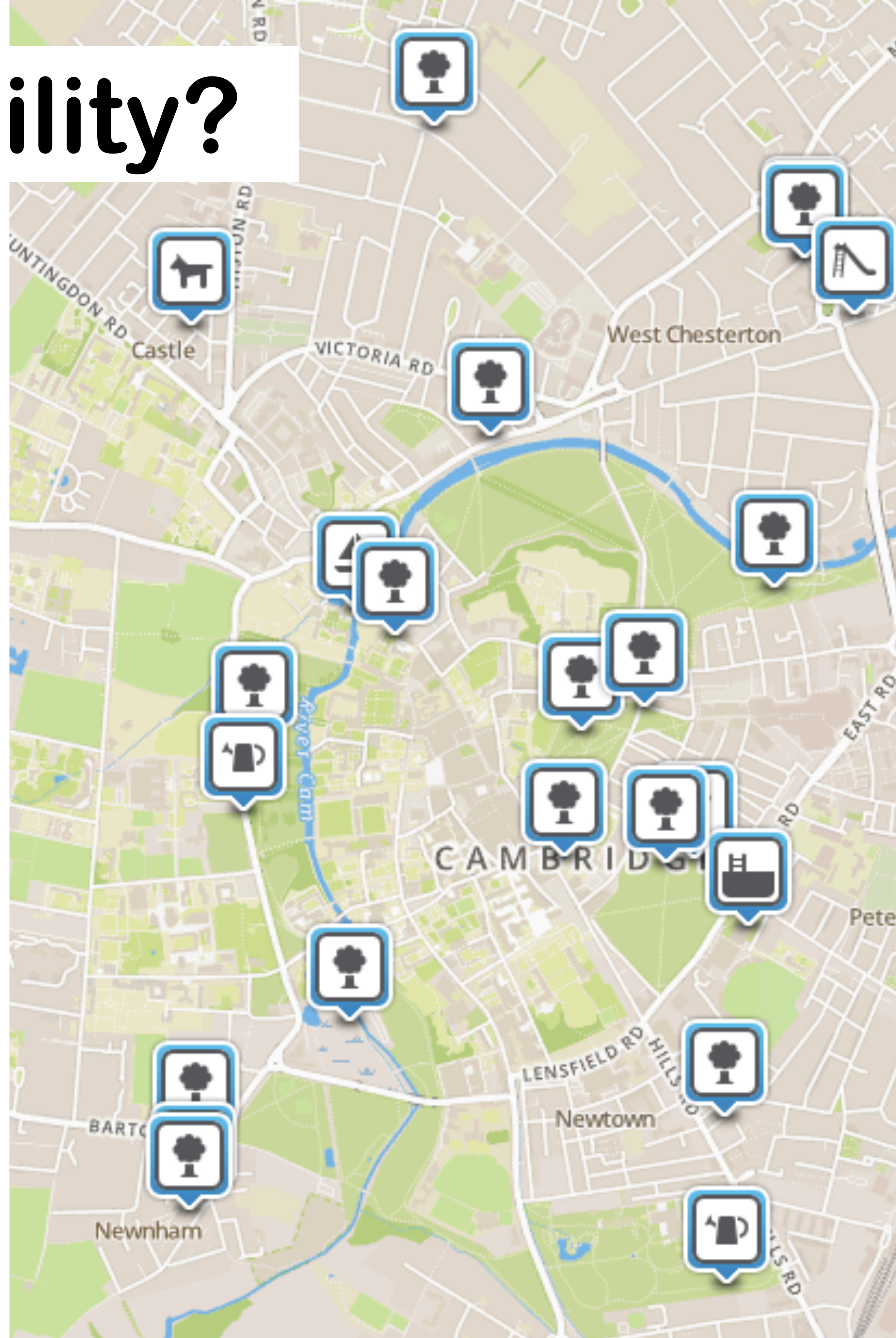
Forky

spoon|rocket
most convenient meal ever

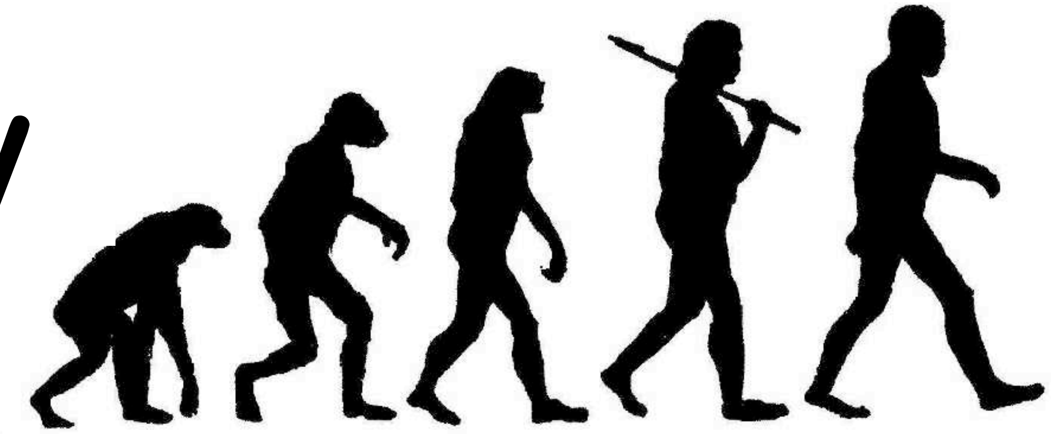
Why human mobility?

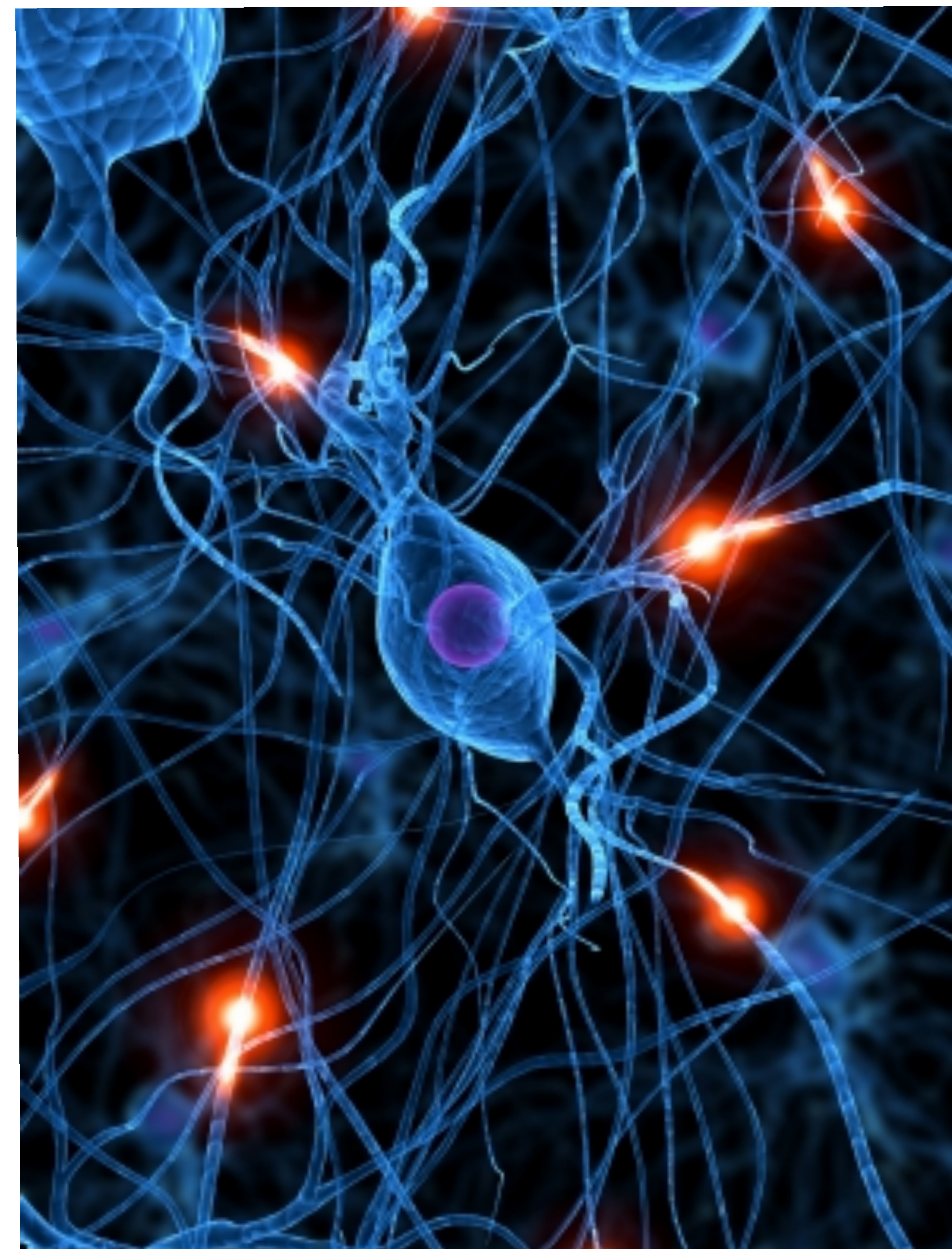
Urban planning :
understand the city and optimise services

Mobile applications and **recommendations:**
study the user and offer services



Why human mobility





Animals have the capacity to search and navigate across space

We can understand how our brain is wired

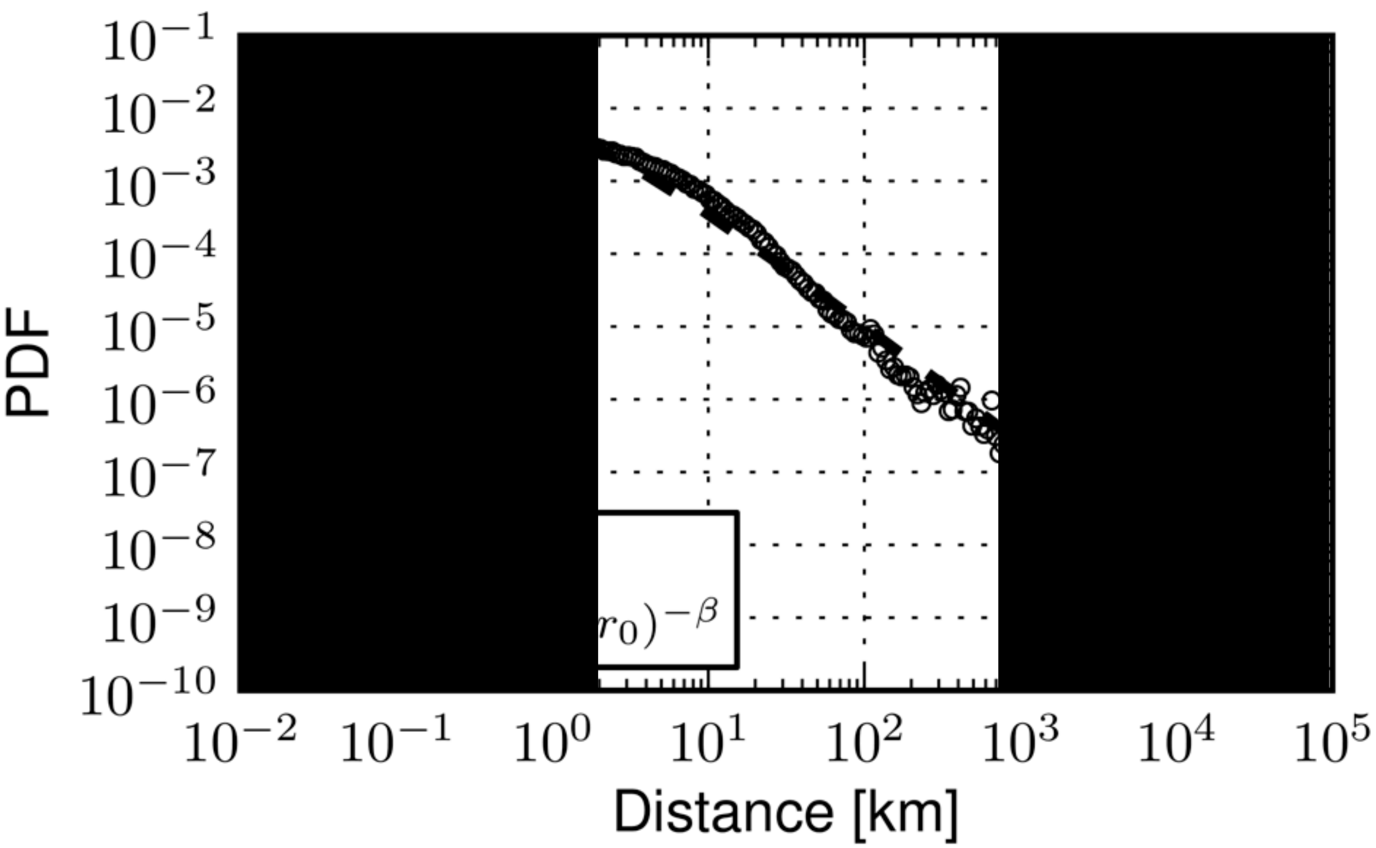
Have we really left the monkey ?

Data on human movement…

Mobile Social	VS	Cellular
GPS accuracy ~ 10 meters		BTS Tower Accuracy ~ 1KM
Global Coverage		Country Coverage
Publicly Available		Private / Corporate

Power-law tales ...

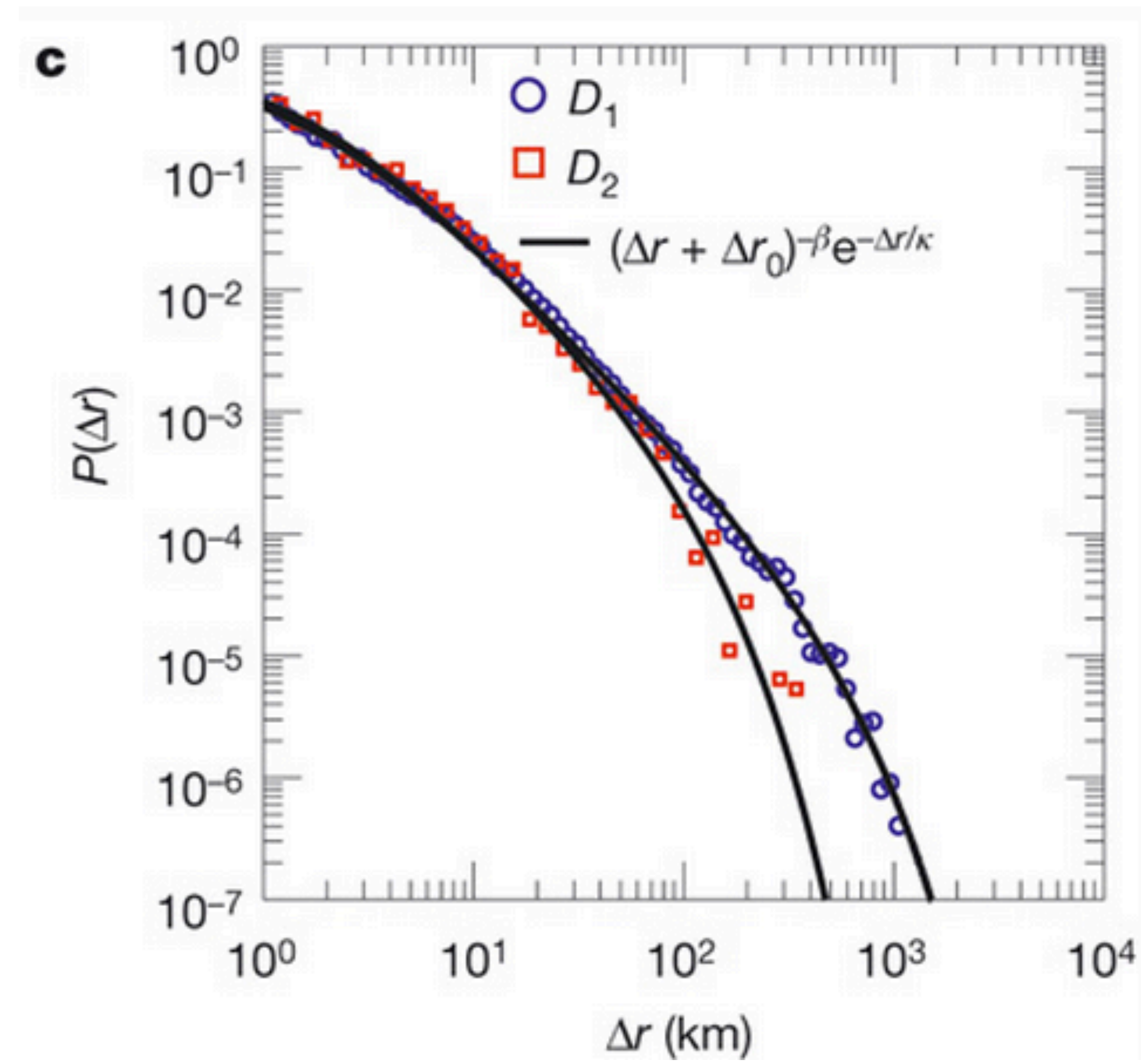
Mobile Social Network Data



$(\Delta r + \Delta r0)^{-\beta}$

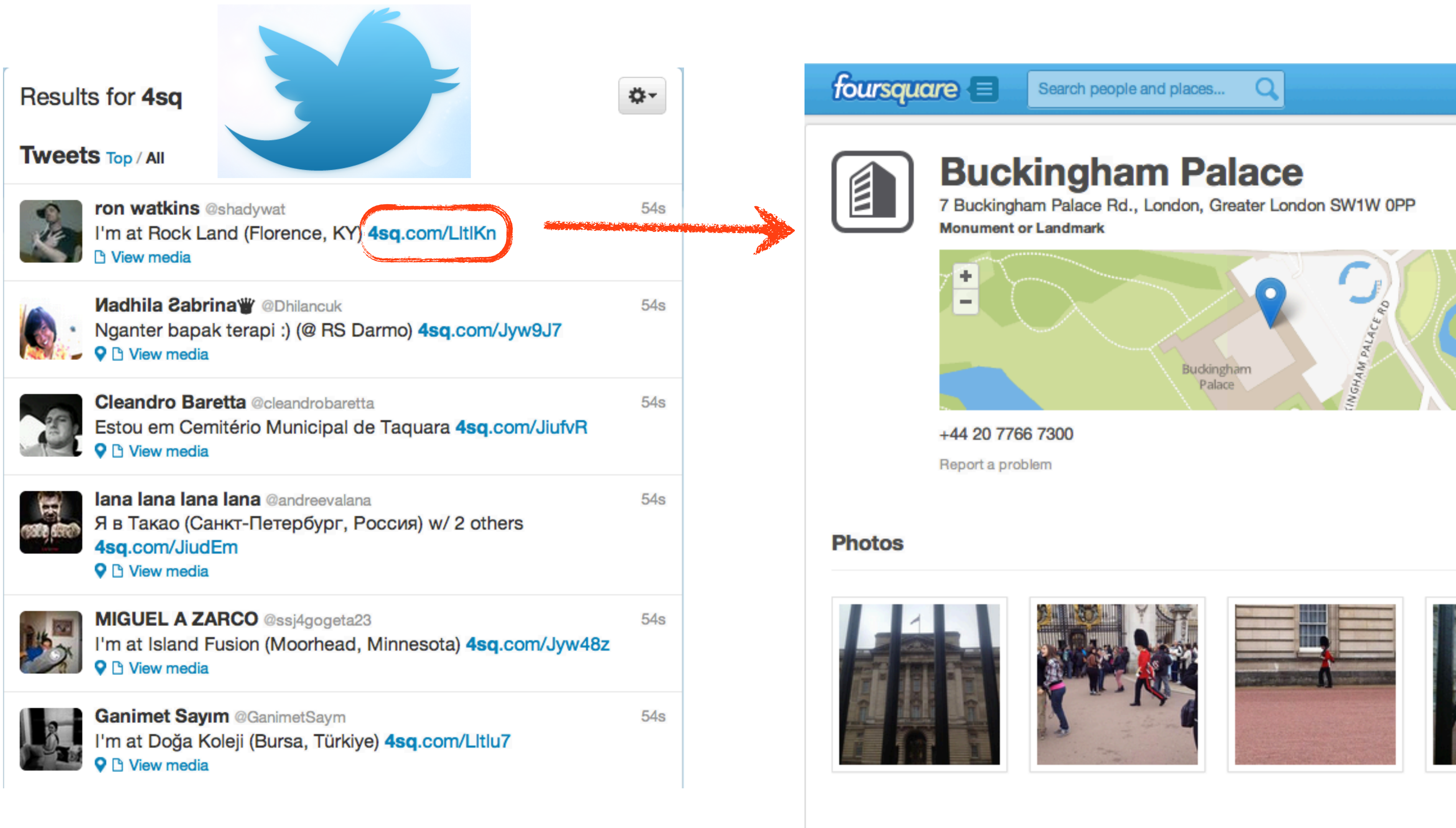
exponent β = 1.50

Nature **453**, 779-782(5 June 2008)



exponent β = 1.75

The Data Crawling Combo …

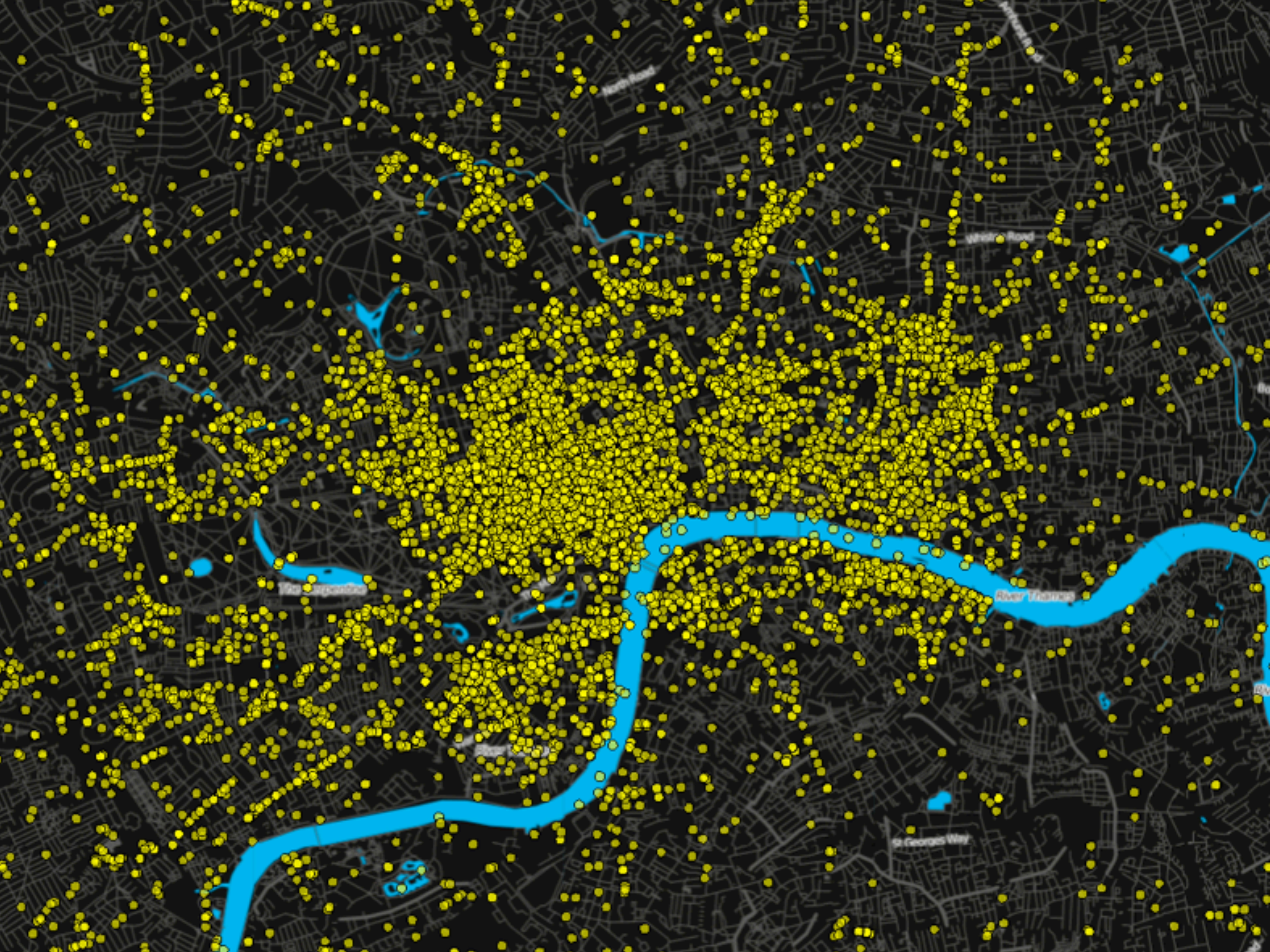


Dataset Statistics

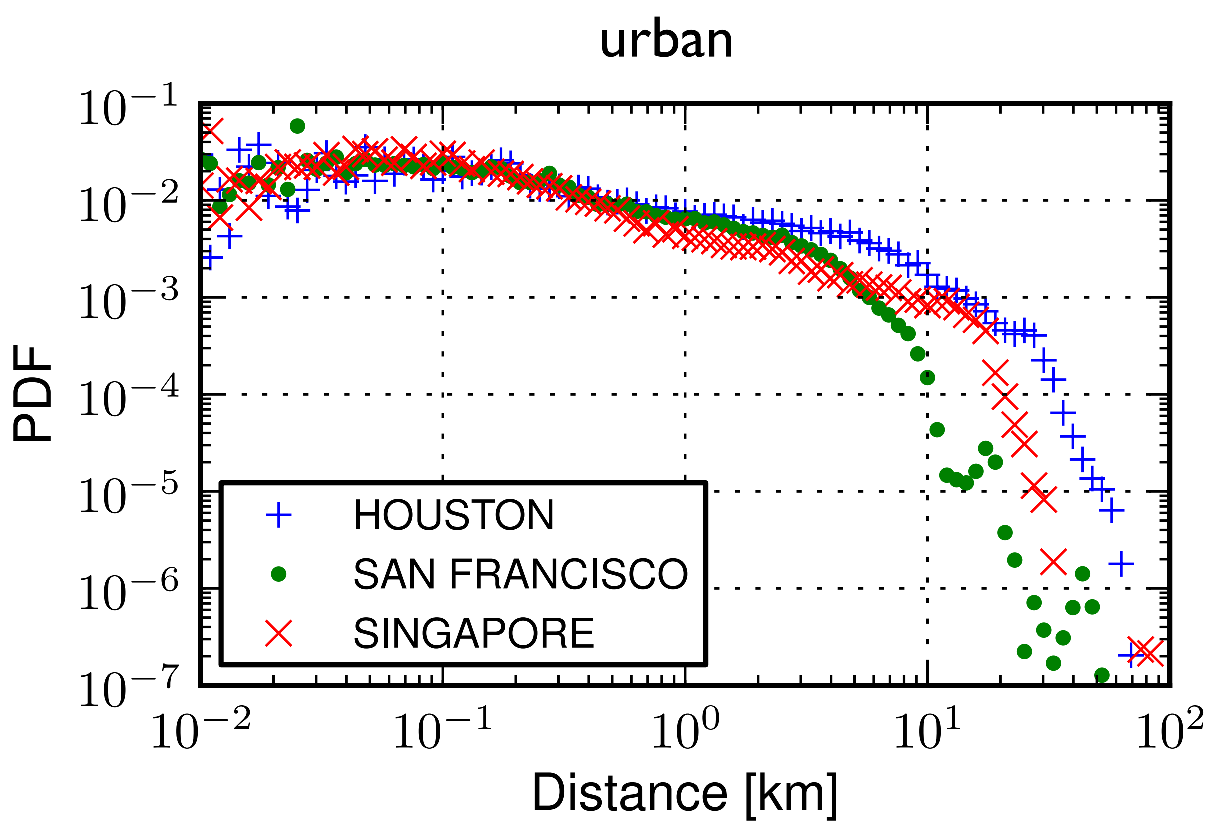
925,030 users around the globe over a period of 6 months in 2010.

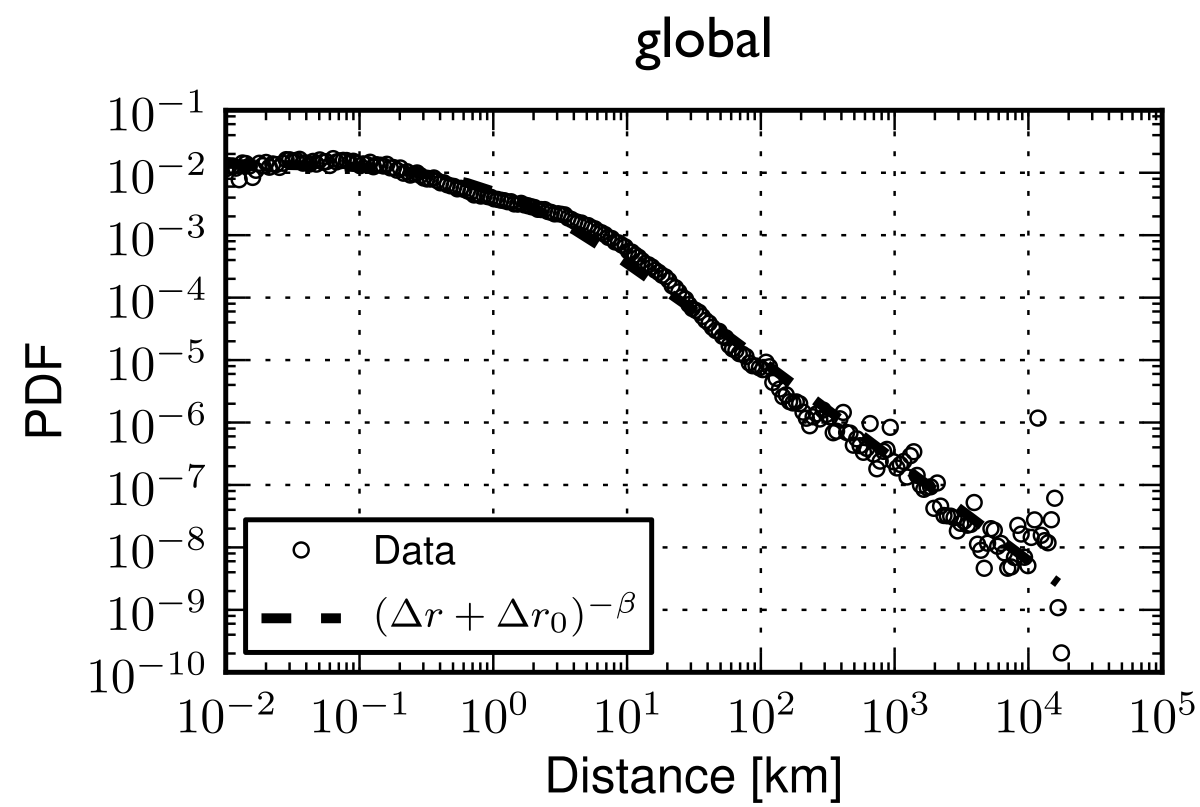
34 Cities that span 4 continents and 11 countries.

A database of ~5 million recorded places.


North Road
River Thames
St Georges Way

Urban vs Global mobility



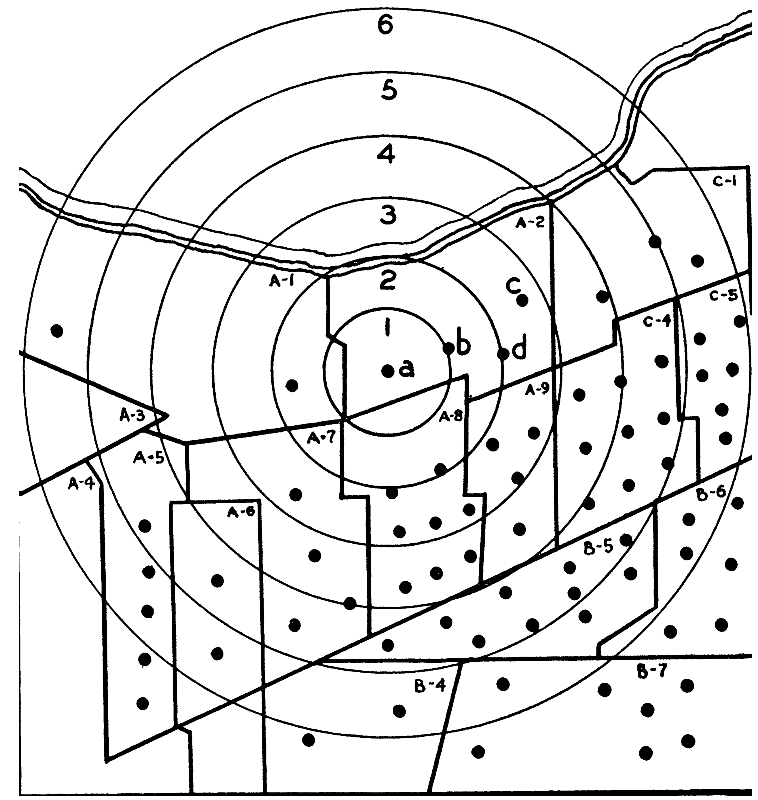


Power law kicks in at 18.42km!!!


VIVA LA RAZA
RACE
BLACK LIKE ME
STATUE OF LIBERTY
HOSTAGE
The Uncertain Giant

Samuel A. Stouffer

Stouffer's **law of intervening opportunities** states, *"The number of persons going a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities."* *

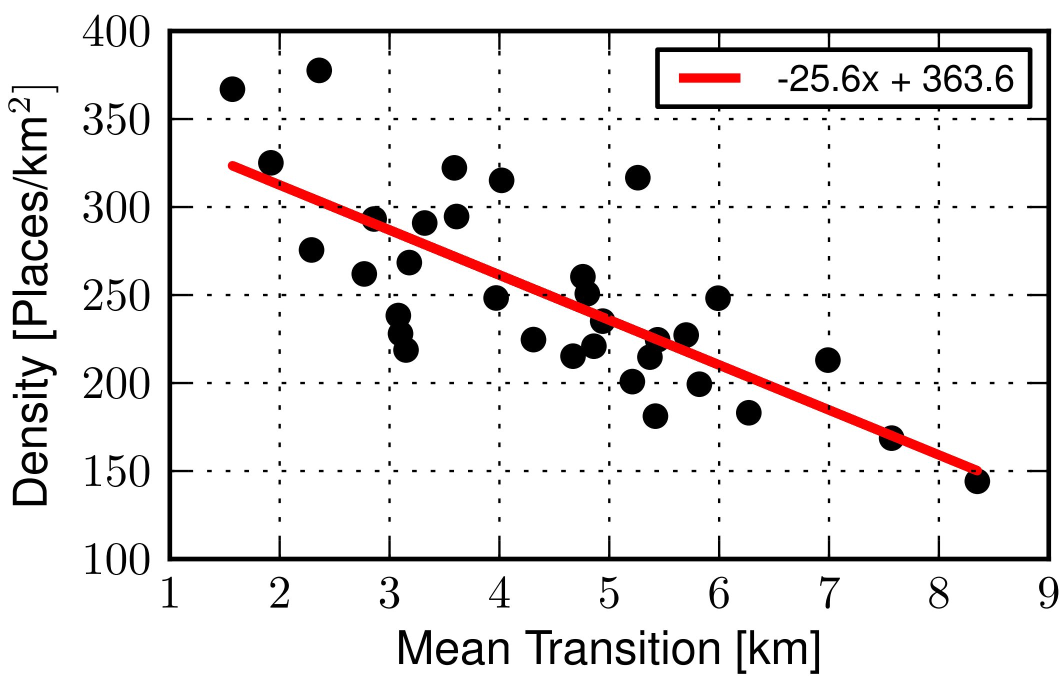


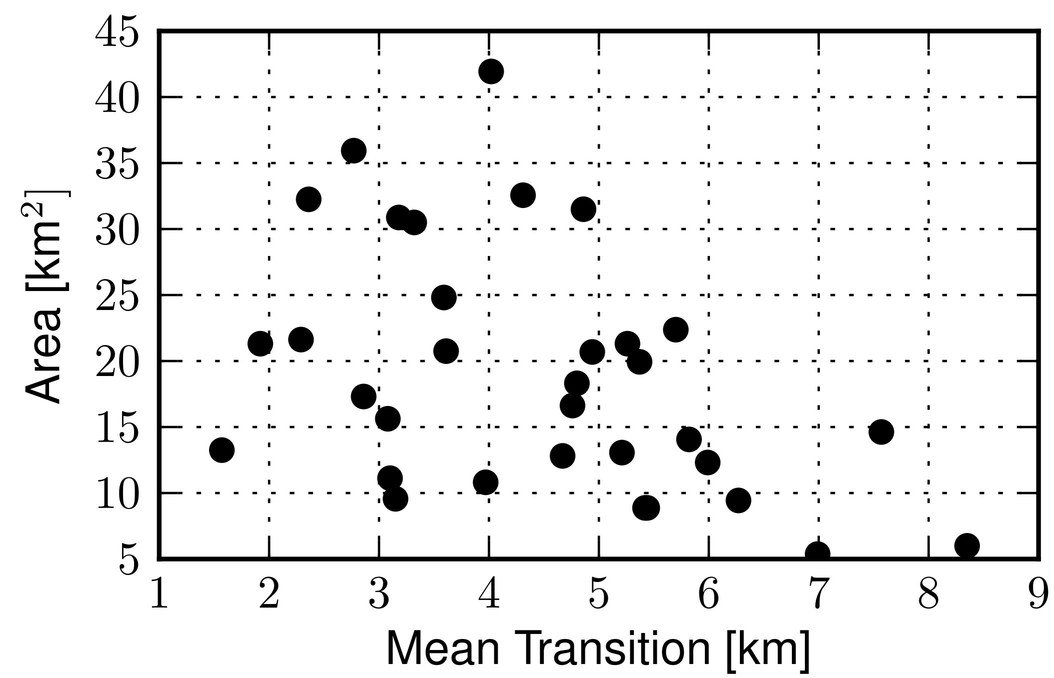
- Empirically proven using data for migrating families in the city of Cleveland.

- We investigate the plausibility of the theory for urban movements in Foursquare.

* S. Stouffer (1940) Intervening opportunities: A theory relating mobility and distance, American Sociological Review 5, 845-867

The importance of density

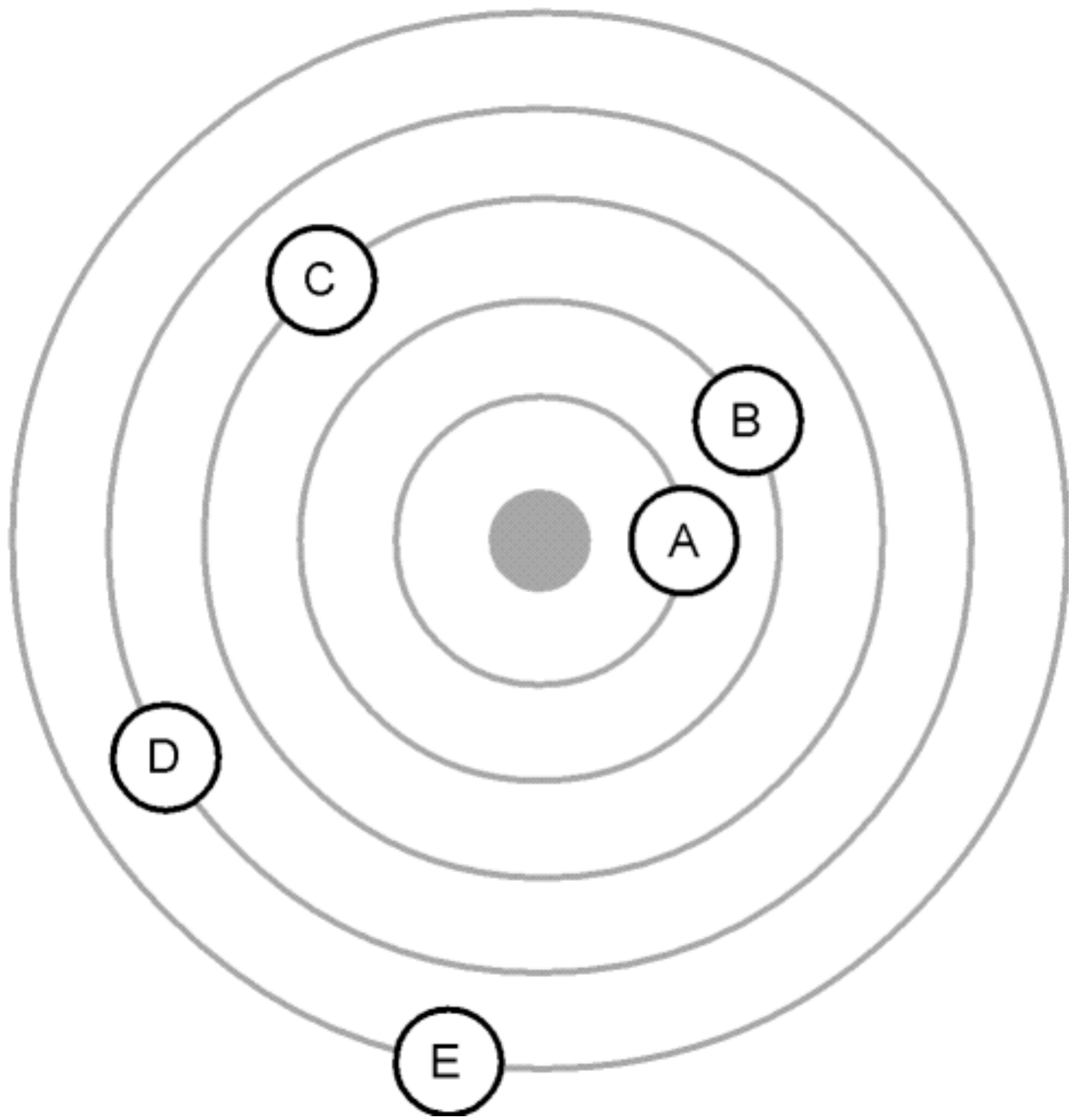




- Stouffer’s Theory of Intervening Opportunities motivated us to inspect the impact of places(=opportunities) in human mobility.

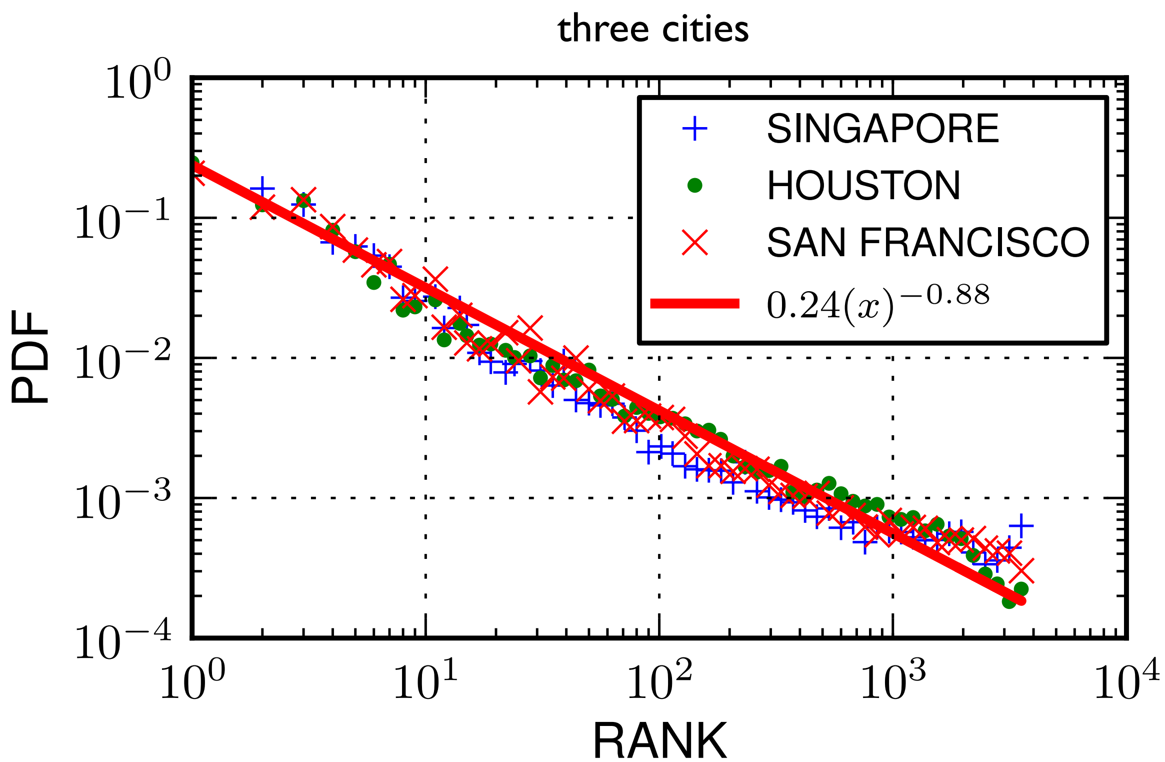
- Place density by far more important than the city area size with respect to mean length of human movements (R^2 = 0.59 and 0.19 respectively).

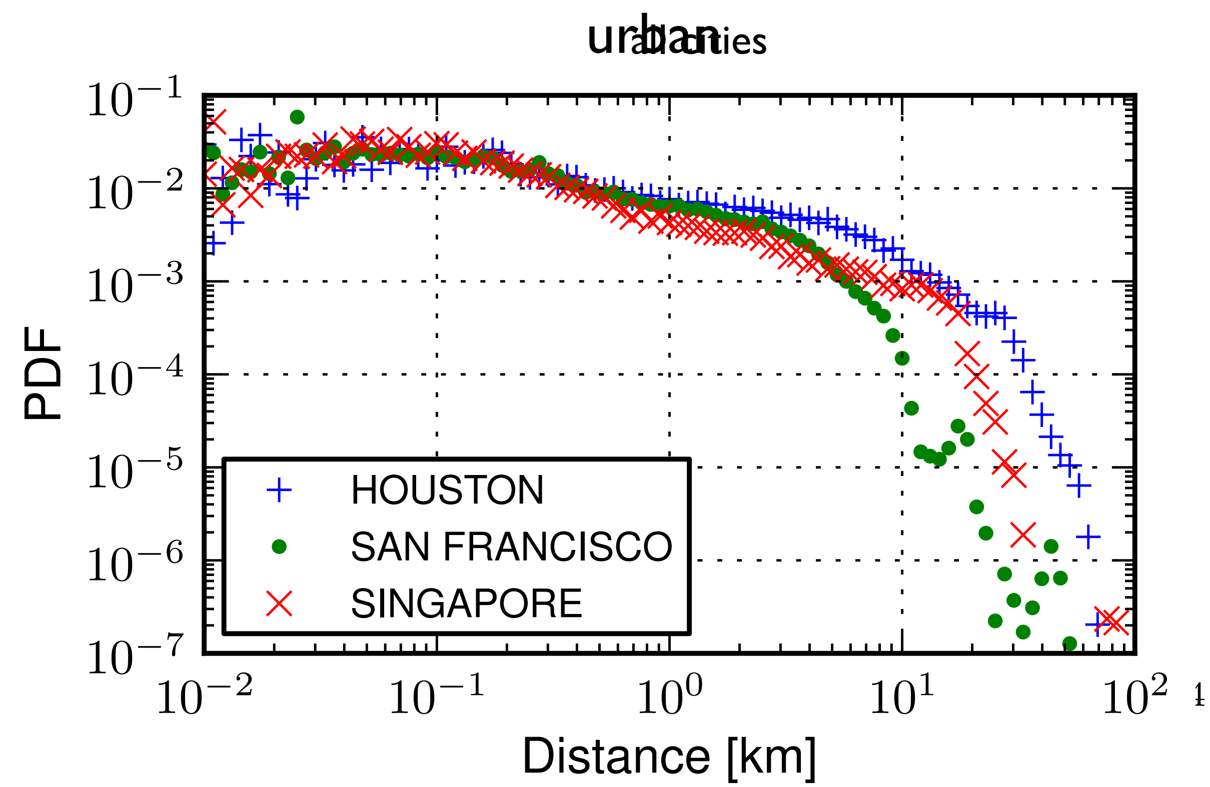
Defining Rank-Distance



$$rank_u(v) = |\{w : d(u, w) < d(u, v)\}|$$

Rank universality



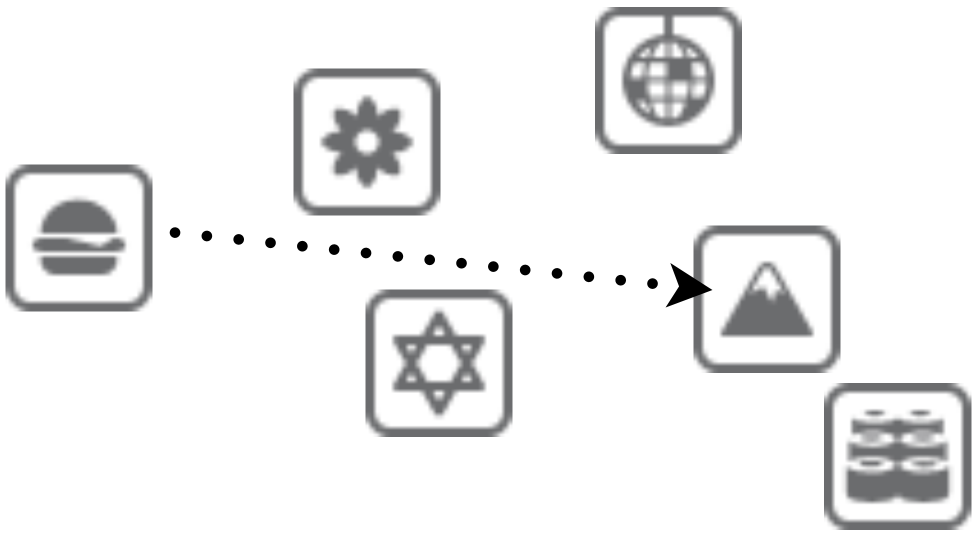


The rank of all cities collapse to a single line.

We have measured a power law exponent $\alpha = 0.84 \pm 0.07$

A new model for urban mobility

soil...



and mind!

$$Pr[u \to v] \propto \frac{1}{rank_u(v)^a}$$

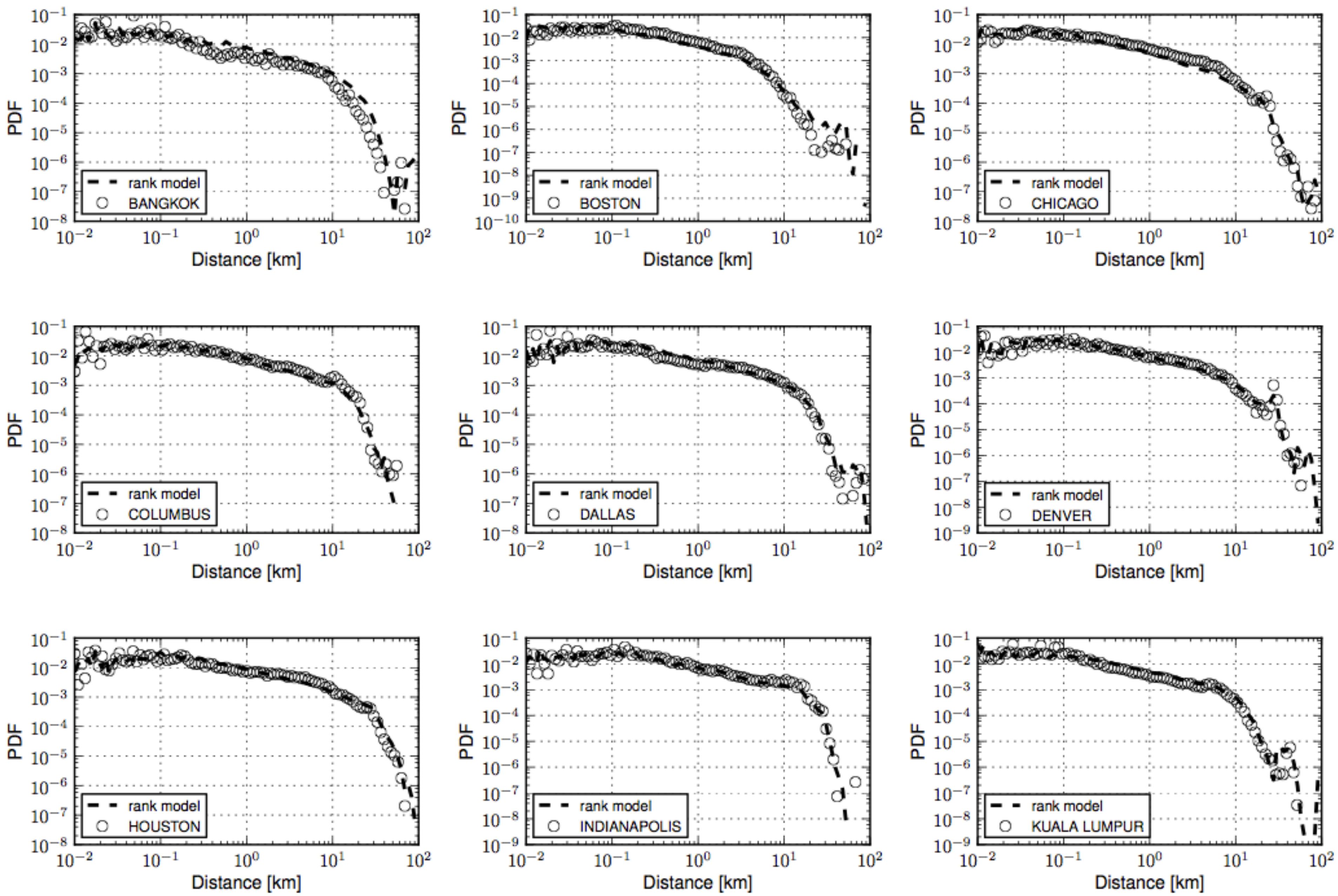
Set … and go!

for all cities we have used the average value $\alpha = 0.84$ for the rank exponent.

all places in the city used as potential starting points for our agents.

the rank element is universal, only the set of places differs from city to city.

Simulation Results …

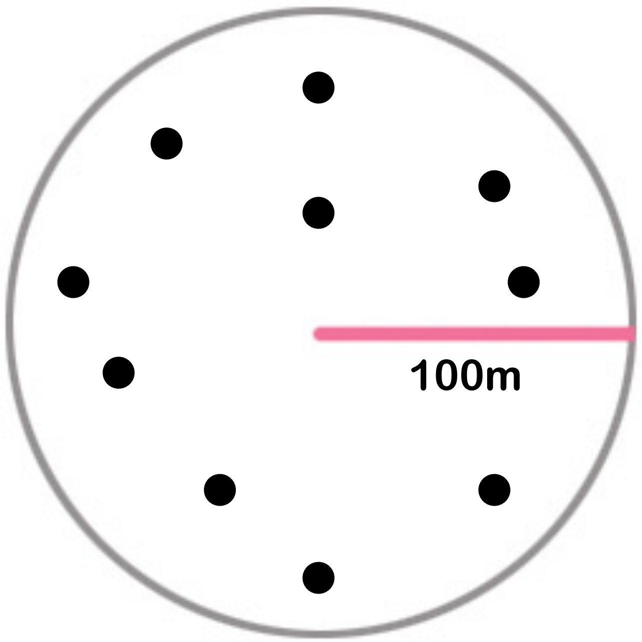


“Zero” Gravity

We have also built a gravity model in the urban context!

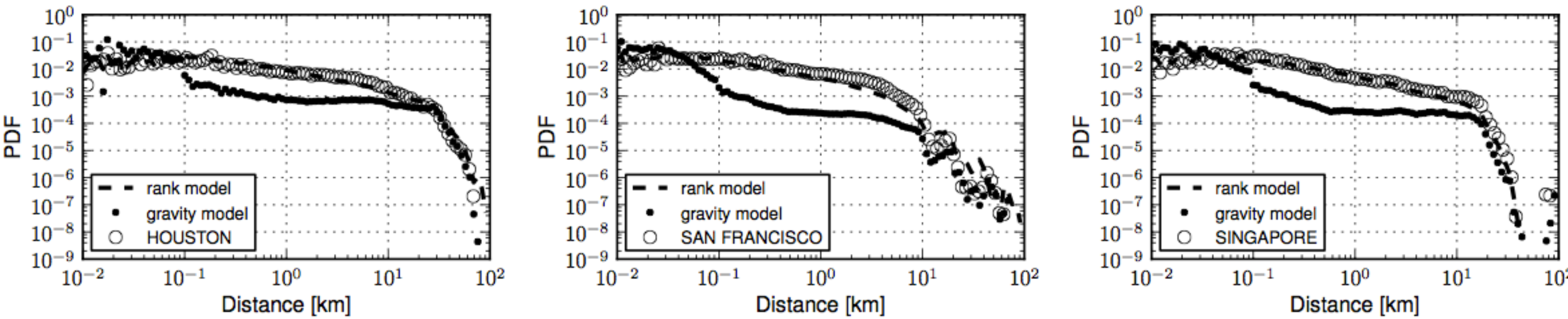
$$P_g[u \to v] \propto \frac{m_u . m_v}{d(u,v)^b}$$

Issue #1: how do we define “mass” in the urban context.



Issue #2: how do we set its parameters?

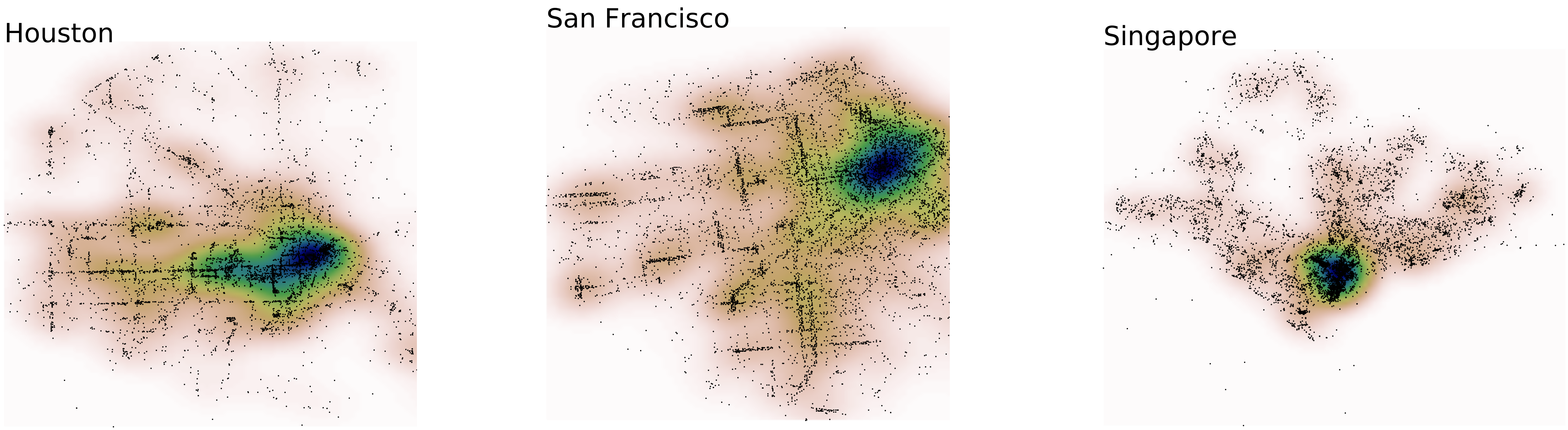
Rank vs Gravity



Rank is simpler and achieves better quality fits for all cities.

Gravity overestimates short transitions ...

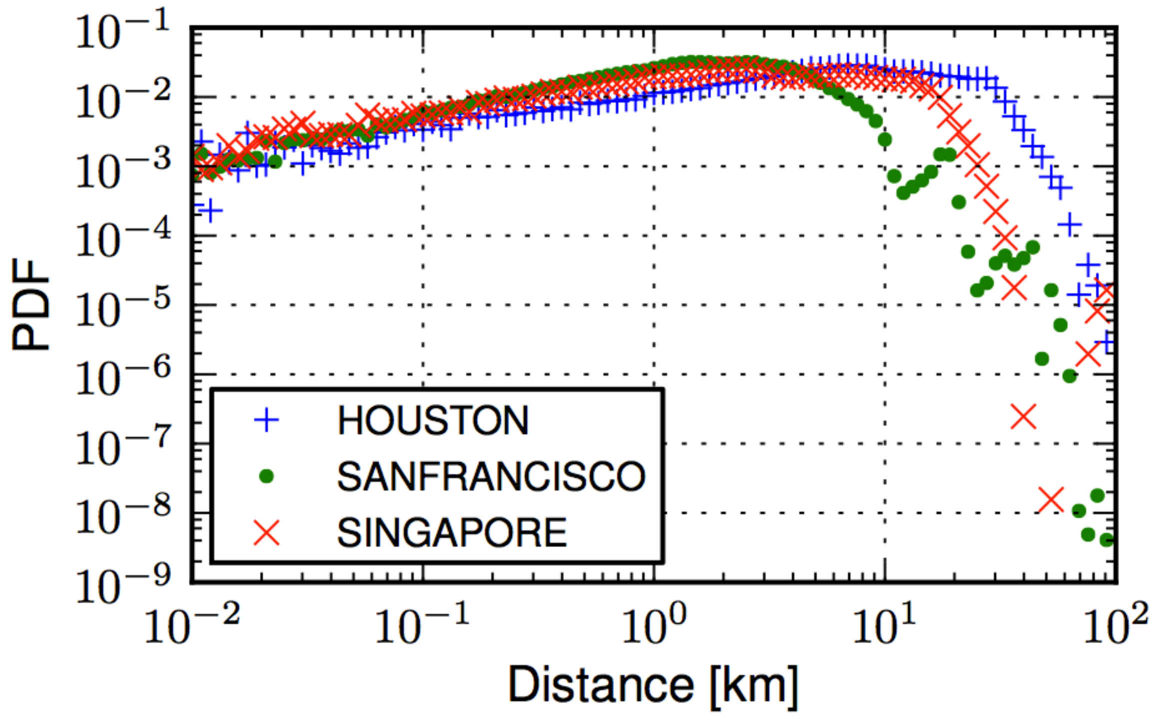
The importance of Geography

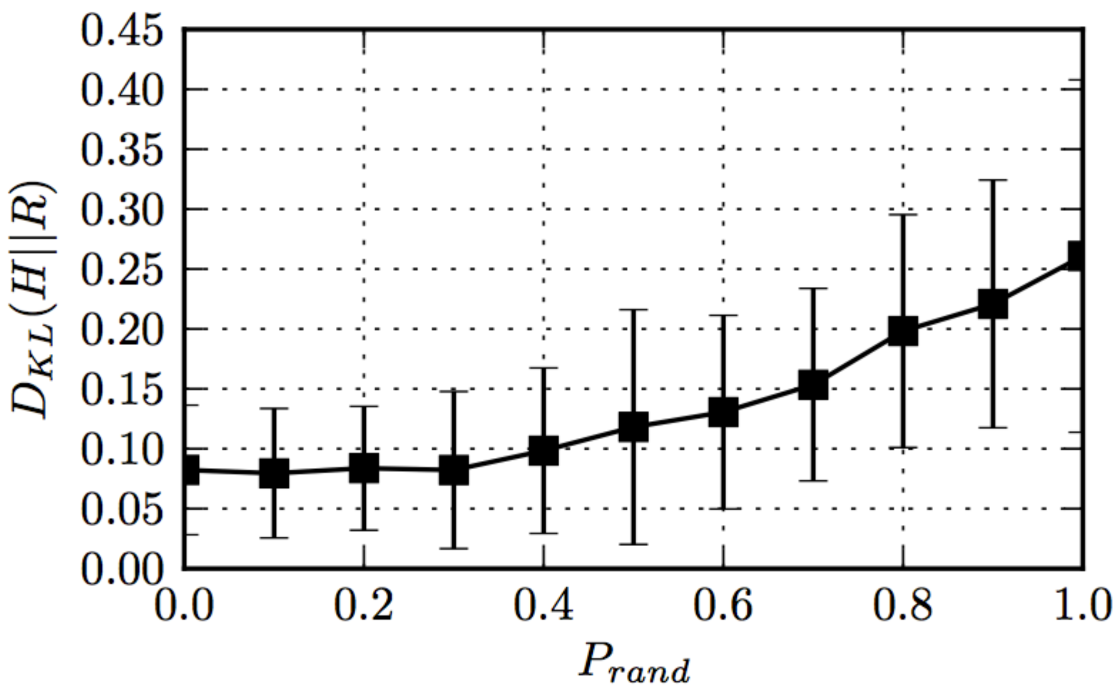


Heterogeneities observed in human mobility is due to geographic variations. Cultural, organisational or other factors do not appear to play a role in urban movements.

The rank model, although simple, can cope with the complex spatial variations in densities observed in urban environments.

Shuffling Urban Geography





Open Questions

- While we have managed to fit human movements in an urban setting, the spatial distribution of places remains unknown.

- Can we model global displacements using the same model?

The Multi-Dimensional Check-In

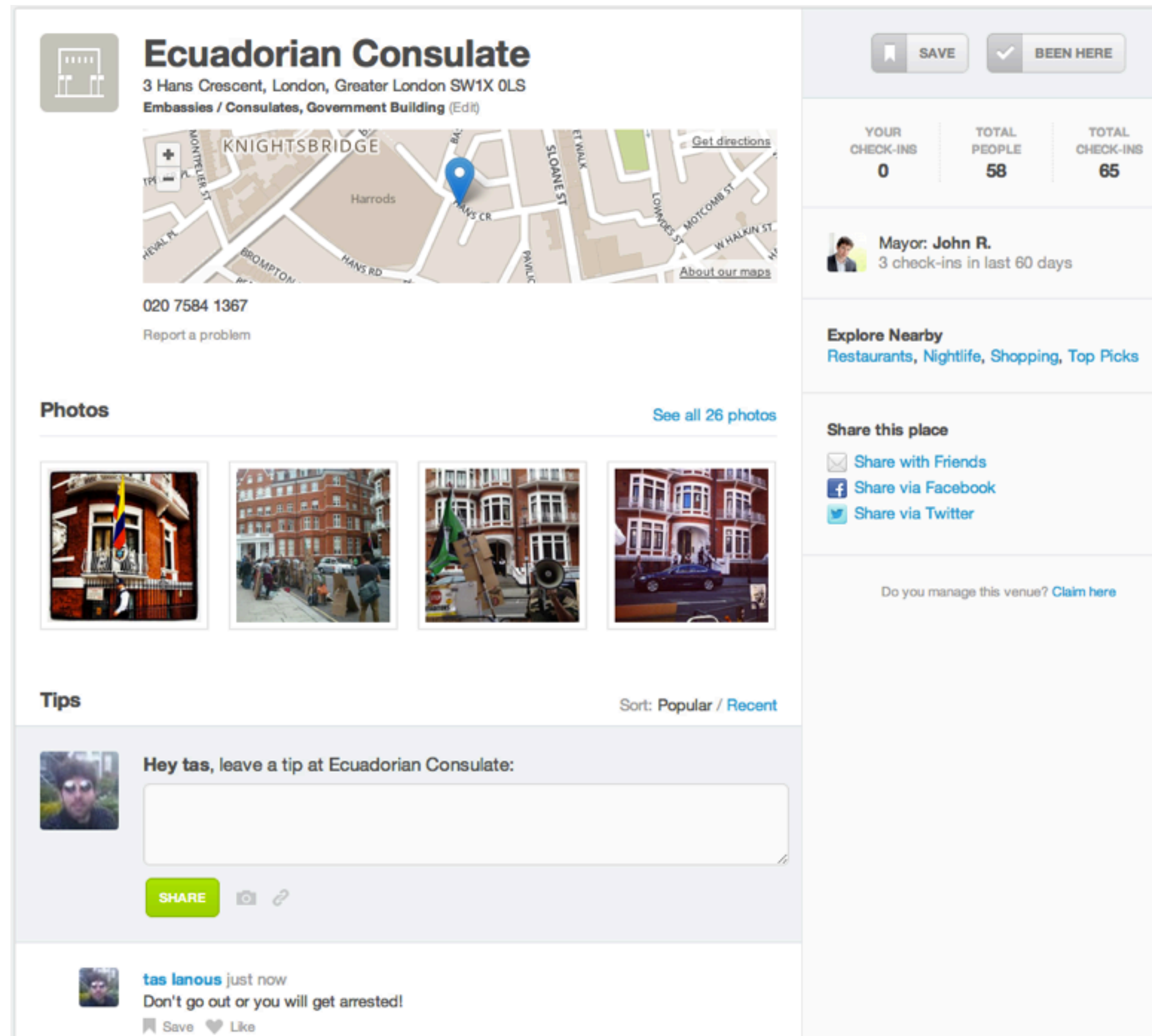
Social

Geo

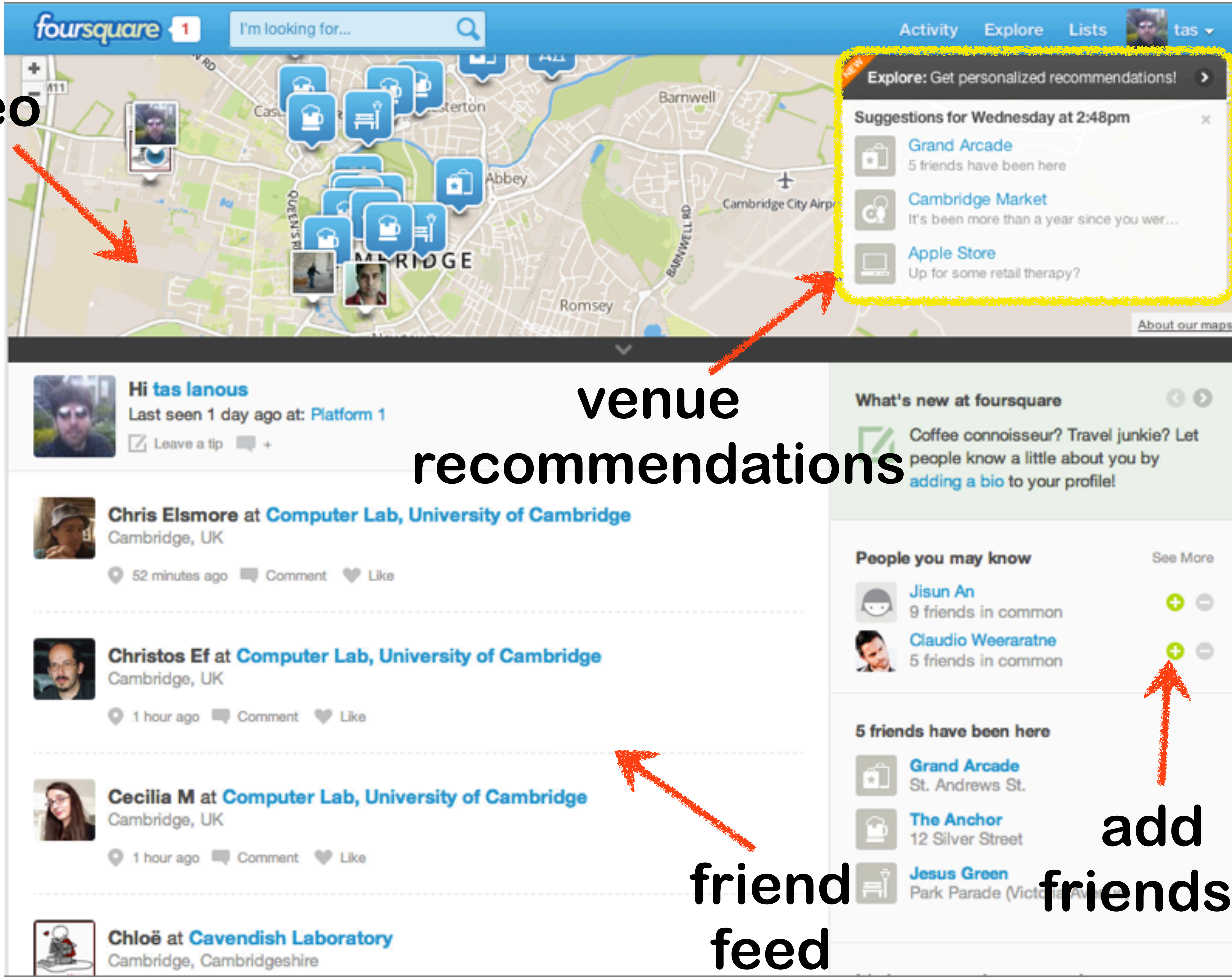
Temporal

Mobile

Linguistic
Multi-Media
Game Features



geo

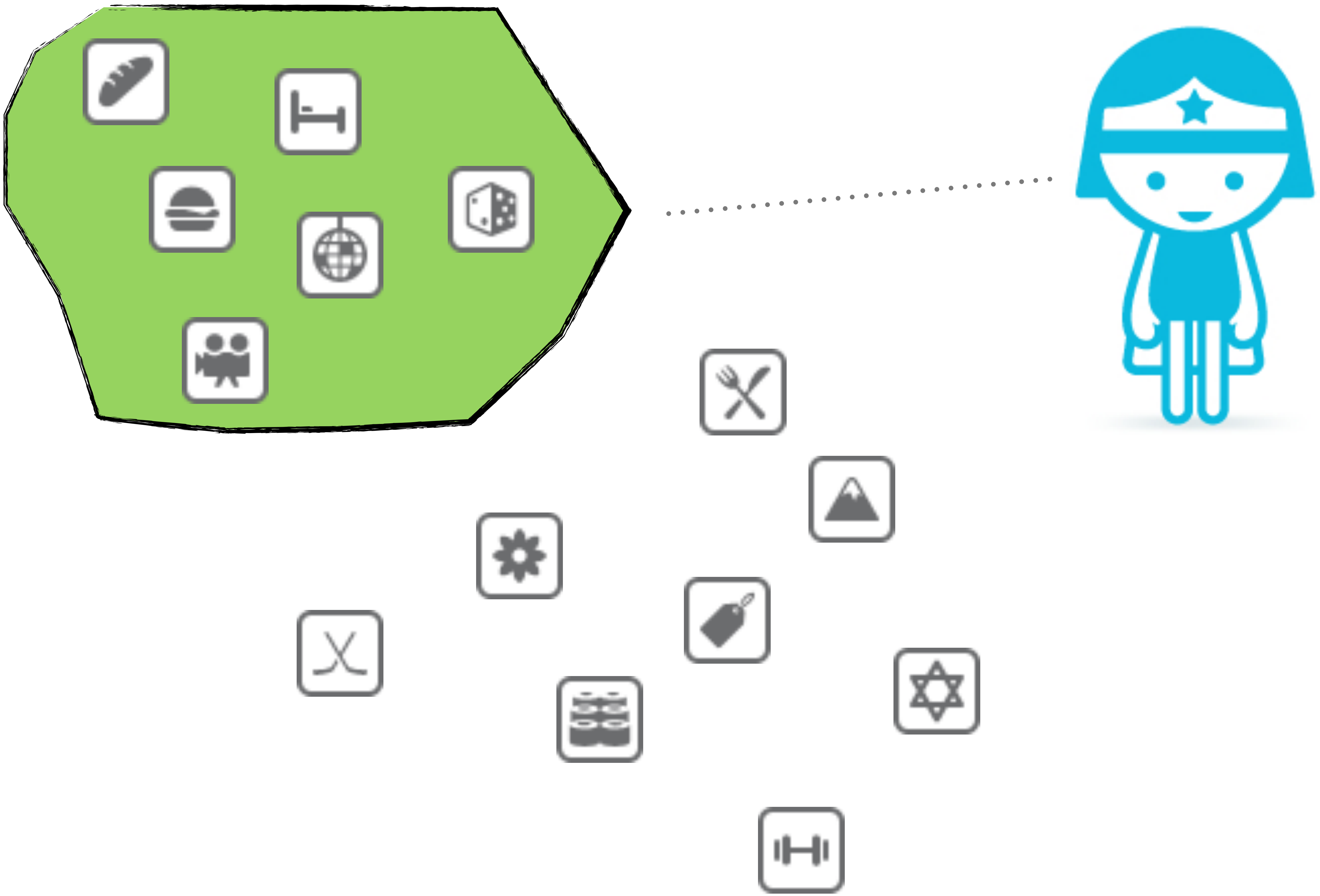


venue recommendations

friend feed

add friends?

New Venue Recommendations



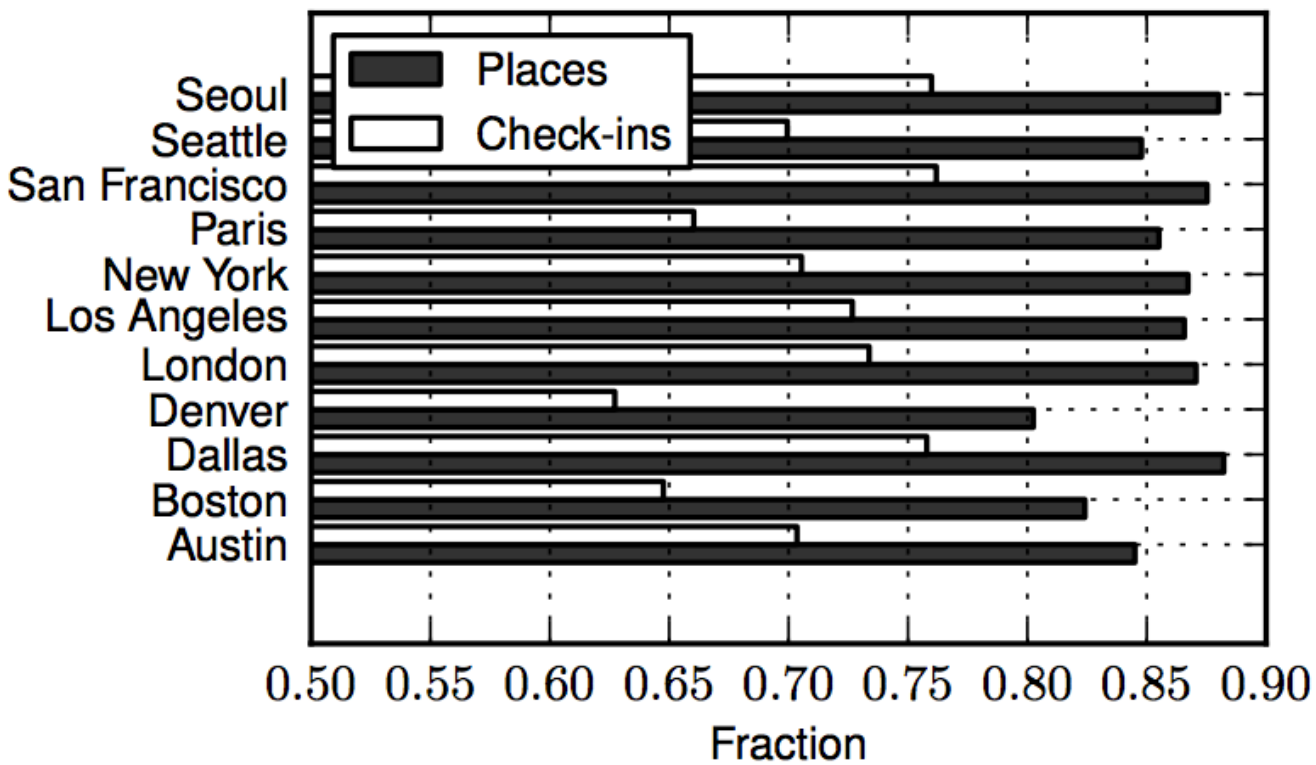
Challenges

LBSN Data represents a new paradigm: **sparse, geographic, implicit feedback.**

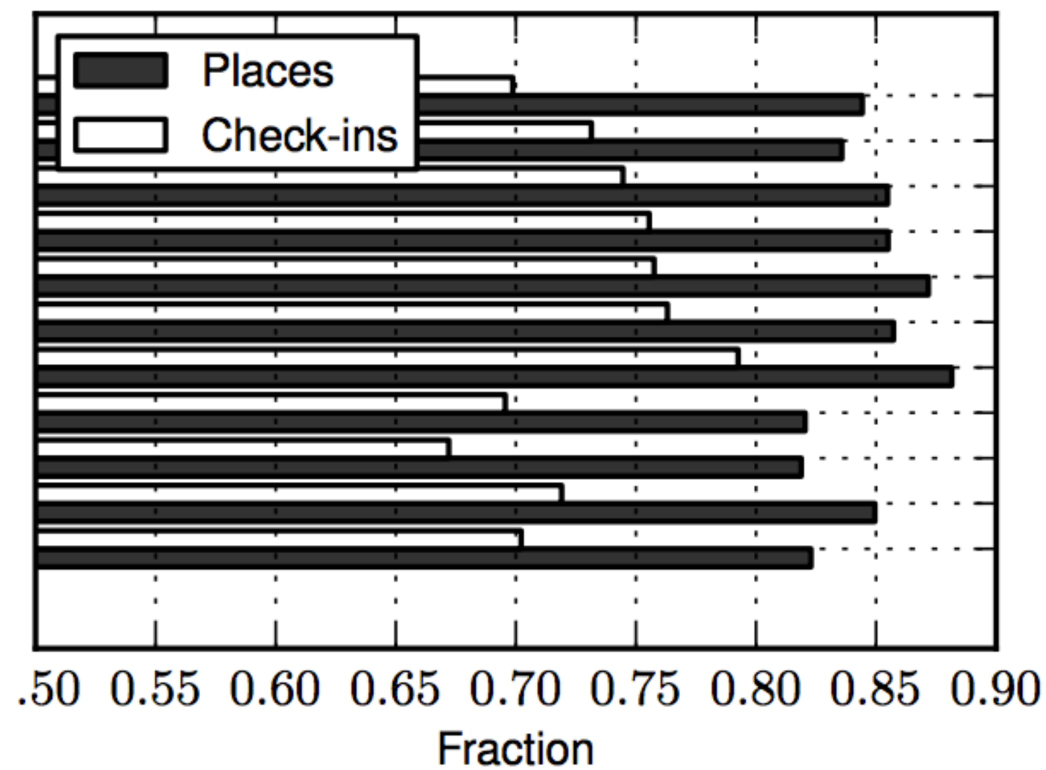
Recommending new **venues** is similar to recommending new **links** in social networks.

Check-Ins at New Venues

foursquare



gowalla



Recommending new, unvisited places to users has an important value, as they often seek to discover new locations.

80-90% of visited places are new places!
60-80% of check-ins occur at new places!

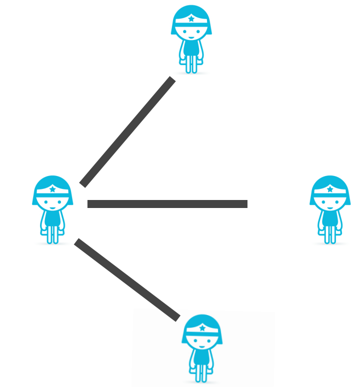
Recommendation Strategies

Popularity	rank places using number of check-ins by any user
Content Filtering	exploit venue type preferences(cafe, bar...)
Social Filtering	rank places using number of check-ins by friends
Home Distance	recommend based on geographic distance from home

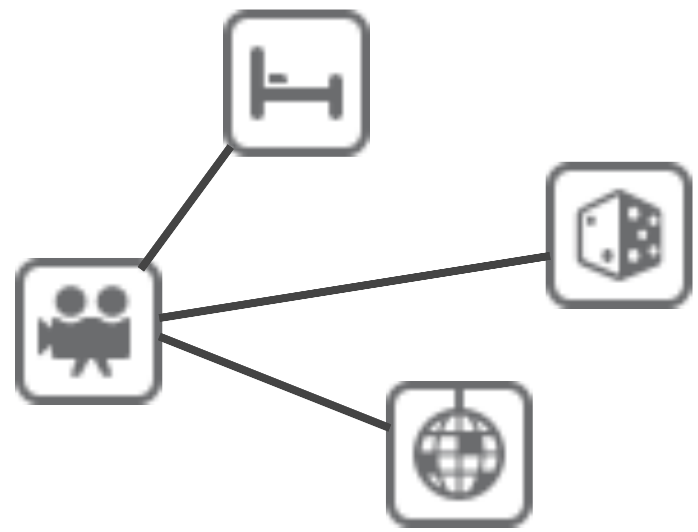
Collaborative Filtering

K-NN user similarity

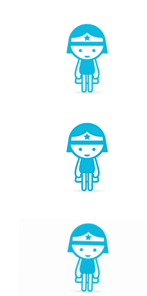
$$\hat{r}_{ij} = \frac{\bar{c_j}}{|\Phi_j|} + \frac{\sum_{n \in U} ((c_{nj} - \bar{c_n}) \times s_{in})}{\sum_{n \in U} s_{in}}$$



Place Network (item similarity)



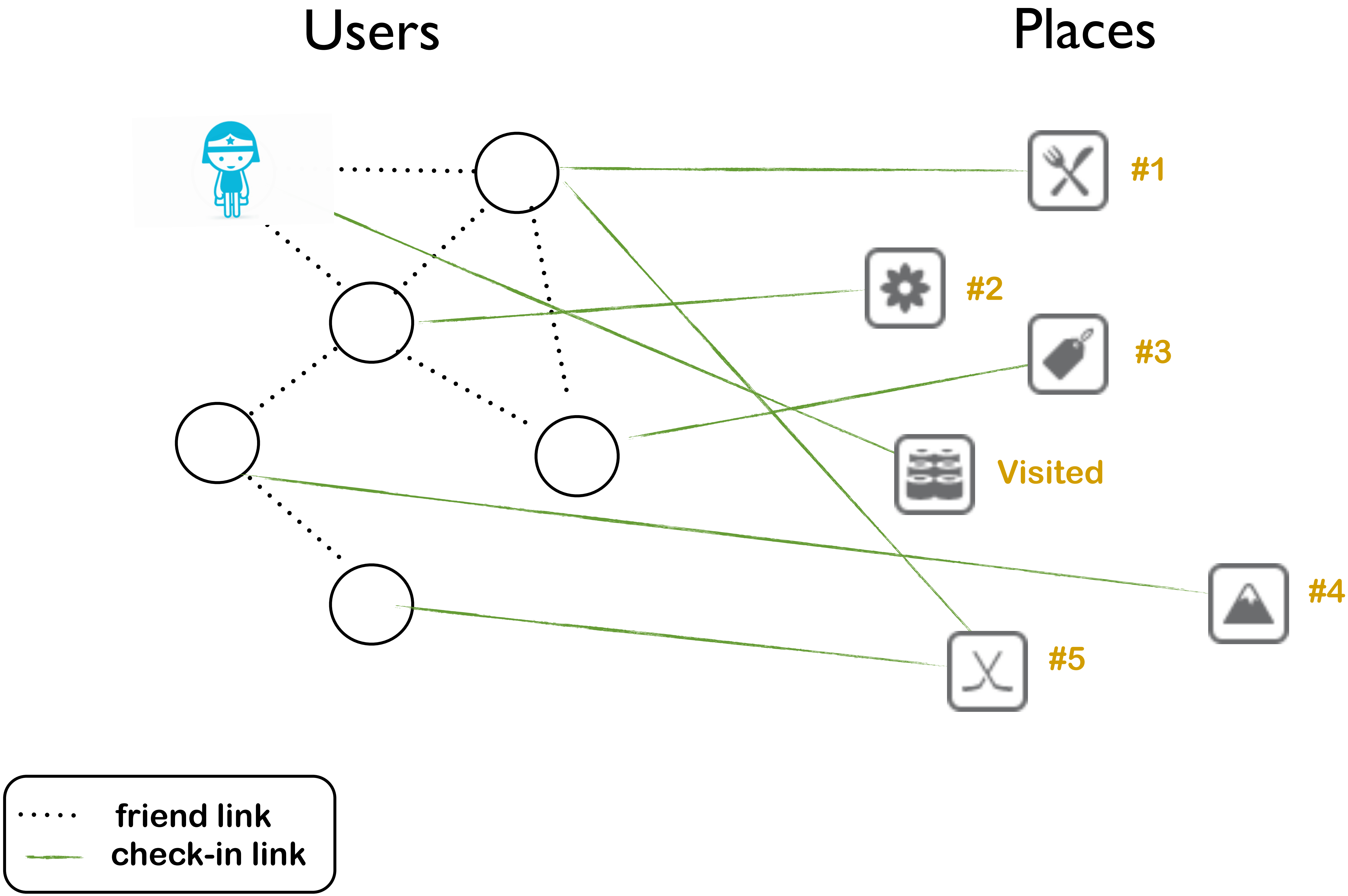
Matrix Factorization



2	3	?
?	3	2
1	2	?

$$E = \sum_{i \in U} \sum_{j \in \Theta_i} (c_{ij} - \mathbf{p}_i \mathbf{q}_j)^2 + \lambda(\|\mathbf{p}_i\|^2 + \|\mathbf{q}_j\|^2)$$

A Random Walk Around The City

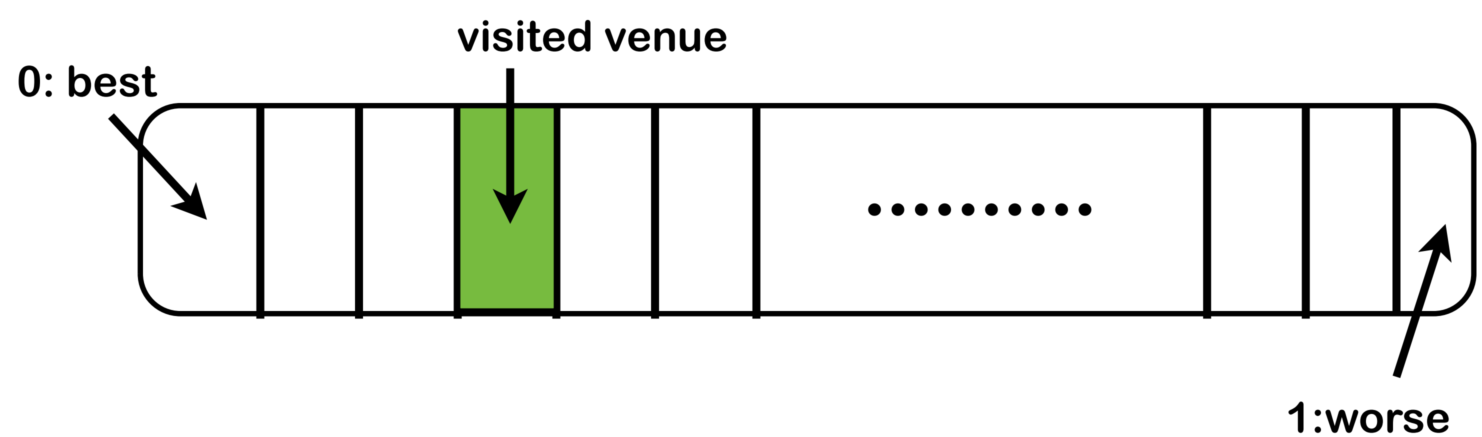


Evaluation

Test and training sets — Monthly Cross Validation

Output — Personalized Venue Recommendation Lists

Average Percentile Rank



Results

Method	Average Percentile Rank
Random Walk	**0.217**
k-NN	0.443
Random	0.500

Lessons:

1. **Collaborative Filtering Fails!**
2. **Popularity/Content Filt. good!**
3. **Random Walk v.good!**

Popularity VS Random Walk

Lessons:
1. Random Walk>> Popularity in 21/22 cases.
2. Better Results where more data: New York & Austin TOP, Gowalla >> Foursquare!

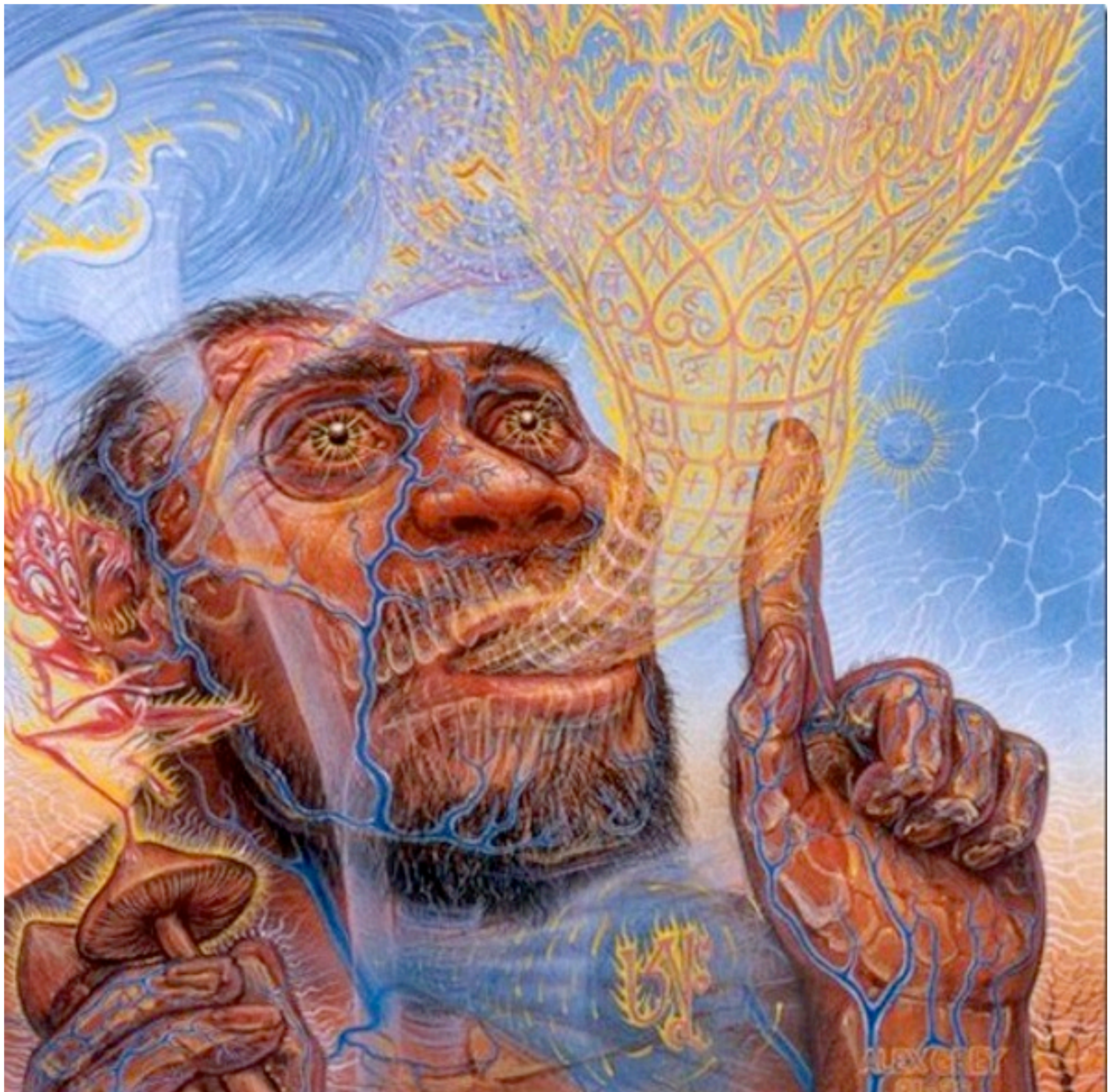
Speculating on RW Victory

Good recommendations can be identified at the trade-off between global and personal behaviour.

Random Walk with Restart achieves that by imposing graph structure (global) and controllable user bias (personalization).

RW resilient to noise! Applicable to users with no check-ins! (cold start).

We then had a vision :)



Predicting your next check-in

thousands of venues in the city to be ranked …. in real time!

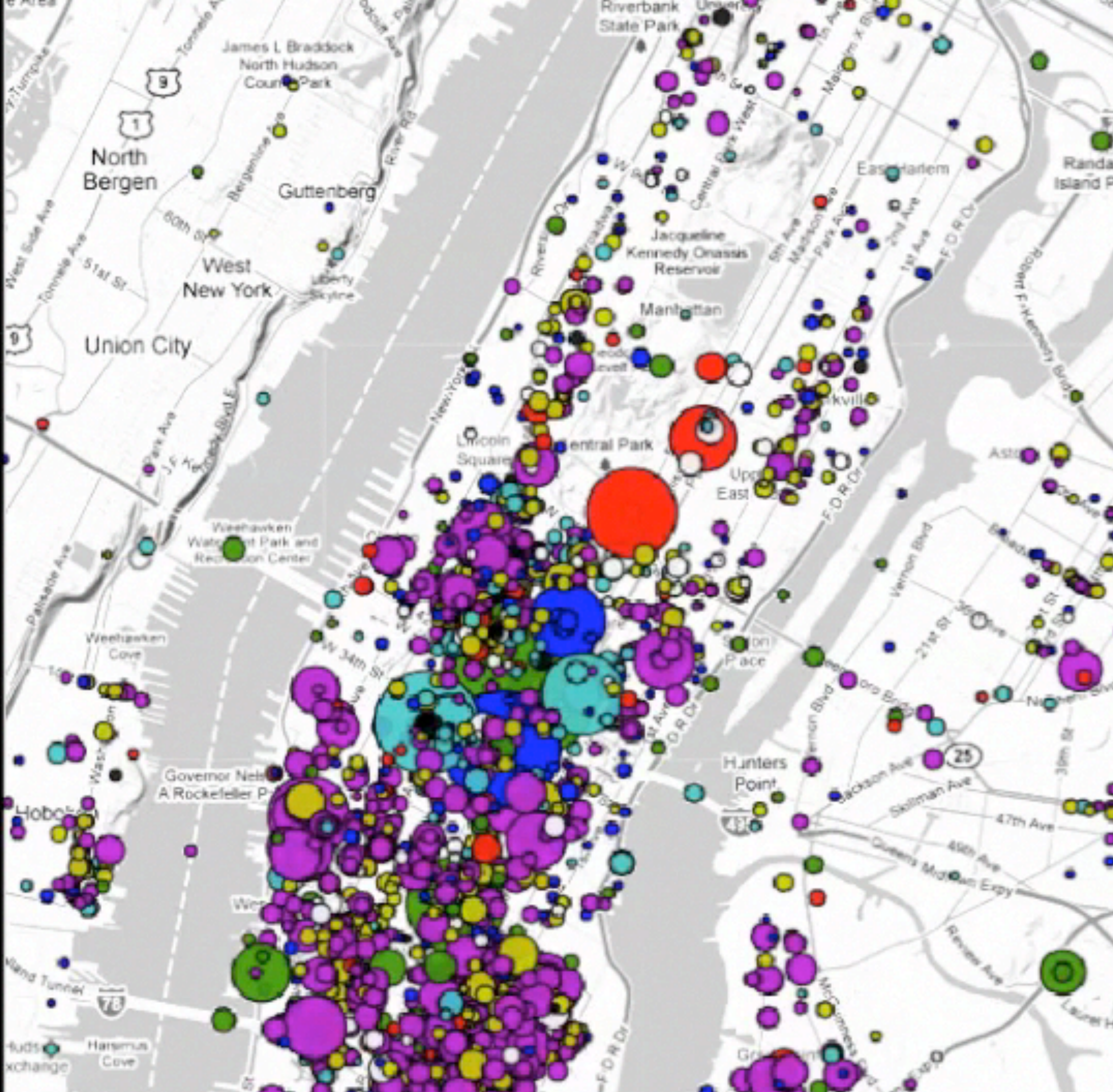
Little Amy checks-in at the Flower Shop on a Saturday Morning. Can we predict where she goes next?

A challenging problem indeed

high class imbalance: one single place to be elected amongst thousands.

extreme sparsity: most users check-in rarely...

very entropic behaviour: physicists believe its impossible to predict!


Riverbank
State Park
James L Braddock
North Hudson
County Park
North
Bergen
Guttenberg
West
New York
Union City
East Harlem
Randall
Island Pa
Jacqueline
Kennedy Onassis
Reservoir
Manhattan
Central Park
Lincoln
Square
Weehawken
Cove
Governor Nelson
A Rockefeller Park
Hunters
Point
47th Ave
Harsimus
Cove

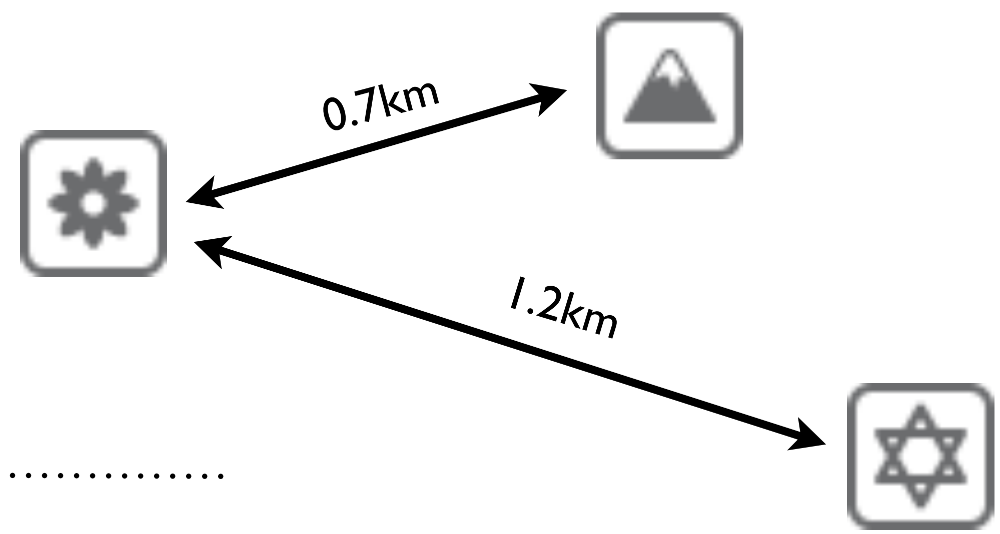
Computer Science at your Service

User Specific features

historic visits
friend check-ins
preferred venue types

Geographic

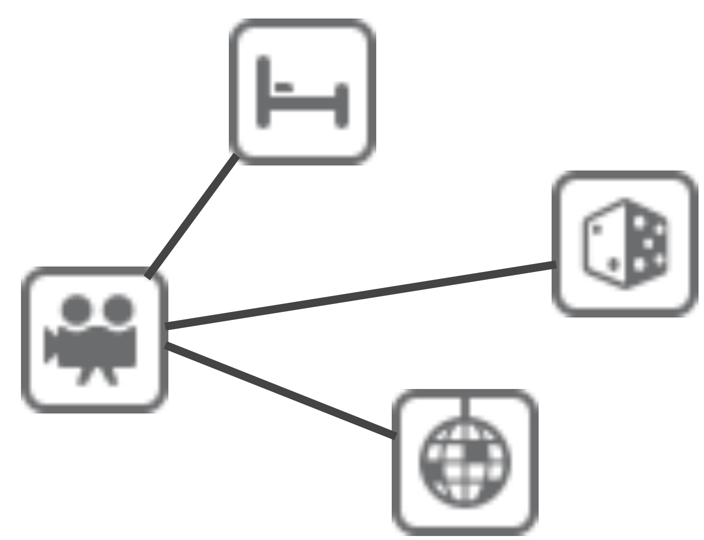
distance and rank-distance



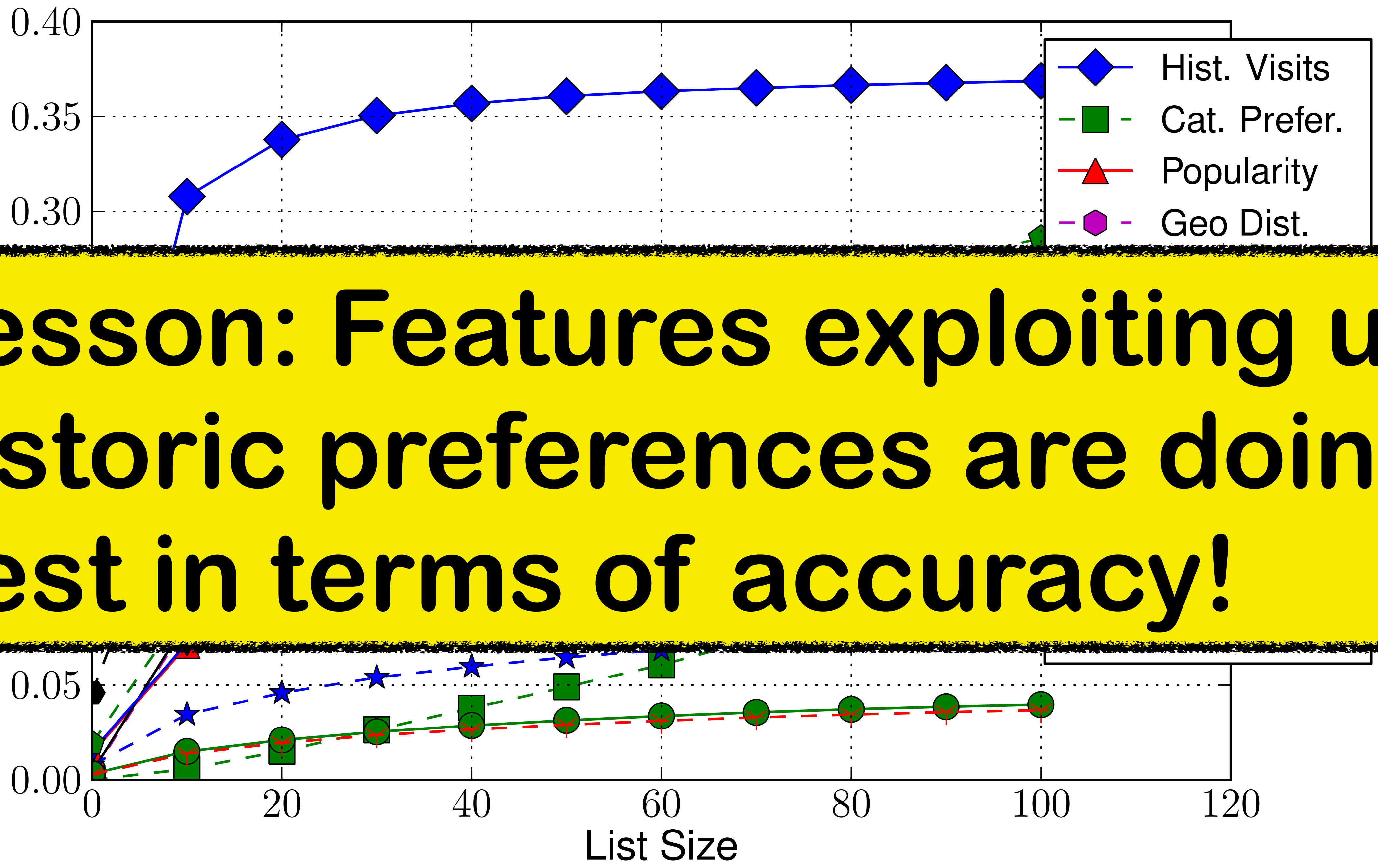
Temporal

trending places
(hour/day)
trending place types
(eg. cinema at nights)

Place Network

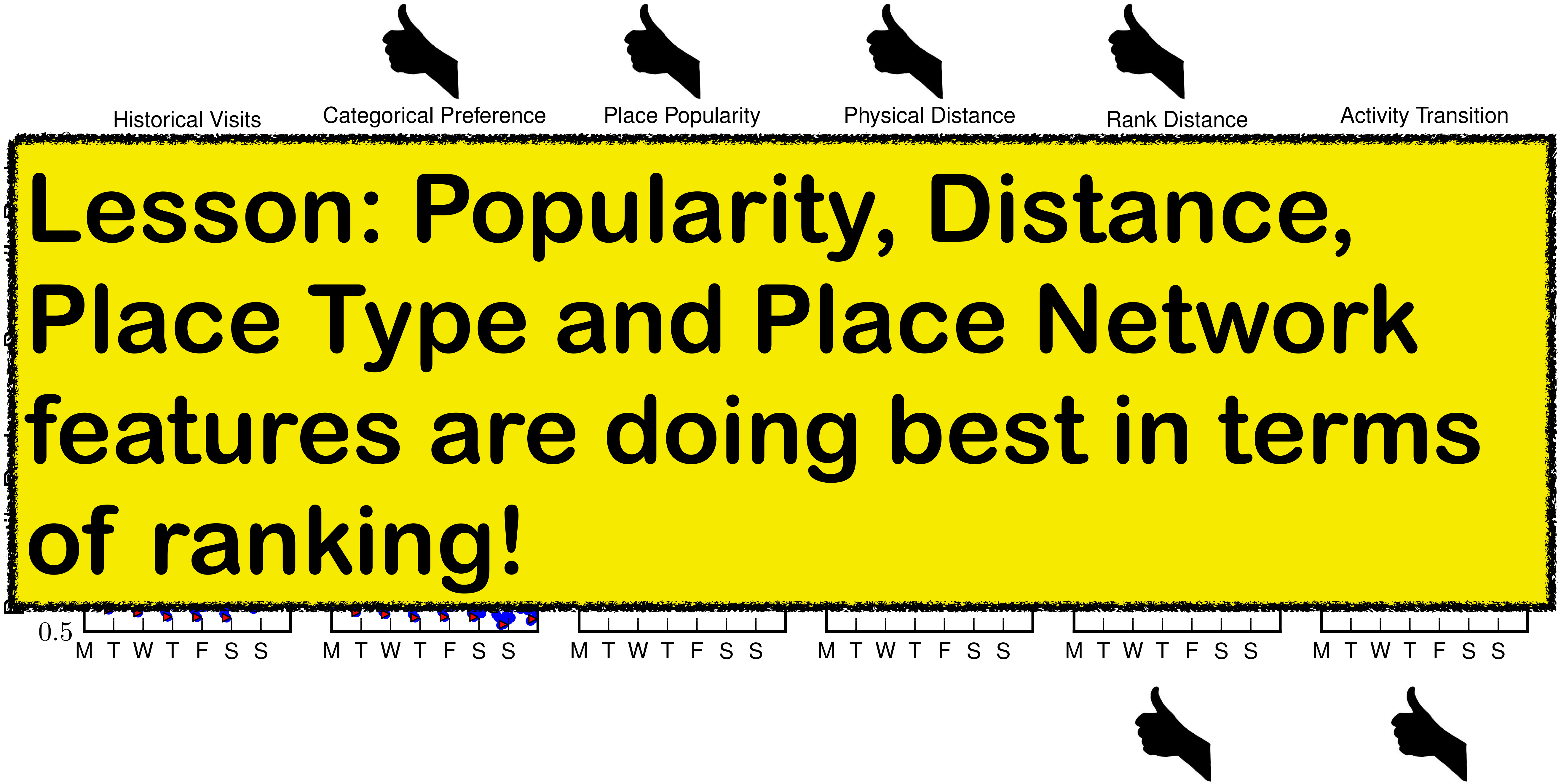


Feature Performance #1



Lesson: Features exploiting user historic preferences are doing best in terms of accuracy!

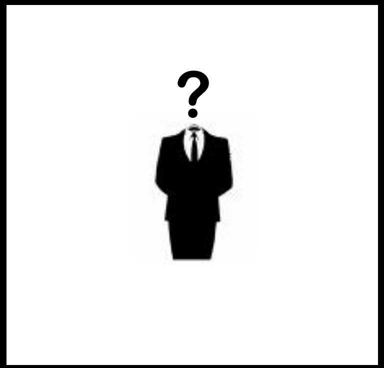
Feature Performance #2



Lesson: Popularity, Distance, Place Type and Place Network features are doing best in terms of ranking!

Duality in Feature Performance

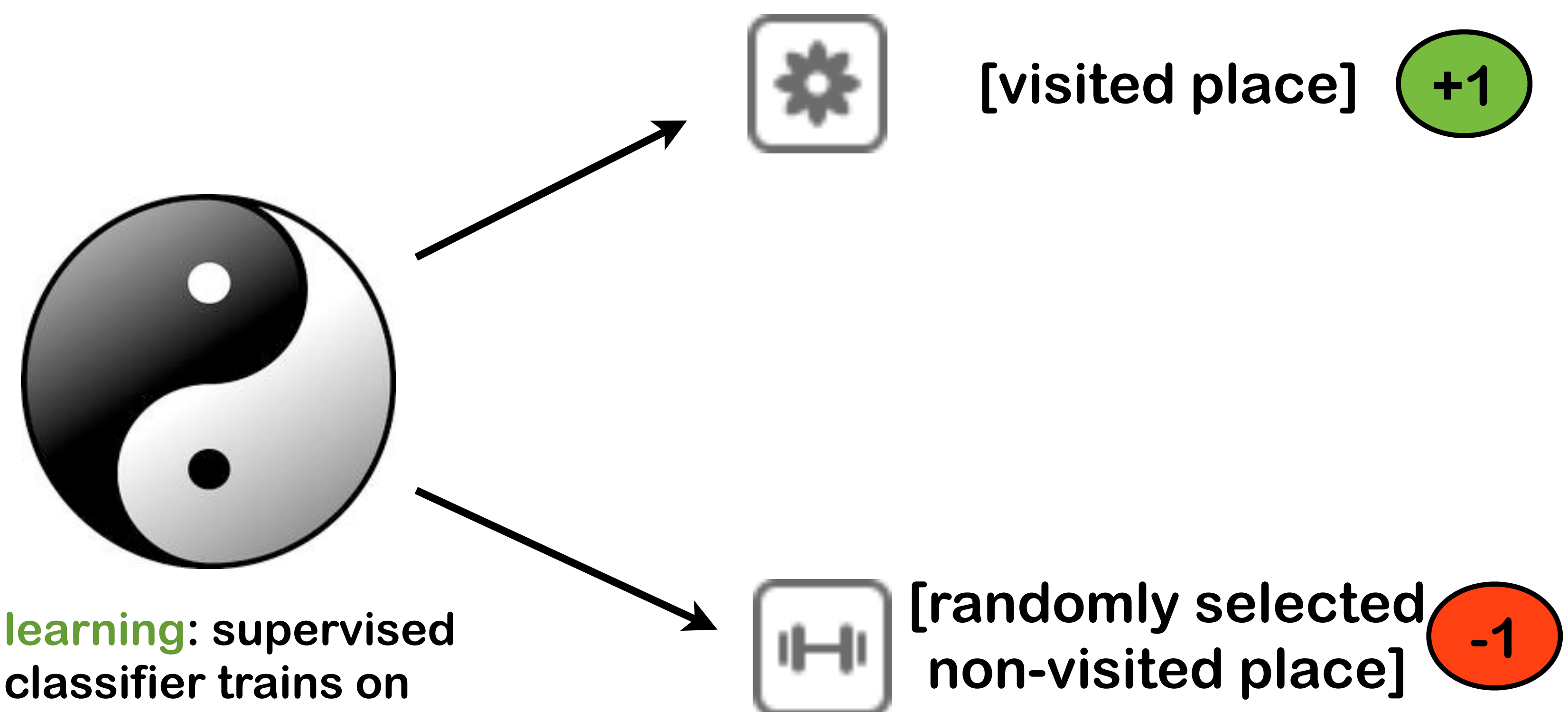
Personalized **VS** **Global**



Features personalized to the user are good to predict historically visited places!

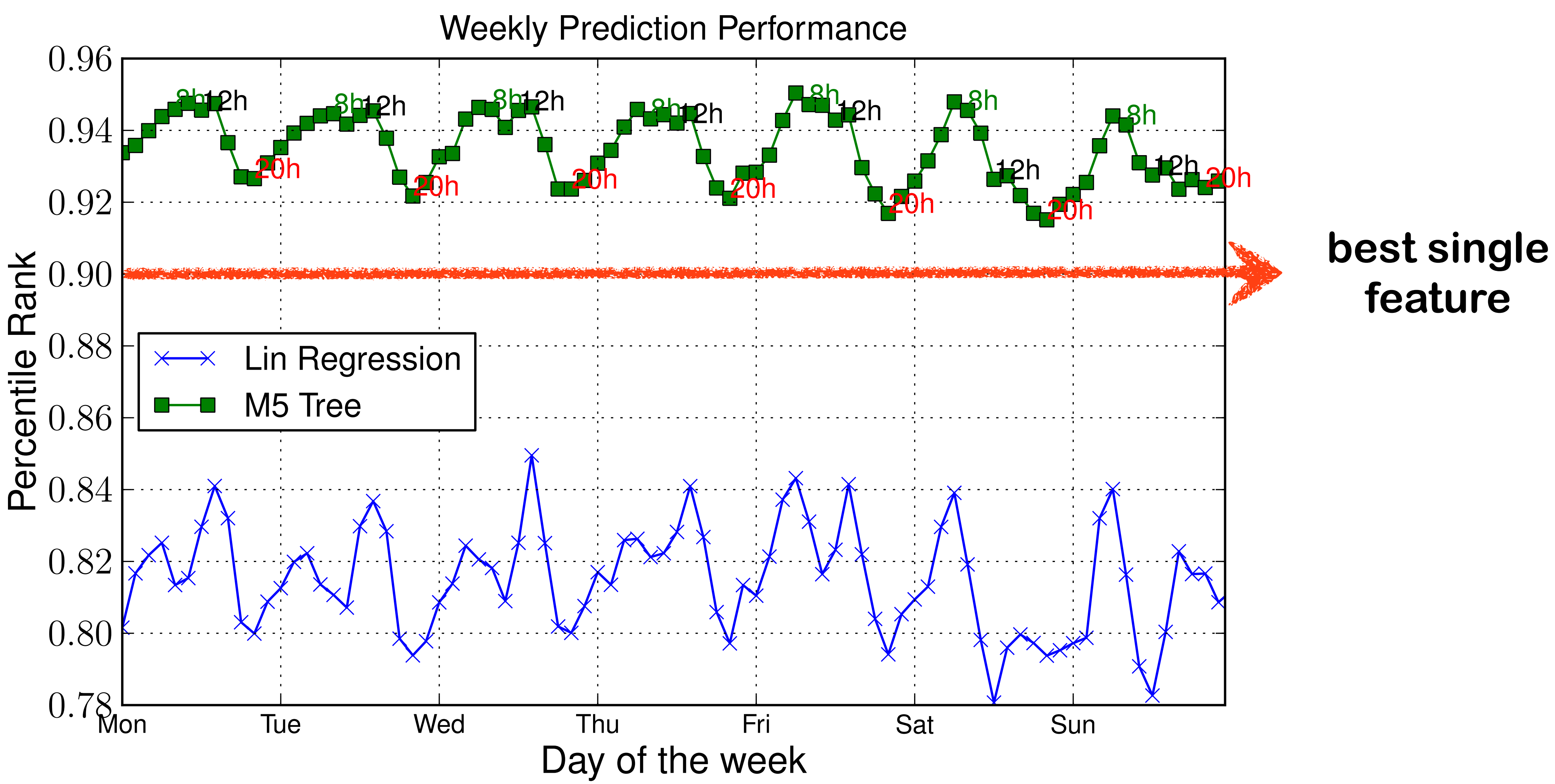
Features exploiting global information on check-in patterns are good to predict previously unvisited (new) places!

Supervised Training: teaching the good and the bad!

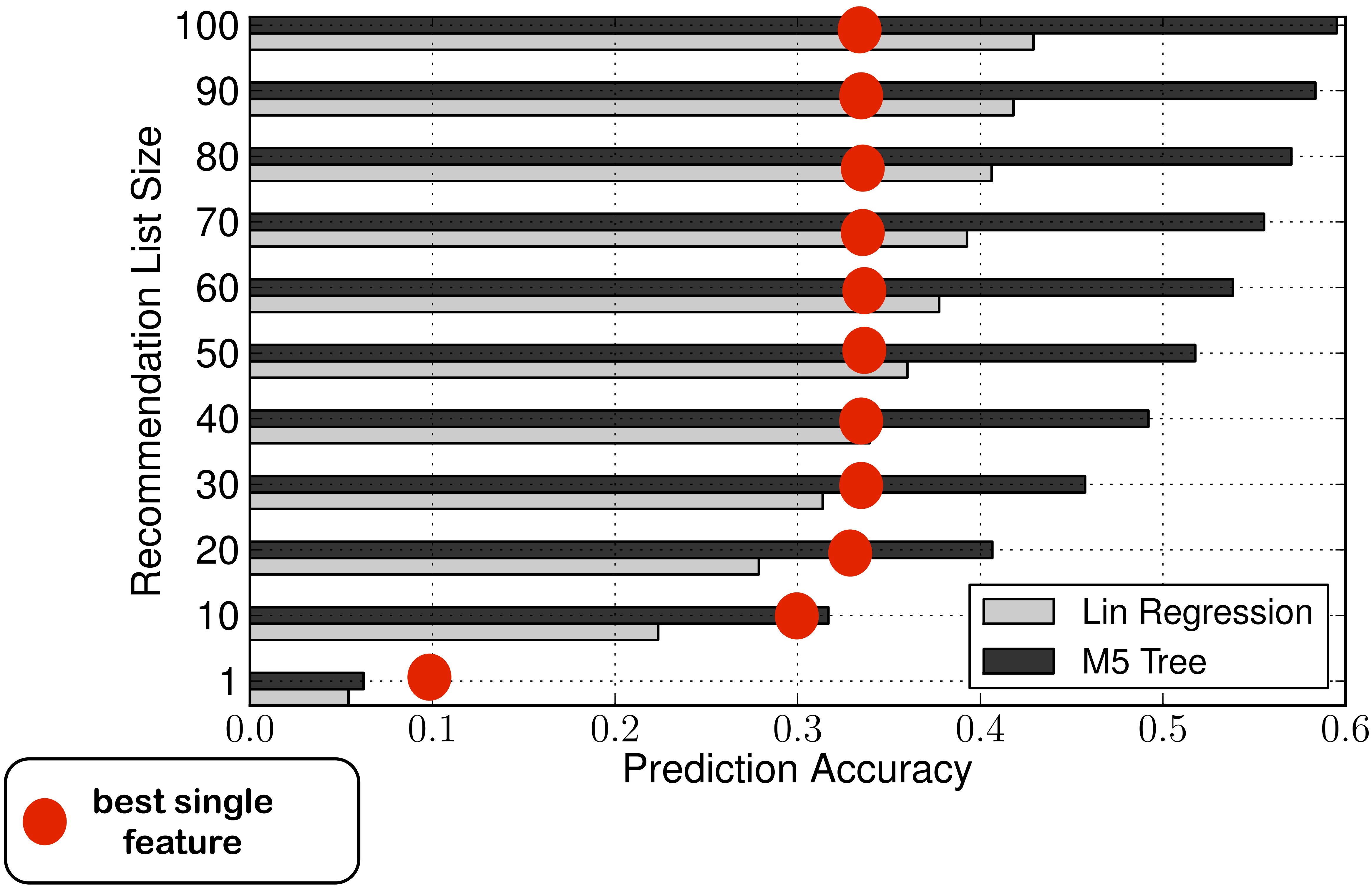


learning: supervised classifier trains on millions of check-ins

Supervised Learning Scores!



The Decision Tree of Life



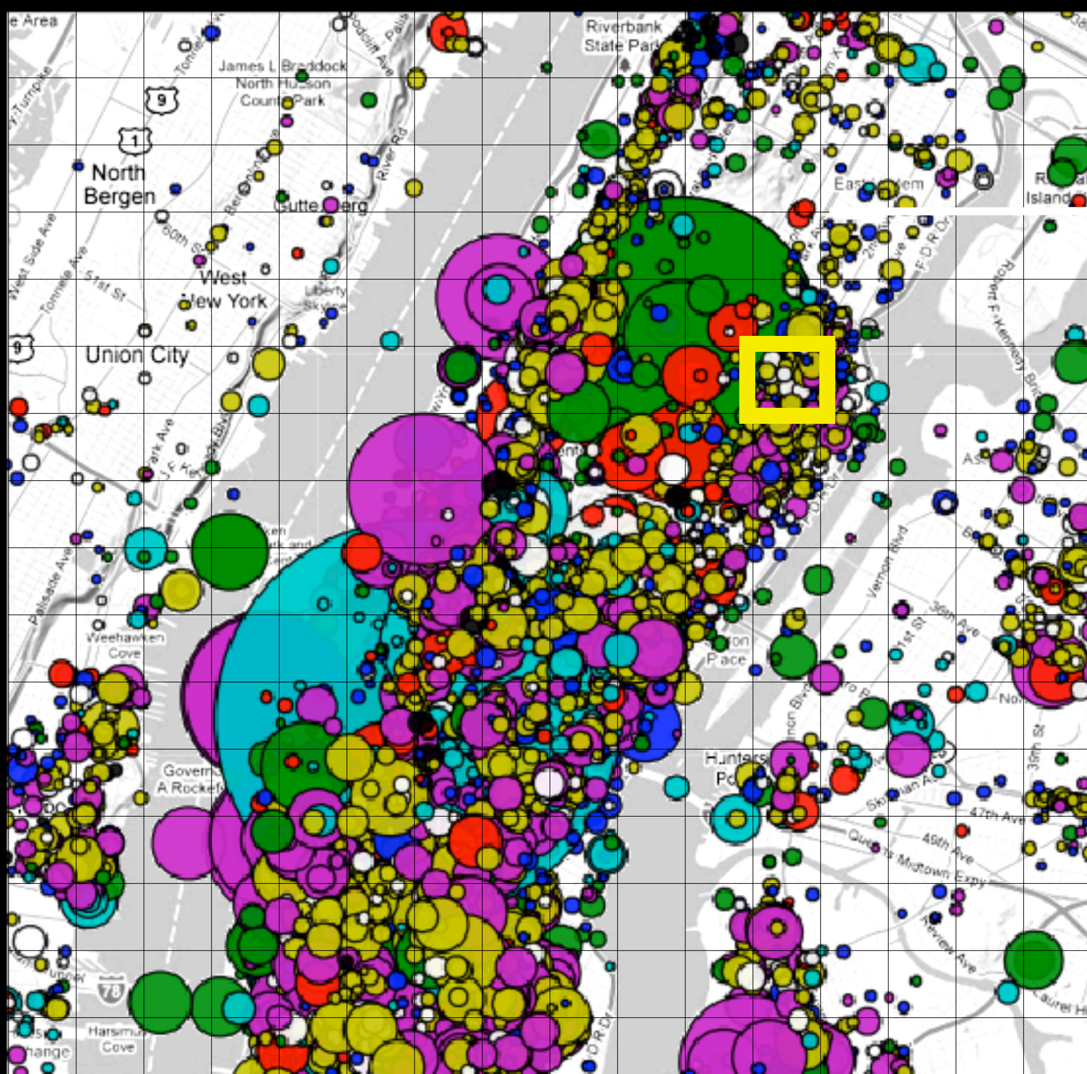
Part 1 END

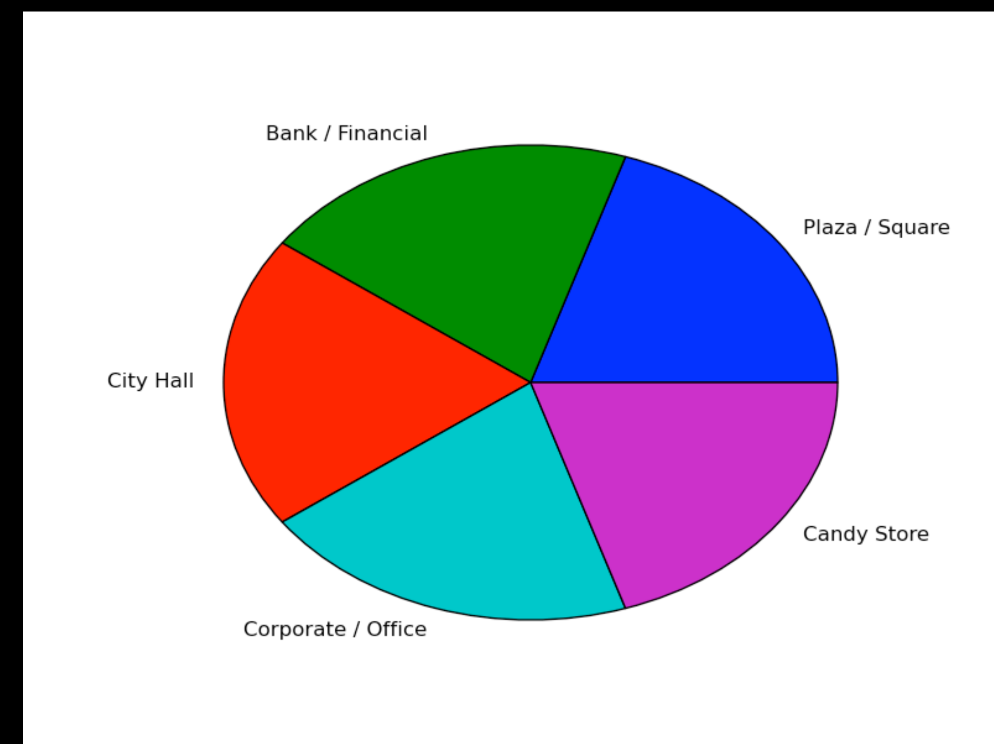

THE END

BRICK LANE E1
ব্রিক লেন
HOT BEIGELS ALL NIGHT
BEIGEL BAKE
BRICK LANE BAKERY
No.159
OPEN 24 HOURS 7 DAYS
Tel.071-729 0616
Fika

An overall consensus among urban studies and public policy researchers defines a “neighborhood” as a contiguous geographic area within a larger city, limited in size, and somewhat homogeneous in its characteristics.

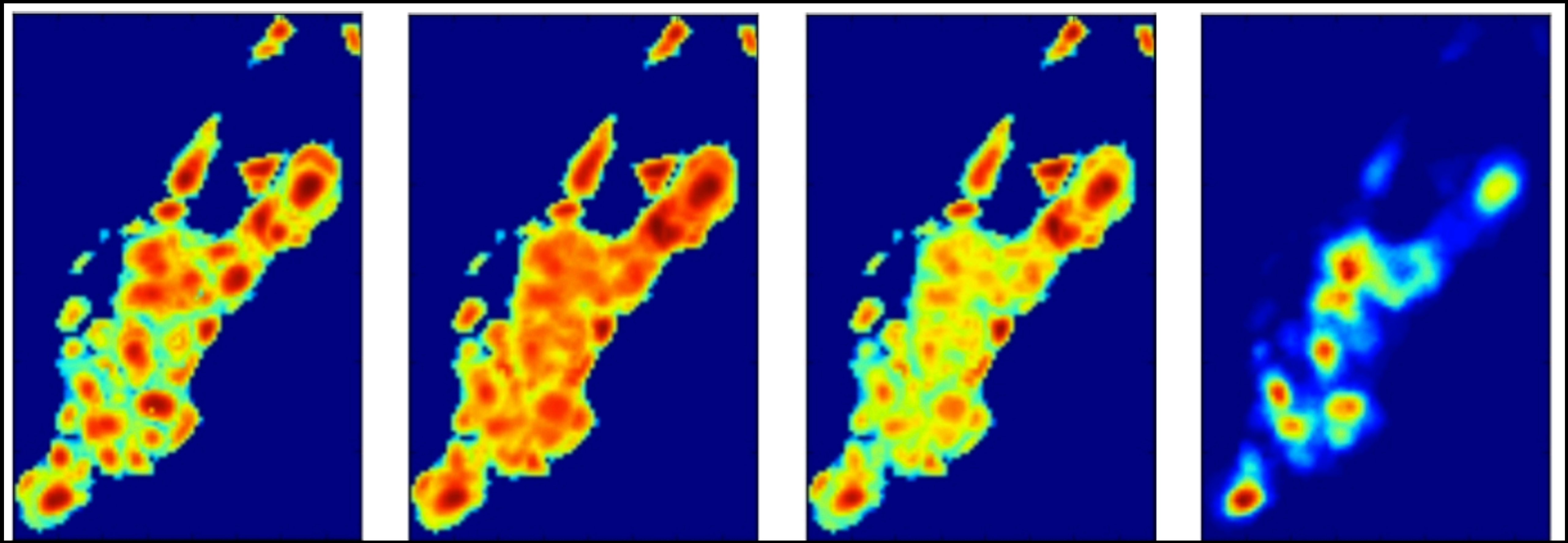
Representing a neighbourhood ...





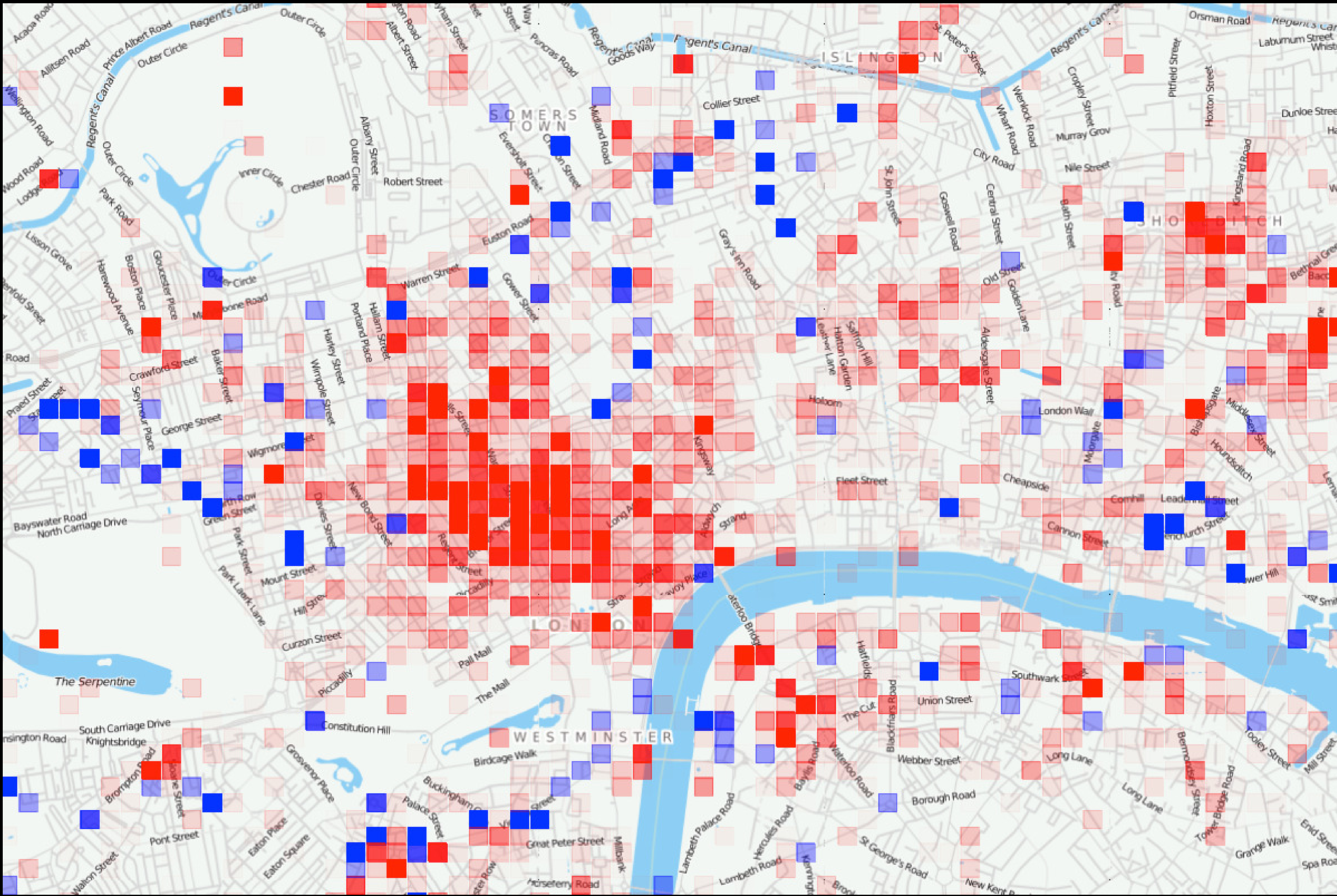
Each square area a multi-dimensional vector

The Principle of Homogeneity

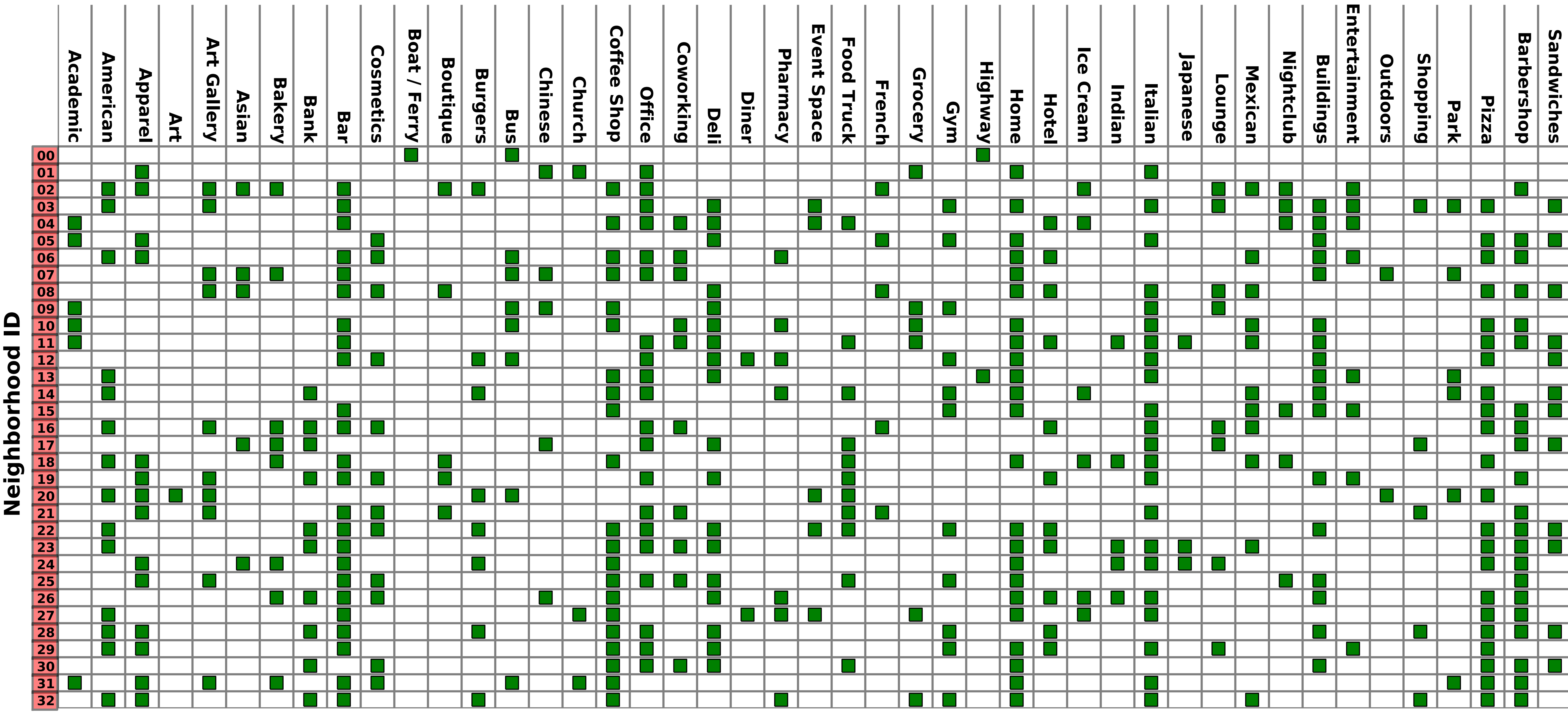


$$\text{H\_Index}(\mathrm{p},\mathrm{r}) = \frac{\sum_{n \in N_{p,r}} \cos(\mathbf{v_p}, \mathbf{v_n}) \times \mathrm{smooth}(\mathrm{n})}{|N_{p,r}|}$$

Locals - Vs - Tourists



Neighborhood Features in New York



Exploiting the network of places for neighborhood detection

Livehoods

Home Maps About Research Press Contact

Re-Imagining the City in the Age of Social Media

Livehoods offer a new way to conceptualize the dynamics, structure, and character of a city by analyzing the social media its residents generate. By looking at people's checkin patterns at places across the city, we create a mapping of the different dynamic areas that comprise it. Each Livehood tells a different story of the people and places that shape it.

> MORE

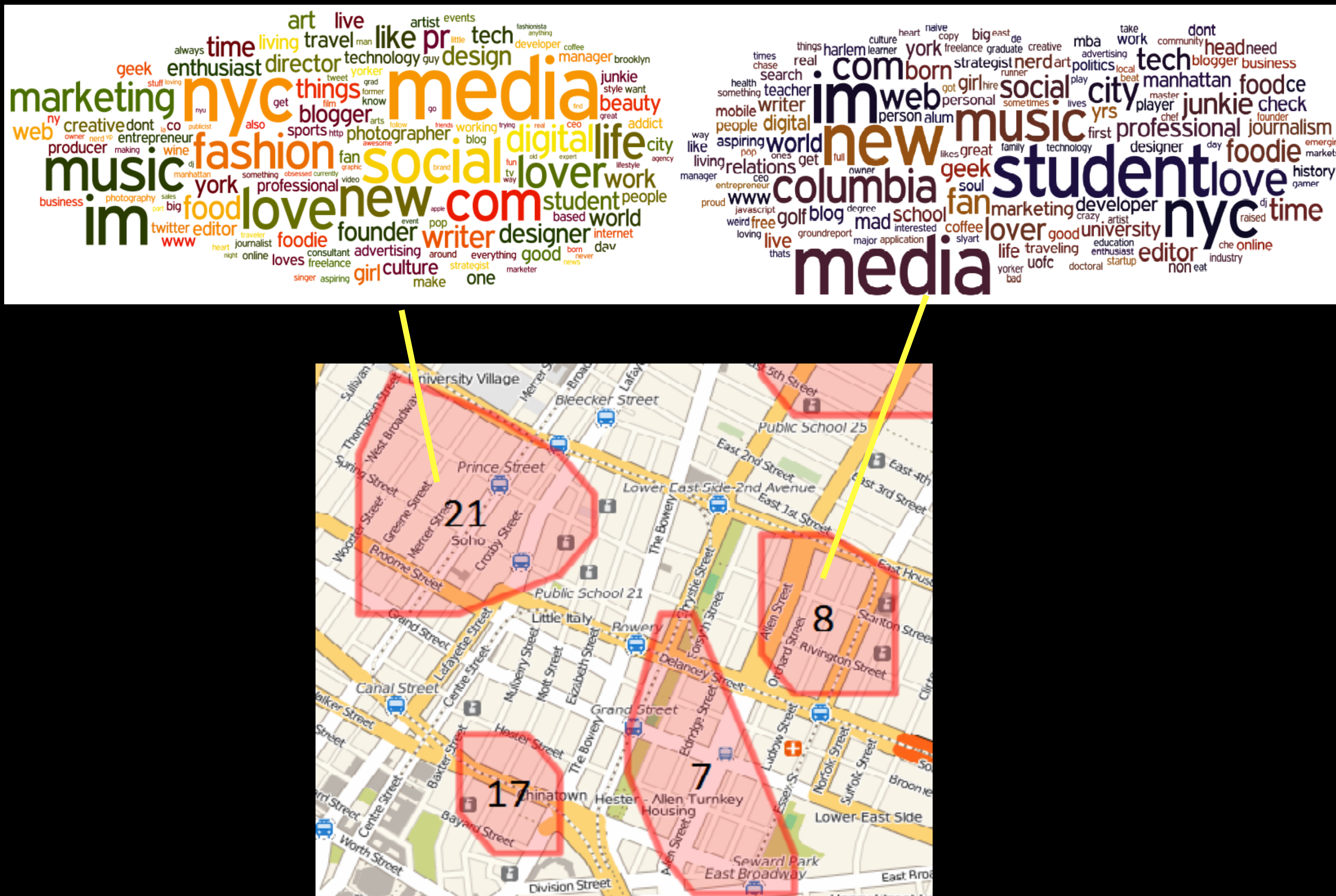
Using Machine-Learning to Study Cities

Our research hypothesis is that the character of an urban area is defined not just by the the types of places found there, but also by the people that make it part of their daily life. To explore this idea, we use data from approximately 18 million check-ins collected from the location-based social network foursquare, and apply clustering algorithms to discover the different areas of the city.

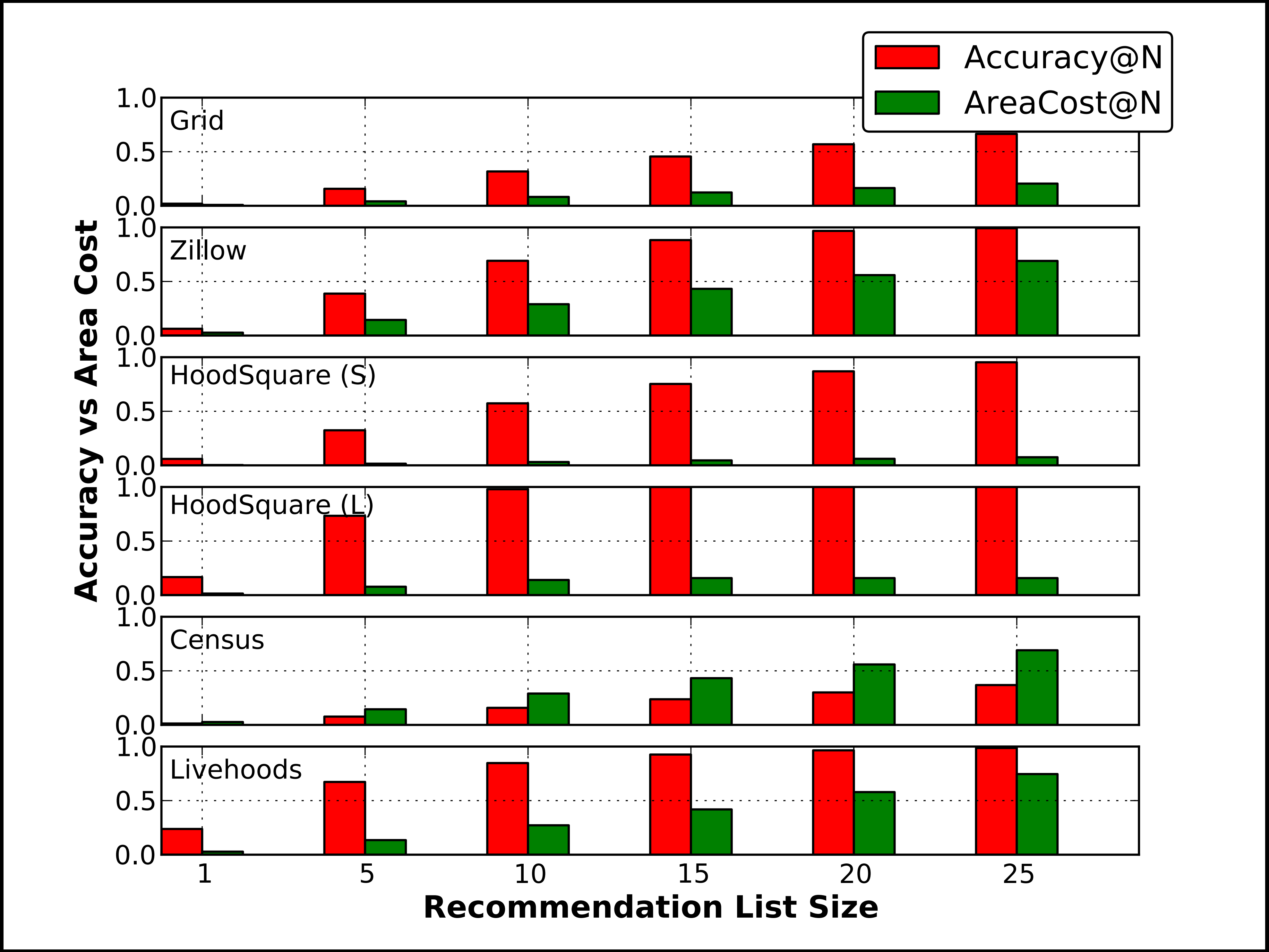
> MORE

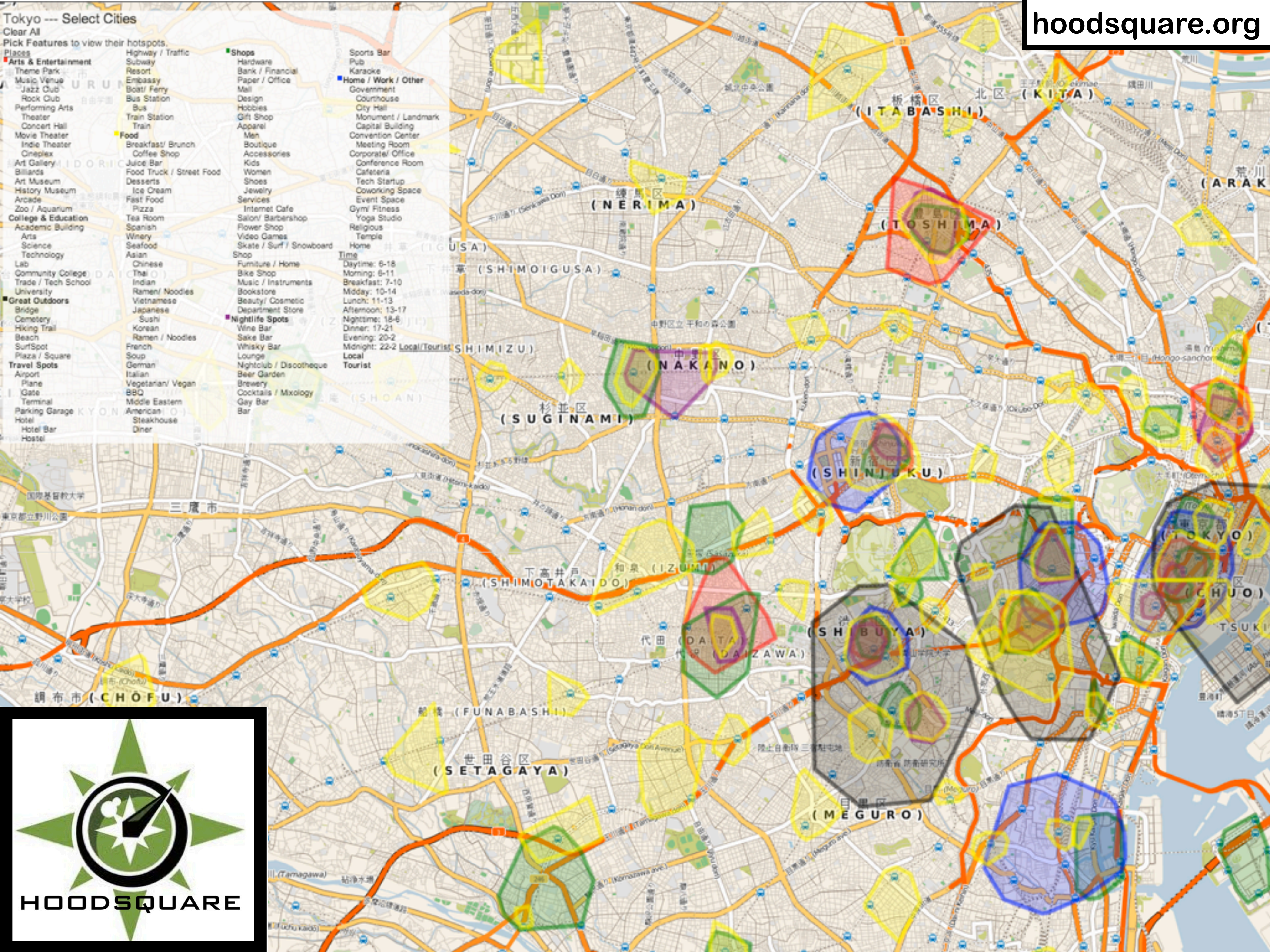
www.livehoods.org

Recommending Neighborhoods

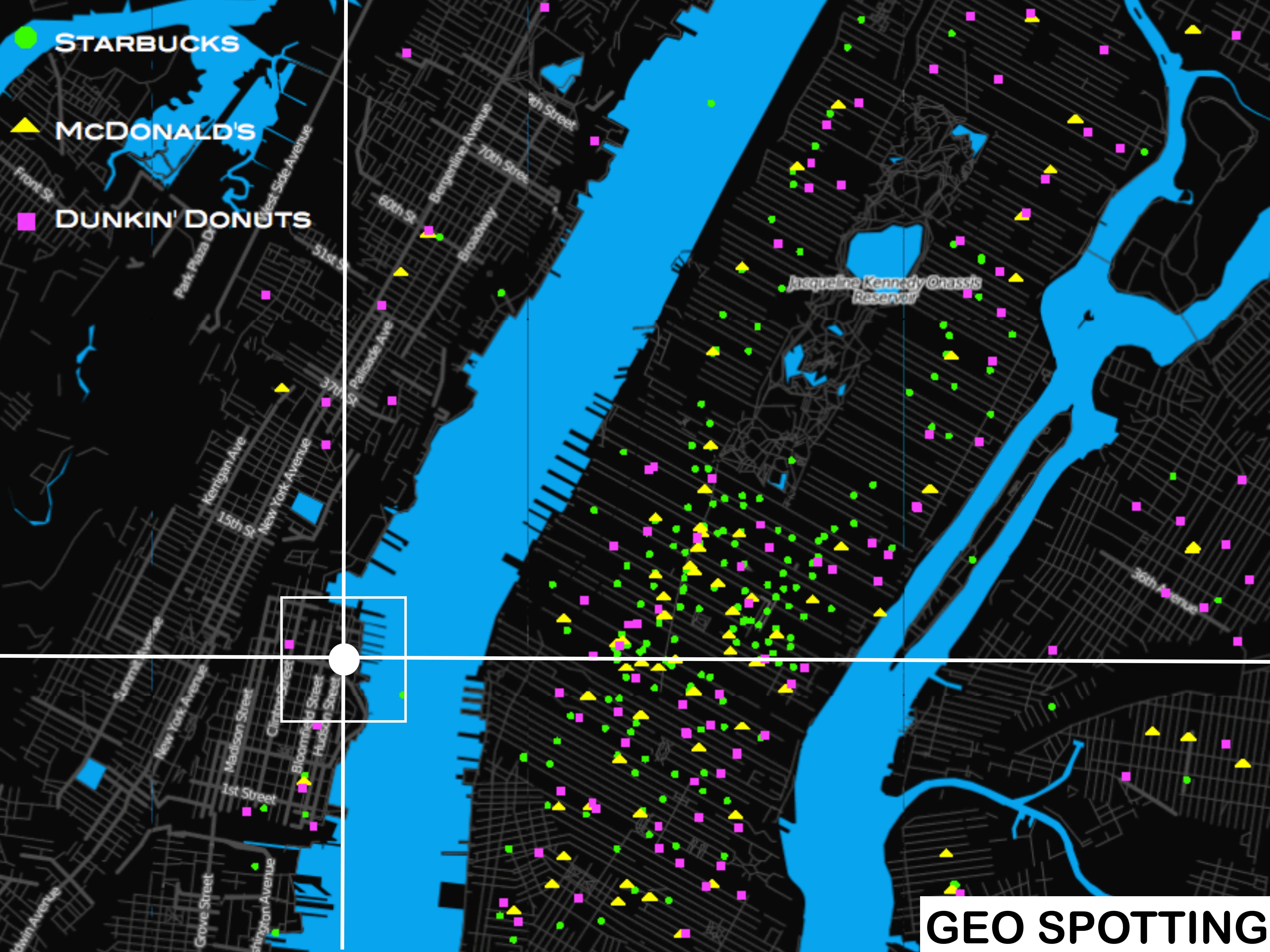


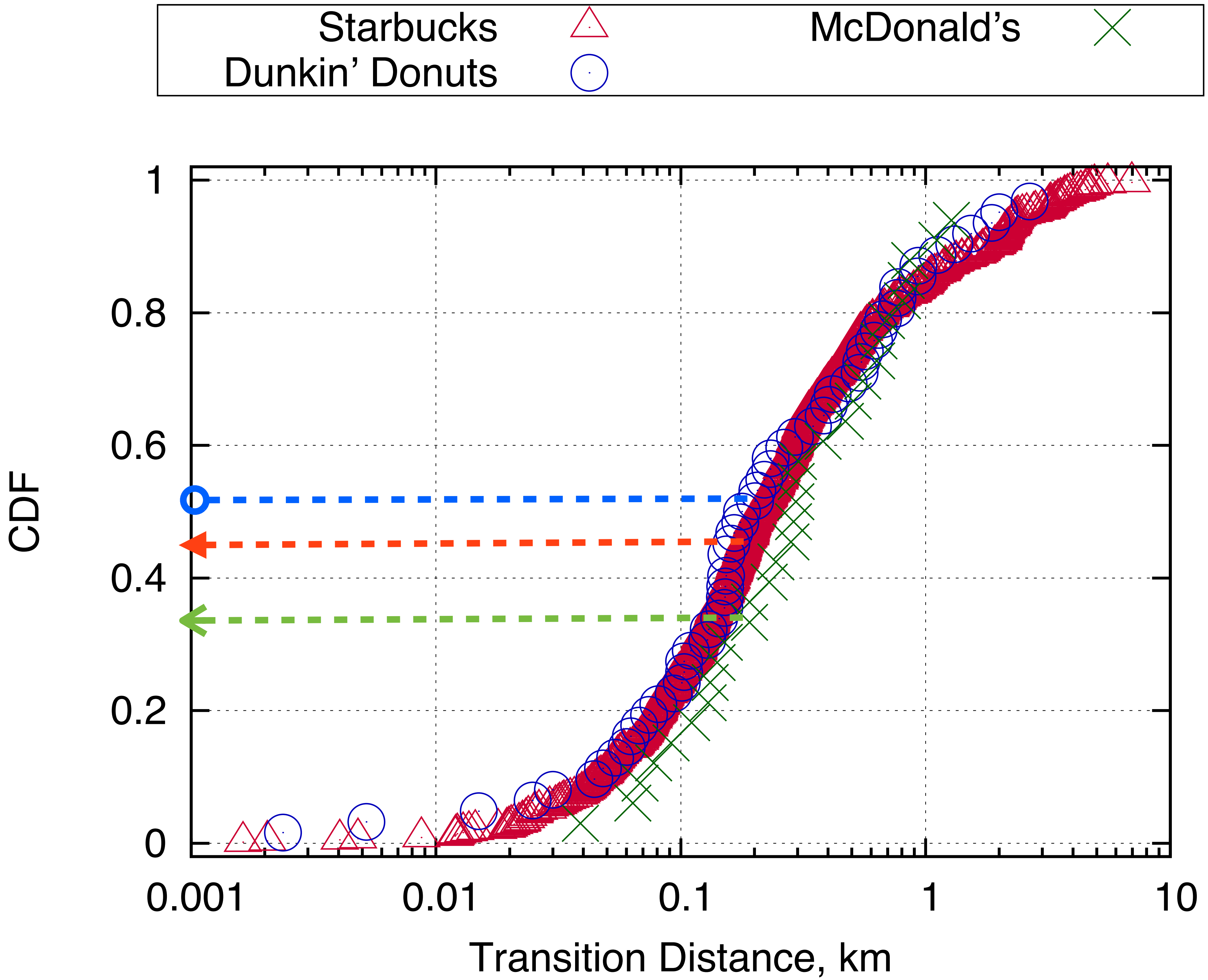
Recommending Neighborhoods





hoodsquare.org
Tokyo --- Select Cities
Clear All
Pick Features to view their hotspots.
Places
Arts & Entertainment
Theme Park
Music Venue
Jazz Club
Rock Club
Performing Arts
Theater
Concert Hall
Movie Theater
Indie Theater
Cineplex
Art Gallery
Billiards
Art Museum
History Museum
Arcade
Zoo / Aquarium
College & Education
Academic Building
Arts
Science
Technology
Lab
Community College
Trade / Tech School
University
Great Outdoors
Bridge
Cemetery
Hiking Trail
Beach
SurfSpot
Plaza / Square
Travel Spots
Airport
Plane
Gate
Terminal
Parking Garage
Hotel
Hotel Bar
Hostel
Highway / Traffic
Subway
Resort
Embassy
Boat/ Ferry
Bus Station
Bus
Train Station
Train
Food
Breakfast/ Brunch
Coffee Shop
Juice Bar
Food Truck / Street Food
Desserts
Ice Cream
Fast Food
Pizza
Tea Room
Spanish
Winery
Seafood
Asian
Chinese
Thai
Indian
Ramen/ Noodles
Vietnamese
Japanese
Sushi
Korean
Ramen / Noodles
French
Soup
German
Italian
Vegetarian/ Vegan
BBQ
Middle Eastern
American
Steakhouse
Diner
Shops
Hardware
Bank / Financial
Paper / Office
Mall
Design
Hobbies
Gift Shop
Apparel
Men
Boutique
Accessories
Kids
Women
Shoes
Jewelry
Services
Internet Cafe
Salon/ Barbershop
Flower Shop
Video Games
Skate / Surf / Snowboard
Shop
Furniture / Home
Bike Shop
Music / Instruments
Bookstore
Beauty/ Cosmetic
Department Store
Nightlife Spots
Wine Bar
Sake Bar
Whisky Bar
Lounge
Nightclub / Discotheque
Beer Garden
Brewery
Cocktails / Mixology
Gay Bar
Bar
Sports Bar
Pub
Karaoke
Home / Work / Other
Government
Courthouse
City Hall
Monument / Landmark
Capital Building
Convention Center
Meeting Room
Corporate/ Office
Conference Room
Cafeteria
Tech Startup
Coworking Space
Event Space
Gym/ Fitness
Yoga Studio
Religious
Temple
Home
Time
Daytime: 6-18
Morning: 6-11
Breakfast: 7-10
Midday: 10-14
Lunch: 11-13
Afternoon: 13-17
Nighttime: 18-6
Dinner: 17-21
Evening: 20-2
Midnight: 22-2
Local/Tourist
Local
Tourist
(ITABASHI)
(KITA)
(ARAK
(NERIMA)
(TOSHIMA)
(SHIMOIGUSA)
(SHIMIZU)
(NAKANO)
(SUGINAMI)
(SHINJUKU)
(TOKYO)
(CHUO)
(SHIMOTAKAIDO)
(IZUMI)
(SHIBUYA)
(DAIZAWA)
(CHOFU)
(FUNABASHI)
(SETAGAYA)
(MEGURO)
HOODSQUARE


STARBUCKS
McDONALD'S
DUNKIN' DONUTS
Jacqueline Kennedy Onassis Reservoir
60th St
70th Street
Bergenline Avenue
Broadway
Palisade Ave
Kerrigan Ave
New York Avenue
15th St
Front St
Madison Street
1st Street
Grove Street
36th Avenue
GEO SPOTTING



Relatively speaking… McDonald's tend to "pull" check-ins from remote locations whereas Dunkin' and Starbucks attract local movements.

r
East 20th Street
East 19th Street
East 18th Street
East 17th Street
East 16th Street
East 15th Street
East 14th Street
East 13th Street
East 12th Street
East 11th Street
2nd Avenue
1st Avenue
Stuyvesant Square Historic District
Stuyvesant Square
Stuyvesant Loop West
Stuyvesant Loop North
STUYVESANT TOWN
PETER COOPER VILLAGE

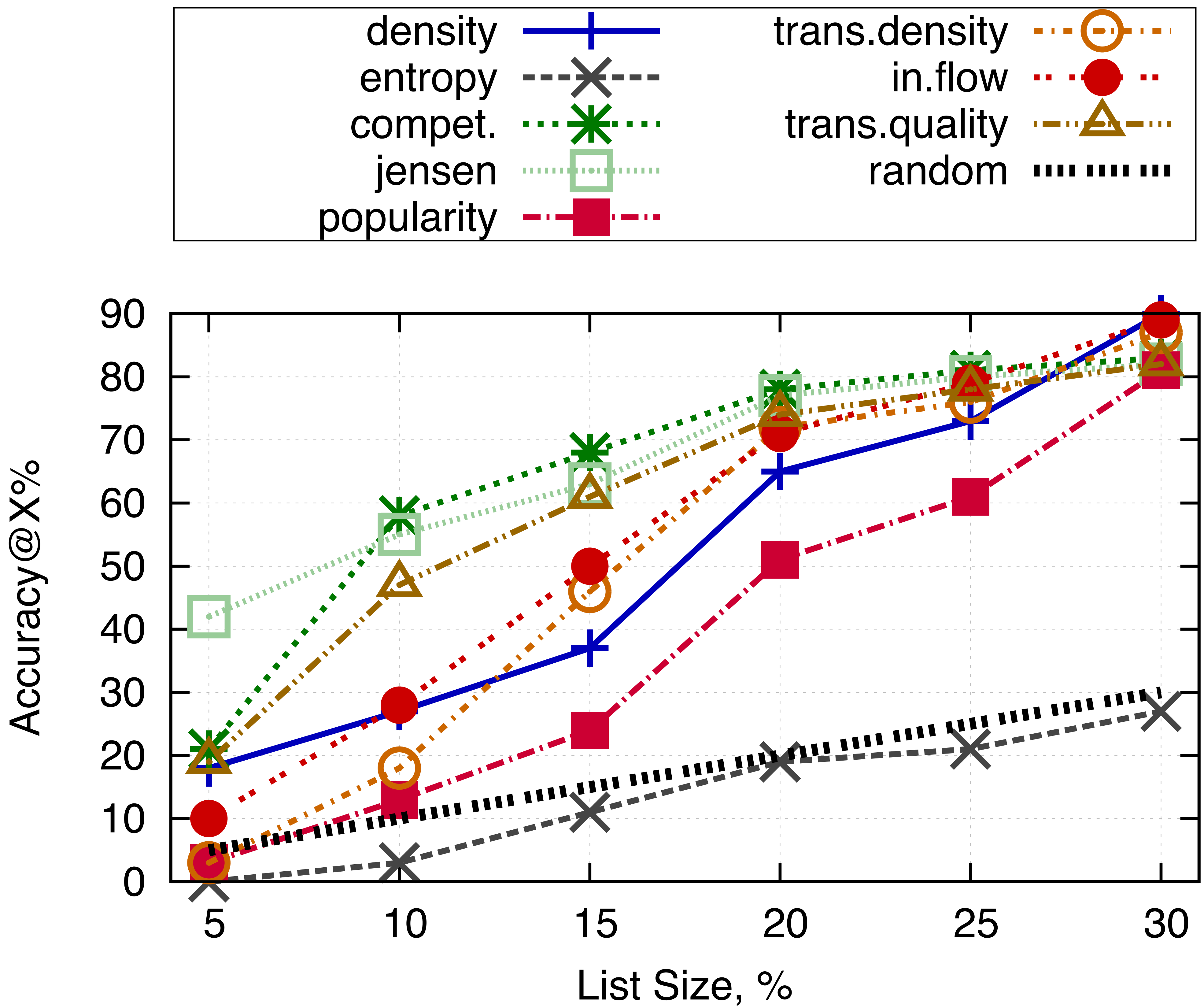
Starbucks		Dunkin' Donuts		McDonalds	
Train Station	11.80	Hostel	5.02	Flower Shop	5.87
Light Rail	8.60	Gas Station	3.05	Office Supplies	3.16
Stadium	7.25	Automotive Shop	2.66	Train Station	3.08
Airport	6.24	Flower Shop	2.36	Theatre	2.84
Museum	5.10	Post Office	2.19	Light Rail	2.32
Convention Centre	4.93	Flea Market	1.84	Gift Shop	2.26
Hostel	4.82	School	1.72	Subway Station	2.21
Corporate Office	4.57	Drug Store	1.70	Department Store	2.17
Hotel	4.13	Subway Station	1.67	Bank / Financial	1.92
Bank / Financial	4.09	Bike shop	1.64	Drug Store	1.89

colocation

Does colocation imply movement?

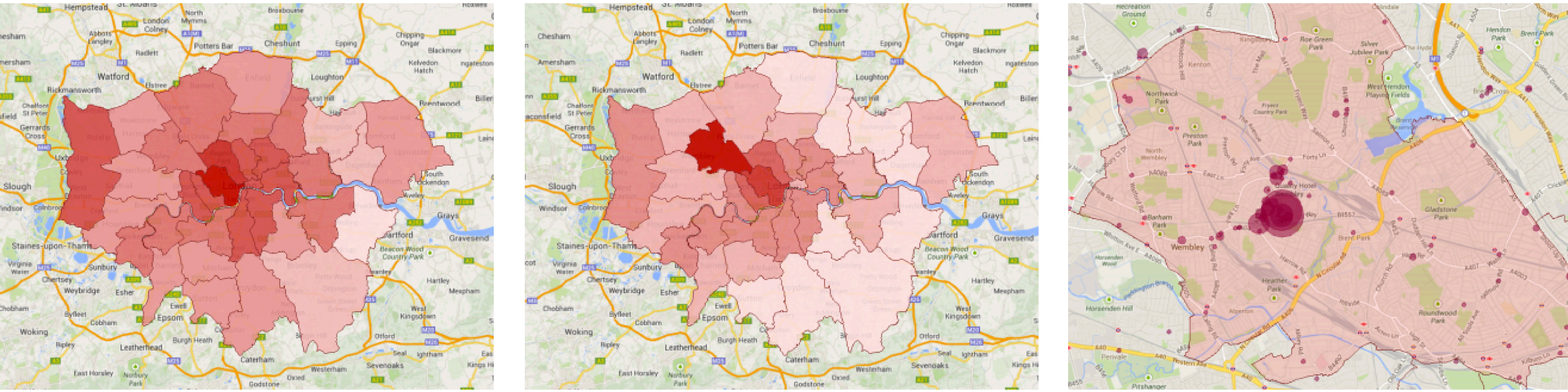
Starbucks		Dunkin' Donuts		McDonalds	
Hostel	15.75	Laundry	11.89	Parks & Outdoor	16.00
Flea Market / Fair	8.38	Drug Store / Pharmacy	5.78	Gas Station	3.72
Sculpture	8.00	Subway Station	2.16	Gift Shop	3.56
Post Office	2.34	Food Shop	1.66	Theatre	3.20
Services	2.20	Medical	1.25	Office Supplies	3.05
Drug Store / Pharmacy	2.20	Home	1.24	Bank / Financial	2.91
Quad / Commons	1.72	Apparel	1.12	Plaza / Square	2.57
Bank / Financial	1.57	Bank / Financial	1.06	Drug Store	2.37
Airport	1.52	School	1.08	Apparel	1.04
Office Supplies	1.52	Post Office	1.07	Home	1.02

mobility


density
entropy
compet.
jensen
popularity
trans.density
in.flow
trans.quality
random
Accuracy@X%
0
10
20
30
40
50
60
70
80
90
5
10
15
20
25
30
List Size, %


EVENTS
How
Who
What
When
Where
Why

Why are events important?



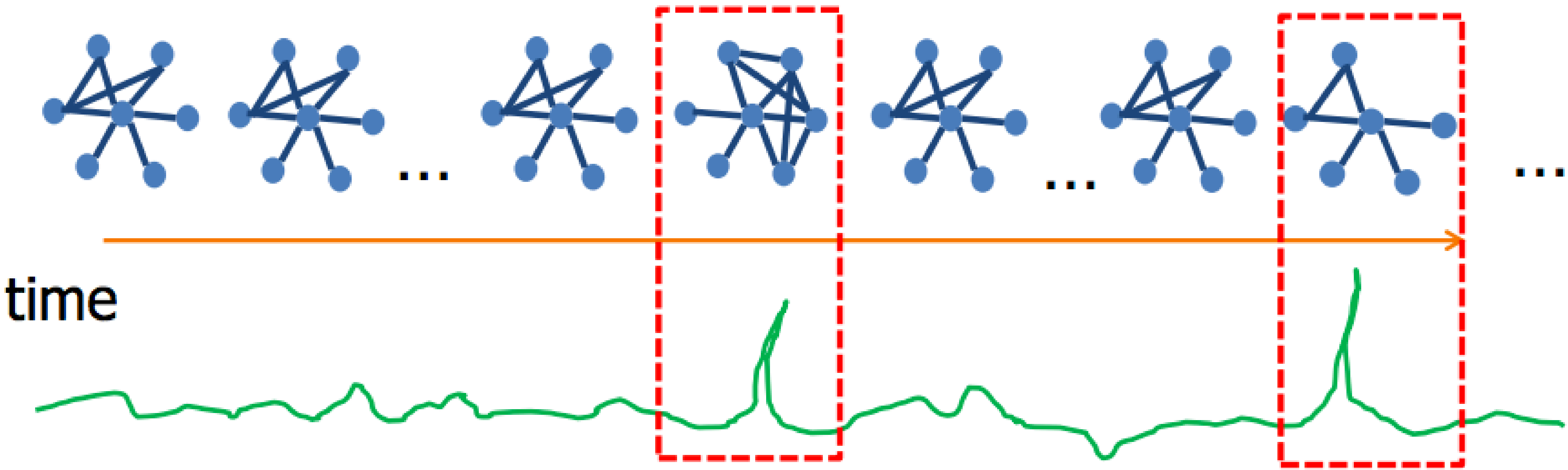
(a) Check-ins, 26 May 2011 (b) Check-ins, 28 May 2011 (c) Check-ins at the Wembley area, 28 May

Social Events are a major disruptor of urban activity.

30-50% of check-ins in the city are generate during events.

Two implications: i) events are important per se
ii) could events affect other measurements?

Event Detection



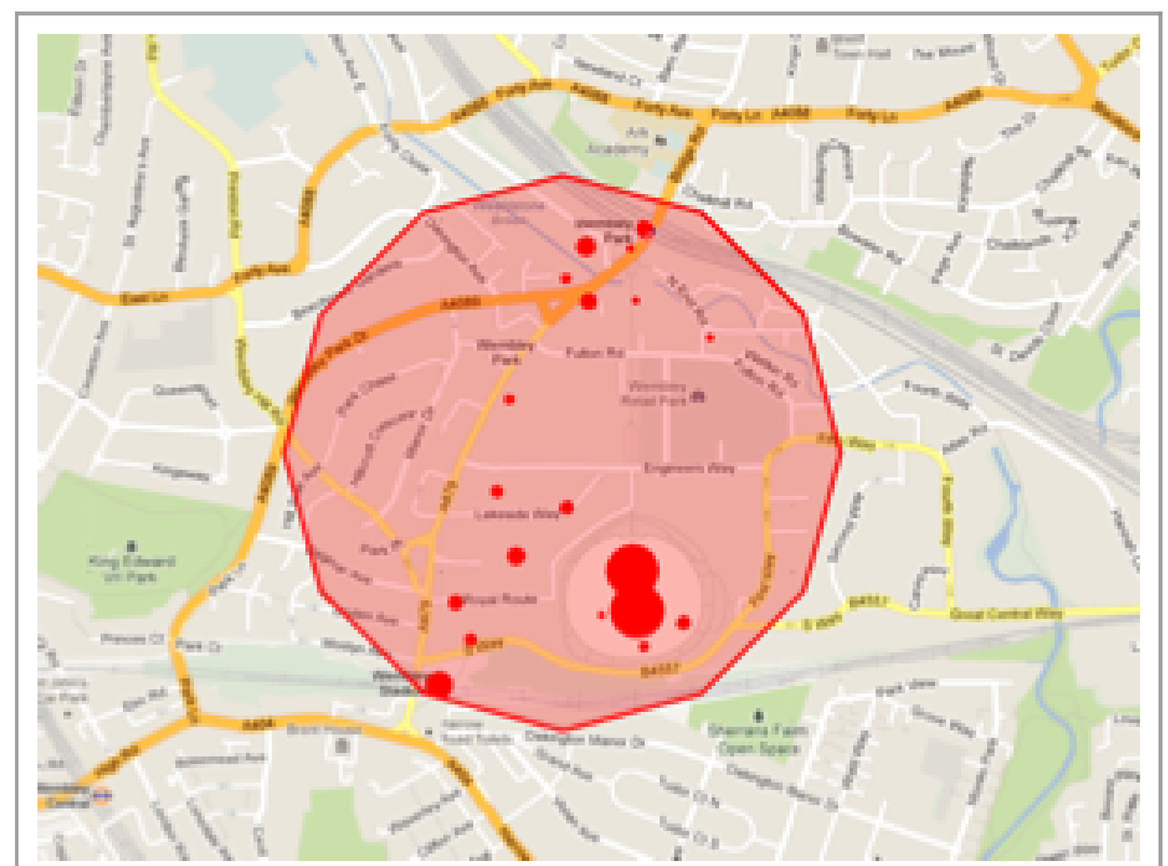
As the case is with all modeling domains, there are plenty of techniques out there....

Our approach is simple and captures the fundamental underlying principle: seek for large deviations of check-ins at places in the city, sort, then validate.

http://www.cs.stonybrook.edu/~leman/icdm12/

Event Scope Definition

(a) MCM Expo (b) UEFA Champions League Final (c) Internet Week (d) Webby Awards (e) DrupalCon (f) Lollapalooza Festival

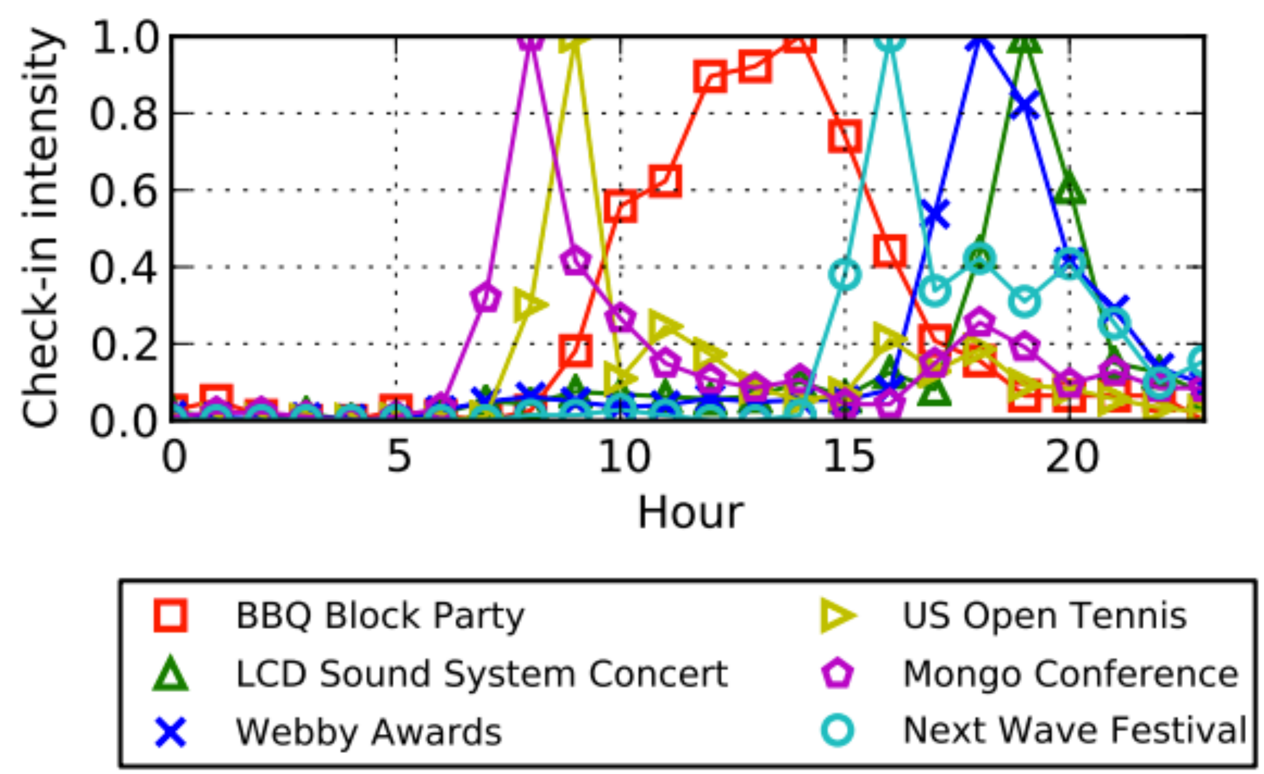


Find the most popular place of an event. Then, add to the event's place list all abnormally active places within 300 meter radius...

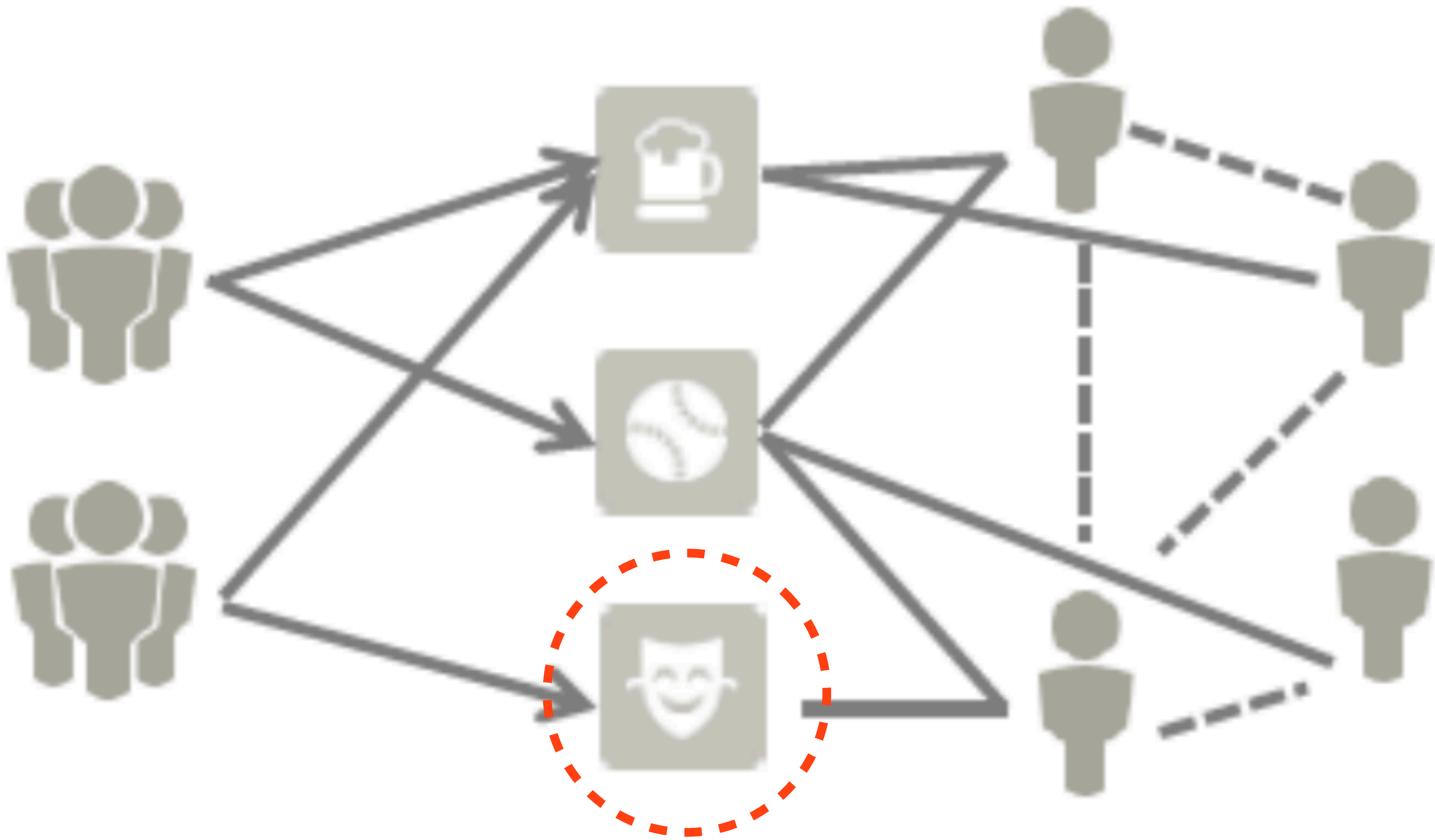
Event Representation (1)

Blogworld Expo		Orioles-Yankees Baseball		Lollapalooza		Chicago Comic Con	
Place type	Score	Place type	Score	Place type	Score	Place type	Score
Convention Center	0.0074	**Baseball**	0.0138	**Music Venue**	0.0947	Indie Theater	0.0106
Event Space	0.0033	Bar	0.0070	Bar	0.0353	**Bookstore**	0.0098
Hotel	0.0025	**Sports Bar**	0.0067	American	0.0195	Convention Center	0.0076
Vegetarian / Vegan	0.0024	Pub	0.0049	Mexican	0.0162	**Cineplex**	0.0072
Train Station	0.0020	Pizza	0.0039	Sports Bar	0.0162	Other - Buildings	0.0059
American	0.0016	**Stadium**	0.0038	Pub	0.0162	Electronics	0.0052
Tech Startup	0.0015	American	0.0031	**Other - Entertainment**	0.0161	Fast Food	0.0047
Corporate / Office	0.0015	Pier	0.0030	Corporate / Office	0.0145	Other - Entertainment	0.0045
Other - Entertainment	0.0014	Coffee Shop	0.0029	Stadium	0.0145	**Movie Theater**	0.0044
Bookstore	0.0013	Gym	0.0029	Burgers	0.0139	Grocery Store	0.0042

Events can be represented by their participants. They also feature characteristic peak times …

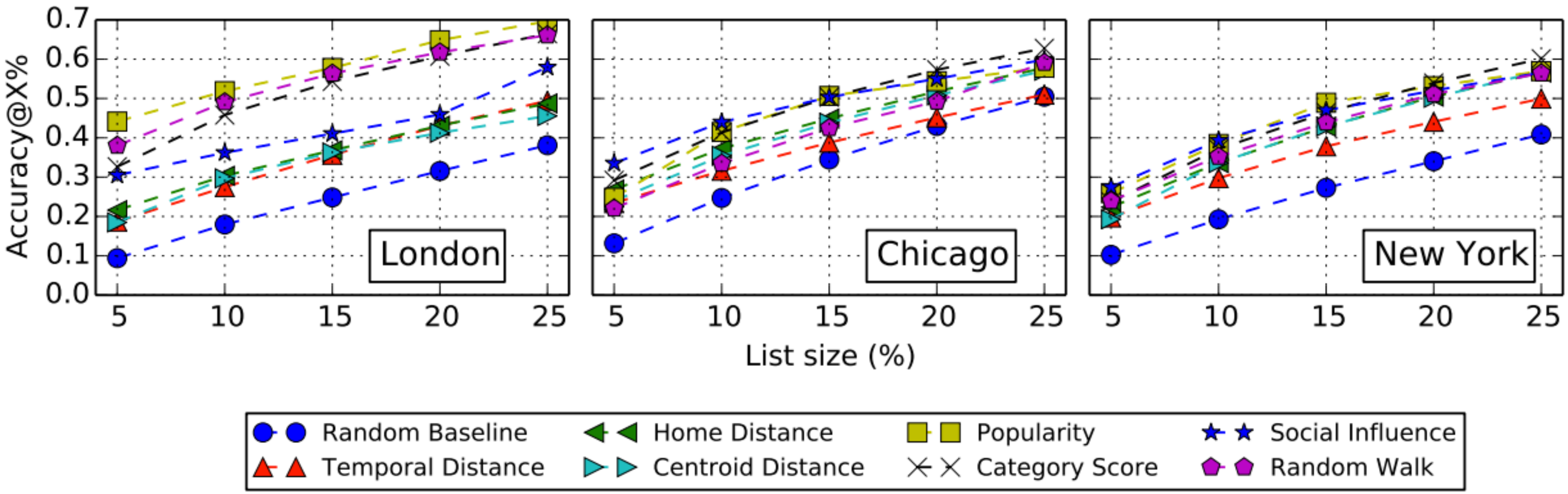


Event Representation (2)



Events are inherently social! We capture two types of social signal: explicit through social connections and an implicit homophily signal from place types.

Recommending Events



Popularity and Social Influence are important factors. But why Random Walk performs so well in London?

Model	London	Chicago	New York
Random	*0.118*	*0.142*	*0.115*
Temporal Dist.	0.203	0.221	0.194
Home Dist.	0.219	0.245	0.223
Category Score	0.315	0.267	0.235
Popularity	**0.411**	**0.275**	**0.262**
Social Influence	0.290	**0.306**	**0.268**
Random Walk	**0.347**	0.221	0.244

“Niche” Events

A niche event is one whose participants exhibit “special” interests with respect to the general population.

There is a negative correlation between the “nicheness” of an event and the accuracy of the Random Walk model.

London	Chicago	New York
-0.50^*	-0.38^*	-0.42^*

This is a kind reminder that events refer to a very abstract concept: their magnitude, duration, style etc. can vary significantly. This may be an indication that employing multiple strategies in a unifying framework
can provide a promising solution to event recommendation.

Supervised learning event recs...

NDCG

Model	London	Chicago	New York
Random	*0.118*	*0.142*	*0.115*
Popularity	0.411	0.275	0.262
Social Influence	0.290	0.306	0.268
LR	0.481	0.311	0.336
M5	0.494	0.346	0.344
LR + RWR	0.505	0.324	0.343
M5 + RWR	**0.528**	**0.363**	**0.367**

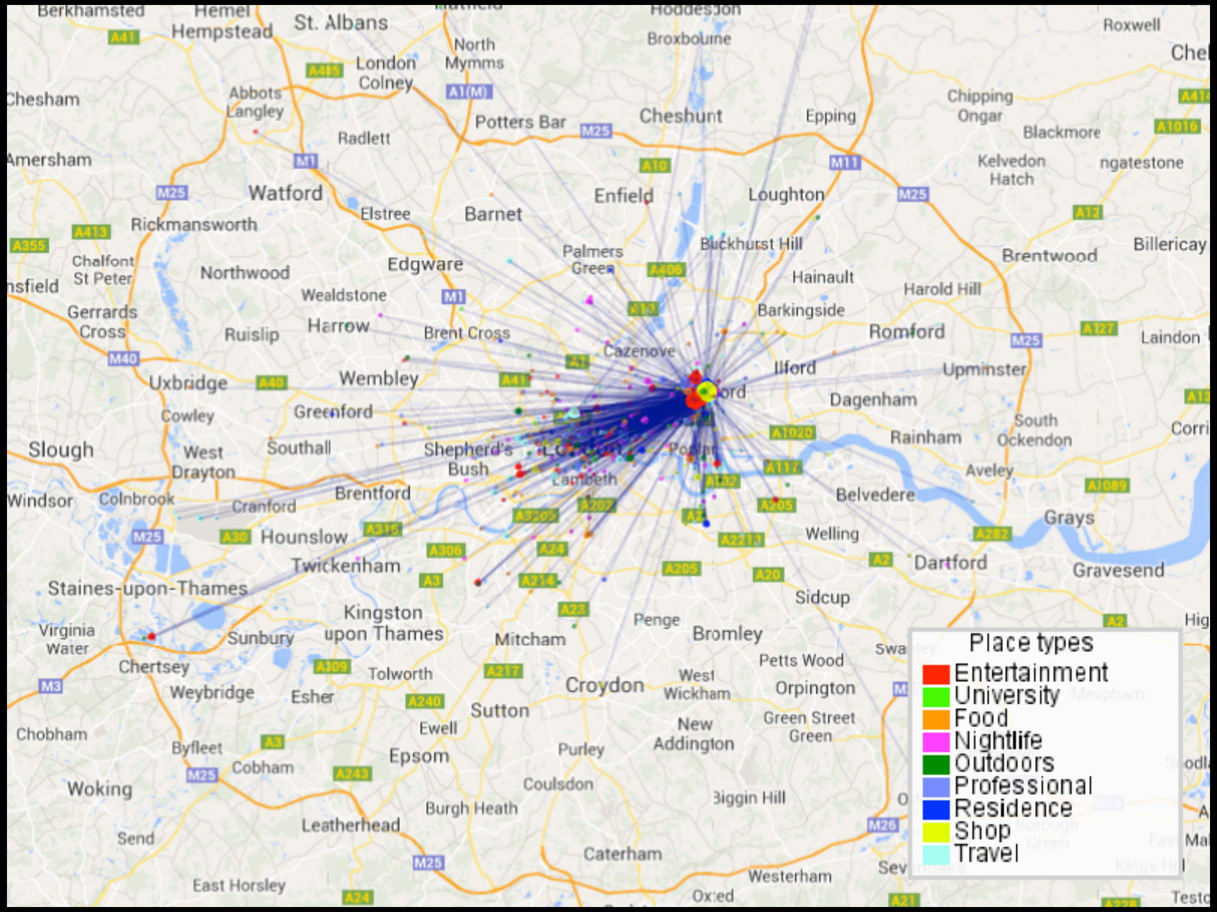
Exploiting the random walk signal improves the results across cities and metrics!

Accuracy

Model	London	Chicago	New York
Random	*0.037*	*0.051*	*0.036*
Popularity	0.267	0.168	0.151
Social Influence	0.220	0.198	0.160
LR	0.293	0.152	0.179
M5	0.344	0.205	0.185
LR + RWR	0.307	0.165	0.182
M5 + RWR	**0.372**	**0.229**	**0.212**

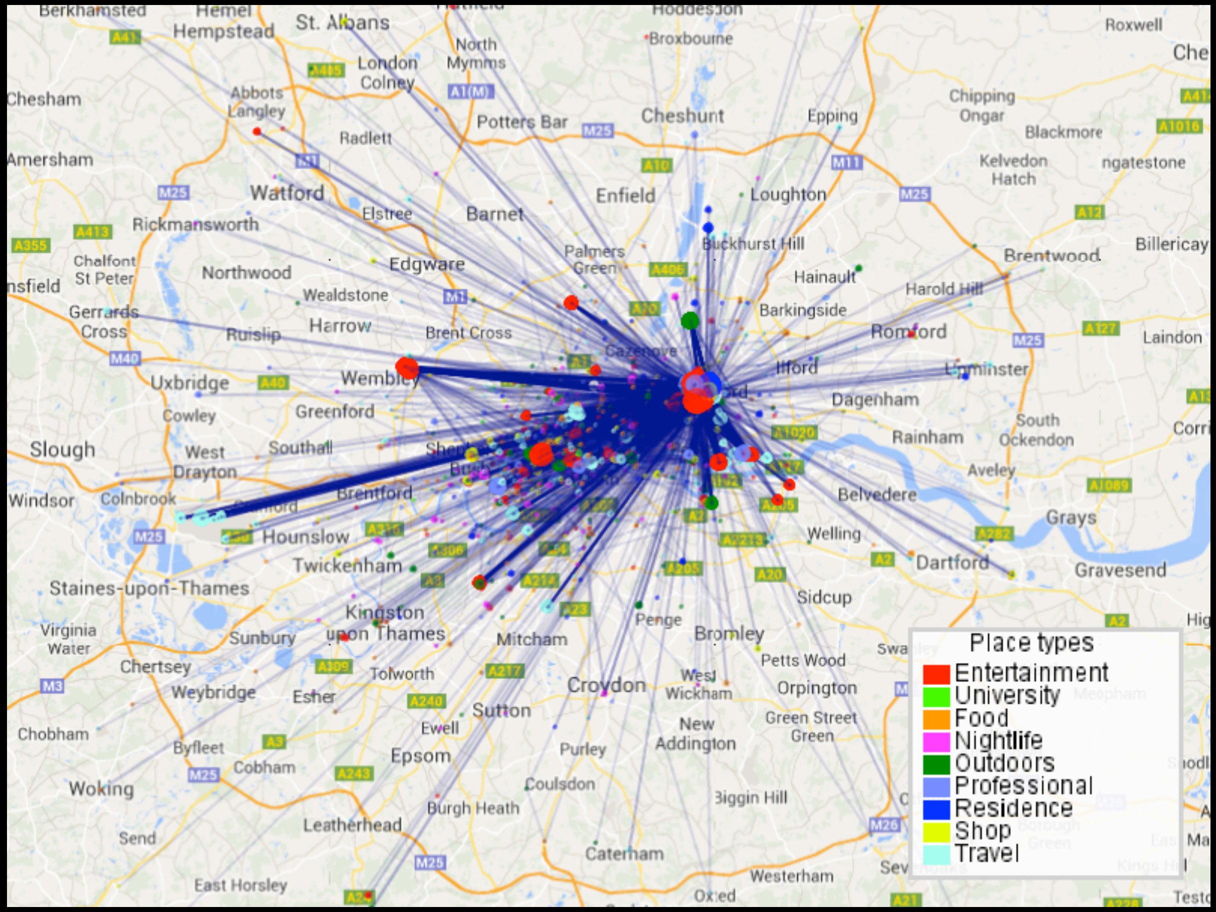
Mega Events: Assessing the impact of the Olympic Games

London



Pre-Olympic Period

London

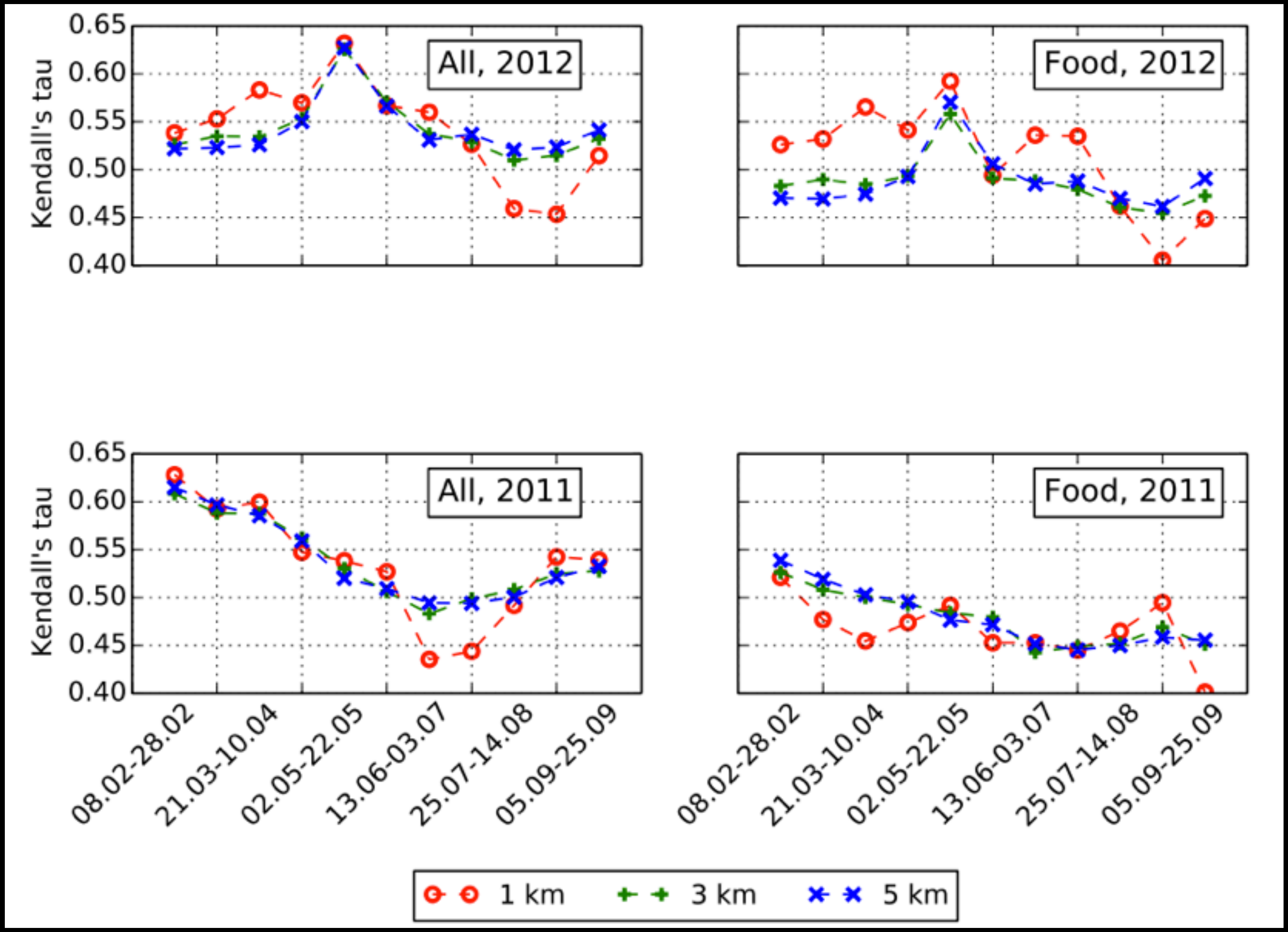


Olympic Period

Most influenced areas are near “hotspots”.

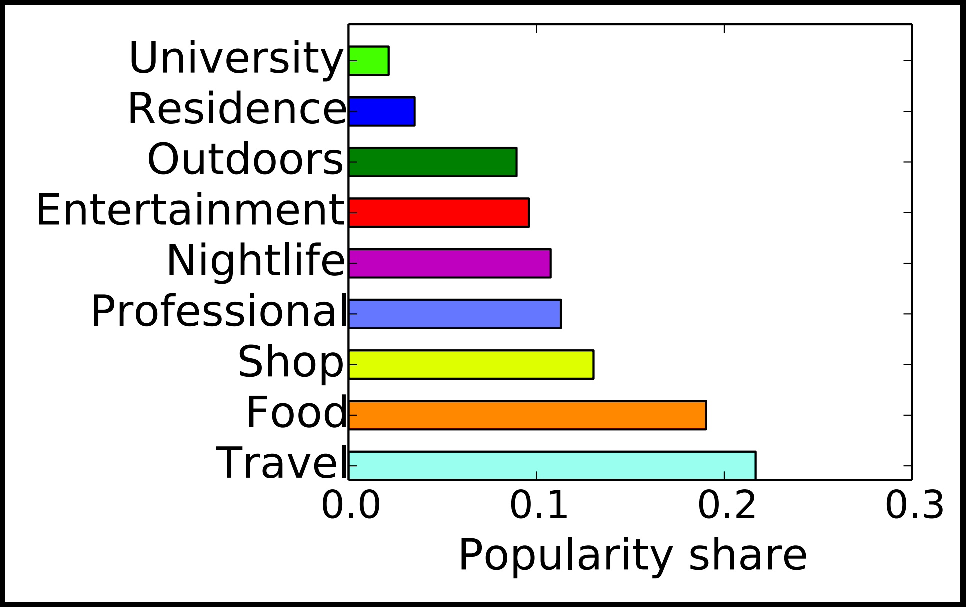
Olympic Hotspots

Place ranking correlations

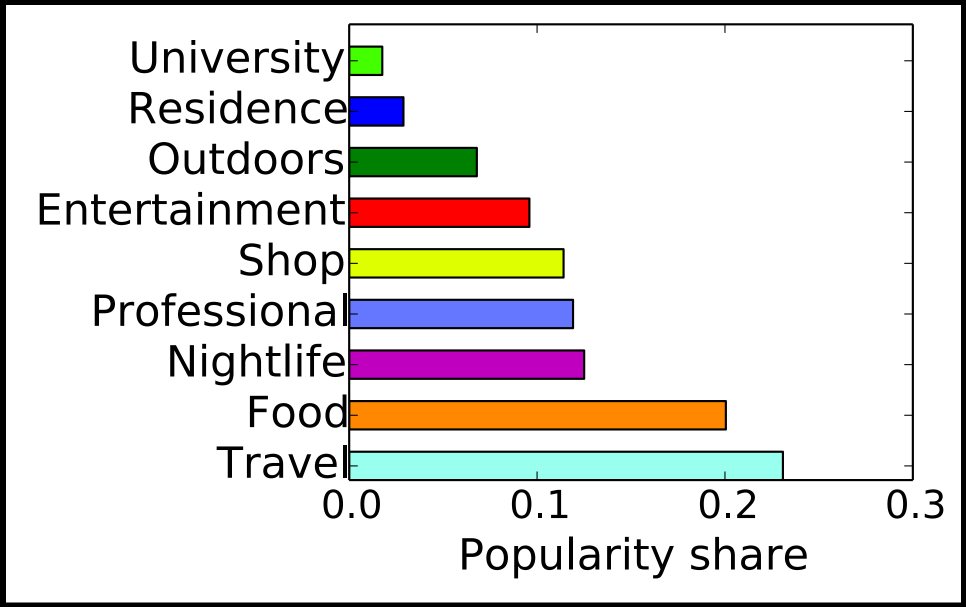


Kendall’s tau correlation coefficient as a function of distance to the nearest Olympic live site for the rankings in popularity of all and Food places between two subsequent periods.

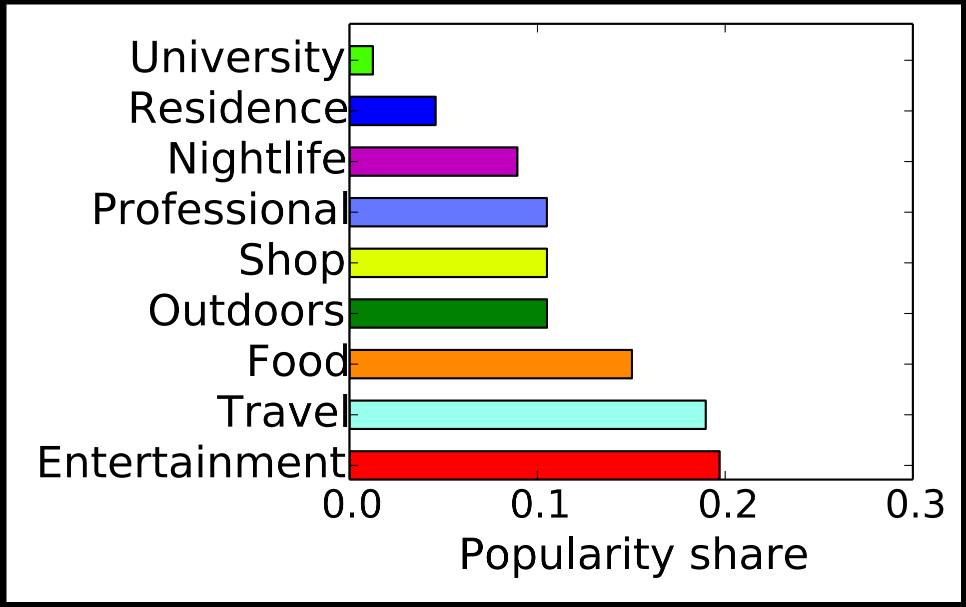
Place Category Popularity Change



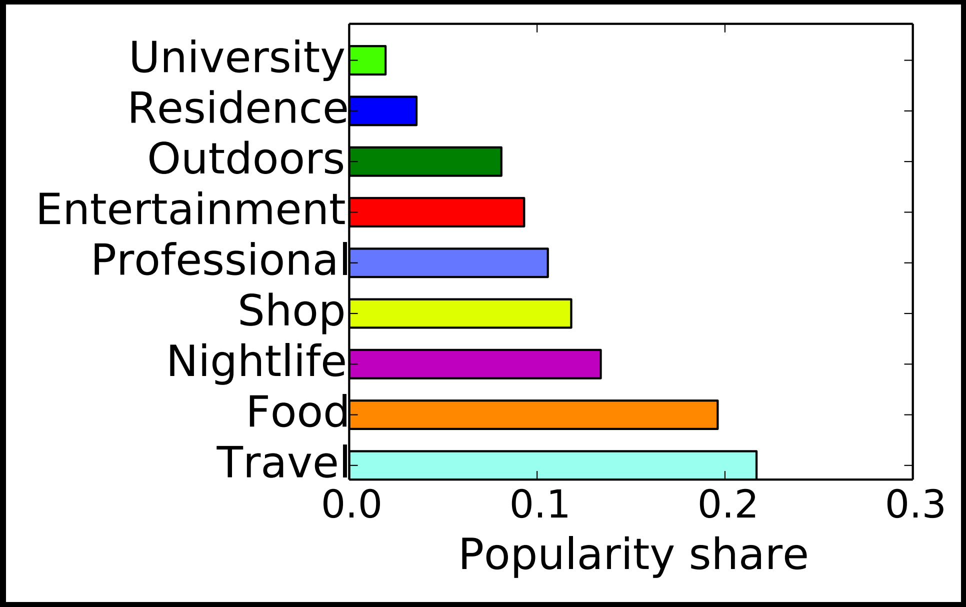
Pre-Olympic Period 2012



Pre-Olympic Period 2011

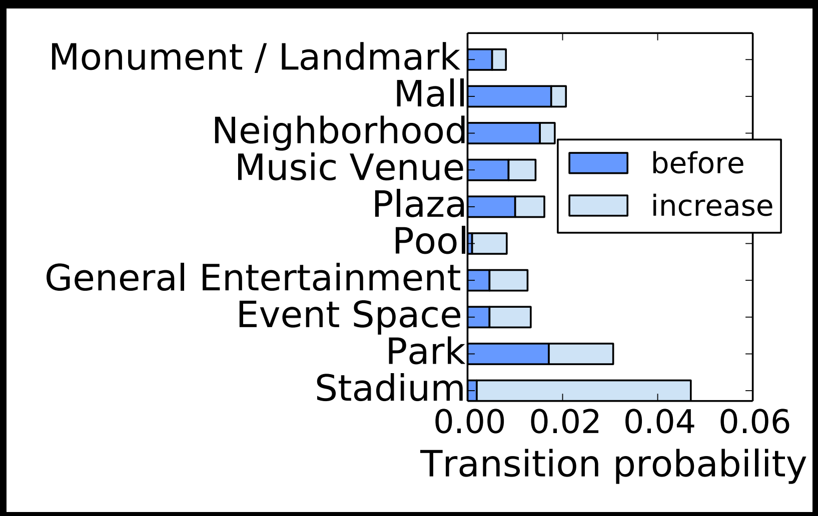


Olympics 2012

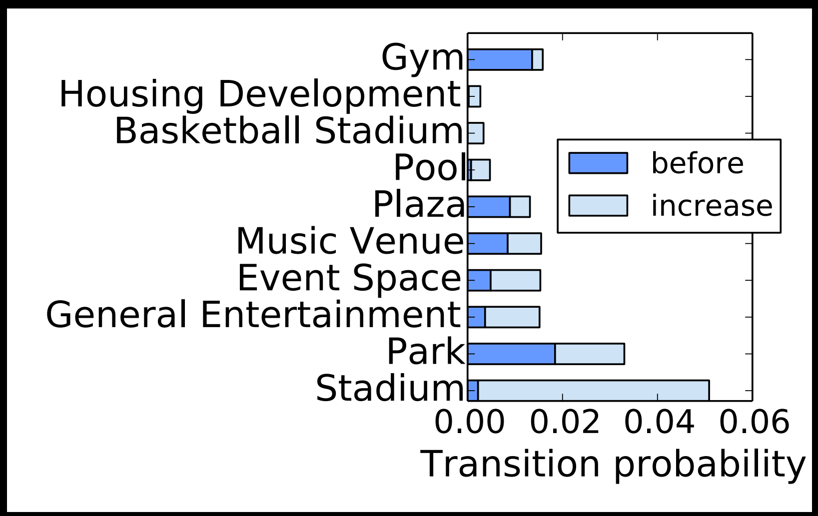


Olympic Period 2011

Network flow changes for food places



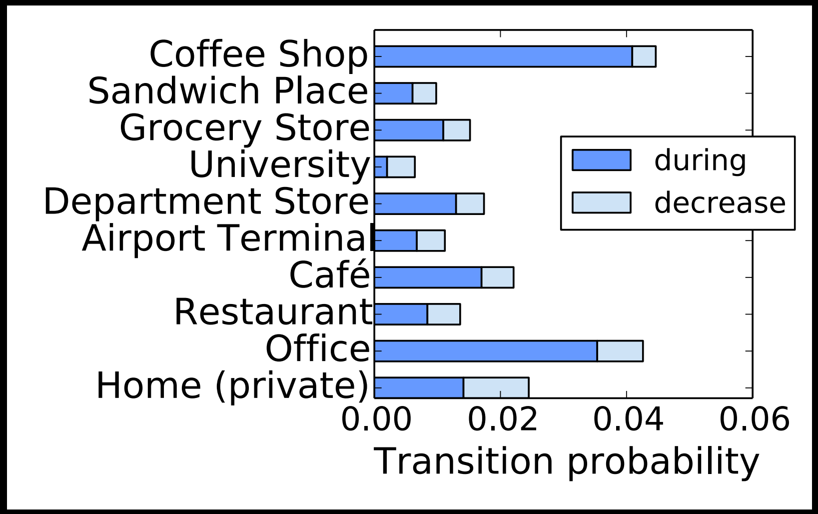
Food, in-flow increase



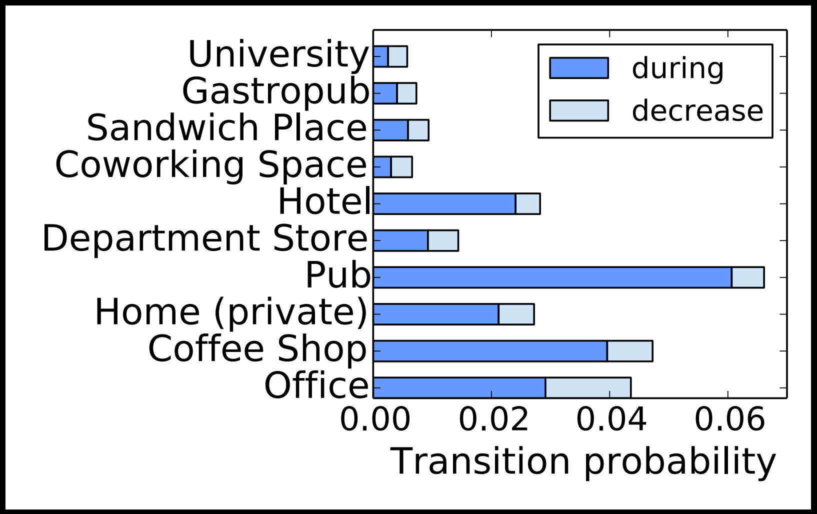
Food, out-flow increase

Transitions from and to Stadiums, General Entertainment facilities, Parks, Pools and Event Spaces are topping the charts with the biggest rise in popularity.

It is notable that these venue types are both sources and targets for the top increases in movements during the Games.



Food, in-flow decrease



Food, out-flow decrease

Data mining features to assess impact (1)

Olympic Distance	Measures the distance between a Food place and the closest event related hotspot.
Stadium Distance	Refines the above measure by measuring the distance to the closest Olympic Stadium.
Nearby Place Entropy	Measures the entropy in terms of place type frequency in the area around the target food spot.
Sponsor Distance	Measures the distance between the target food place and the nearest McDonald's restaurant.

Data mining features to assess impact (2)

Jensen Quality

Place type (t_v)	$k_{t_p \to t_v}$	**Place type** (t_v)	$k_{t_p \to t_v}$
Wine Shop	11.620	Rock Club	0.040
Tanning Salon	10.554	Mosque	0.046
Technology Building	9.582	Comedy Club	0.049
Car Wash	5.418	Dance Studio	0.055
Fish Market	4.217	Multiplex	0.057
Liquor Store	3.784	Flower Shop	0.063
BBQ Joint	3.700	History Museum	0.064
Latin Am. Restaurant	3.363	Fire Station	0.074
Library	3.342	Museum	0.081
Camera Store	3.320	Adm. Building	0.087

Entertainment Flow

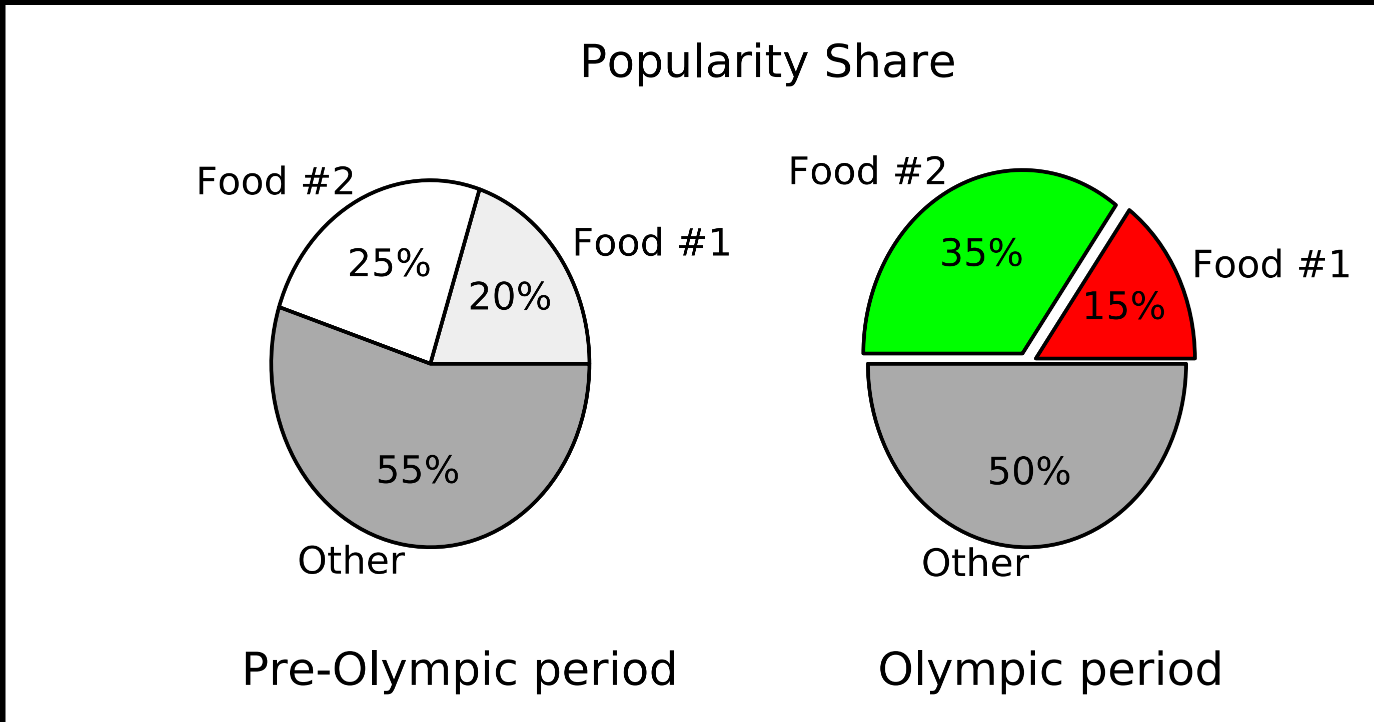
Measures the mean empirical probability of observing a user transition from an entertainment to another venue in the target area.

Area sociability

Simply count the number of friend pairs that have visited an area before the Olympic Games.

The city stock market

Can we exploit those features in order to decide which areas will increase their 'market' share and which will not?



Doing regression and aiming to predict the exact check-ins a place will receive could be an option but data is sparse and distributions are skewed.

$$AR_{i\tau} = R_{i\tau} - E_{i\tau}$$

Abnormal Returns are equal to actual returns minus the expected returns during the olympic period.

Assessing the predictability of multiple signals

Feature	Description	AUC
Random	Random case baseline	0.50
Geographic		
Olympic Distance	Distance to nearest hot spot	0.48
Stadium Distance	Distance to nearest stadium	**0.72**
Jensen Quality	Nearby area attractiveness	0.69
Nearby Place Entropy	Activity diversity in the area	**0.72**
Sponsor Distance	Distance to McDonald's	0.68
Mobility		
Popularity	Pre-Olympic # check-ins	0.56
Entertainment Flow	Transitions to ent. places	**0.71**
Social Area	# friend pairs in the area	**0.71**

Combing features (again :-))

Algorithm	Set	Precision	Recall	AUC
Naïve Bayes	G	0.60	0.74	0.69
	M	0.69	0.44	0.72*
	GM	0.74	0.63	0.72*
Random Forest	G	0.61	0.65	0.72*
	M	0.62	0.63	0.68
	GM	0.74	0.67	0.78*
SVM	G	0.68	0.65	0.74*
	M	**0.81**	0.74	0.79*
	GM	0.71	**0.76**	**0.80***

Events can be an urban earthquake.
Can we develop automated techniques to detect them?

How do we formalize this process?
What is the primary data structure?
What problems/applications do we target?

Place networks represent a very attractive direction for analysis and modelling: lineage of techniques/literature, data in place.

The epic split!

foursquare

FOURSQUARE

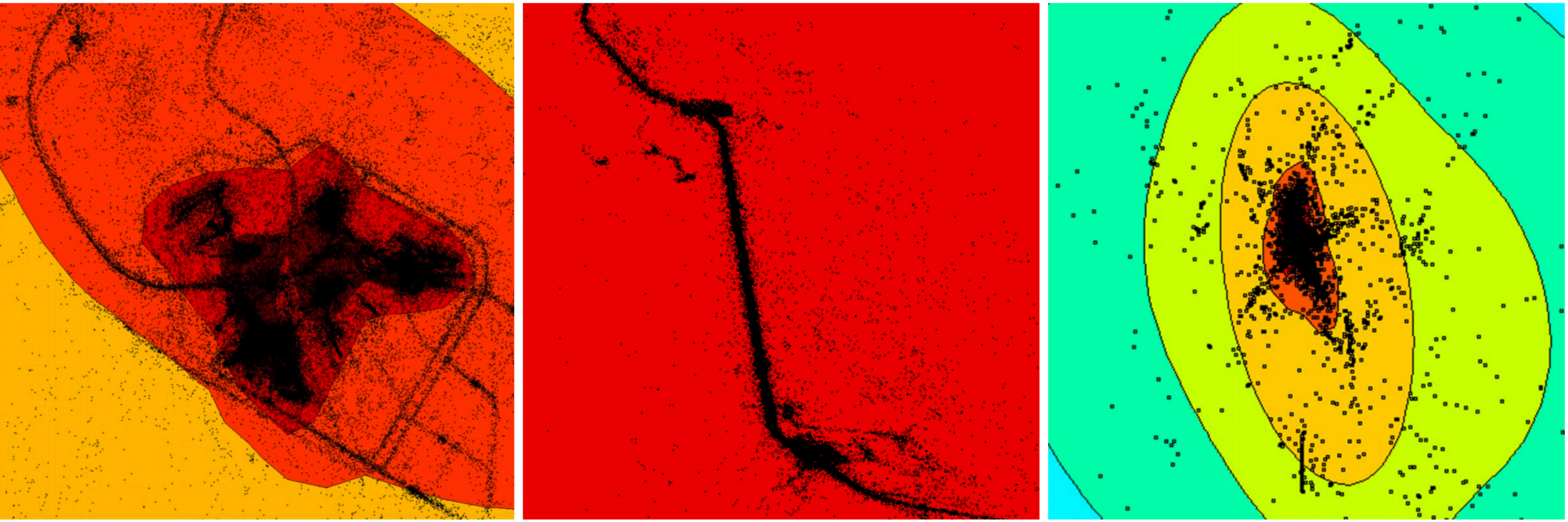
Swarm by FOURSQUARE

Mobile Sensing for Place Recommendations

Mobile Sensing & foursquare

- Accelerometer
- Ambient light
- Bluetooth
- Camera
- GPS
- Magnetometer
- Microphone
- Proximity
- WiFi

The "real" shape of places



(a) JFK Airport (b) Golden Gate Bridge (close up) (c) The Blind Tiger

http://courses.cse.tamu.edu/caverlee/csce670_2013/ltr-foursquare-paper.pdf

Bridging the Physical with the Digital 2.0

mobile sensing

on-line recommendations

foursquare

data and learning go both ways!

Opportunity: Exploit Foursquare's Venue database and add a sensing layer to it by aggregating samples from mobile users.

Challenges:
Sensing is costly. A lot of trade-offs to consider (energy).
Are users ready to share more of data (privacy)?

Big (Data) Disruptions in the City

or

WHICH TAXI SHALL I TAKE IN NYC?

founded: 1st of March 2009

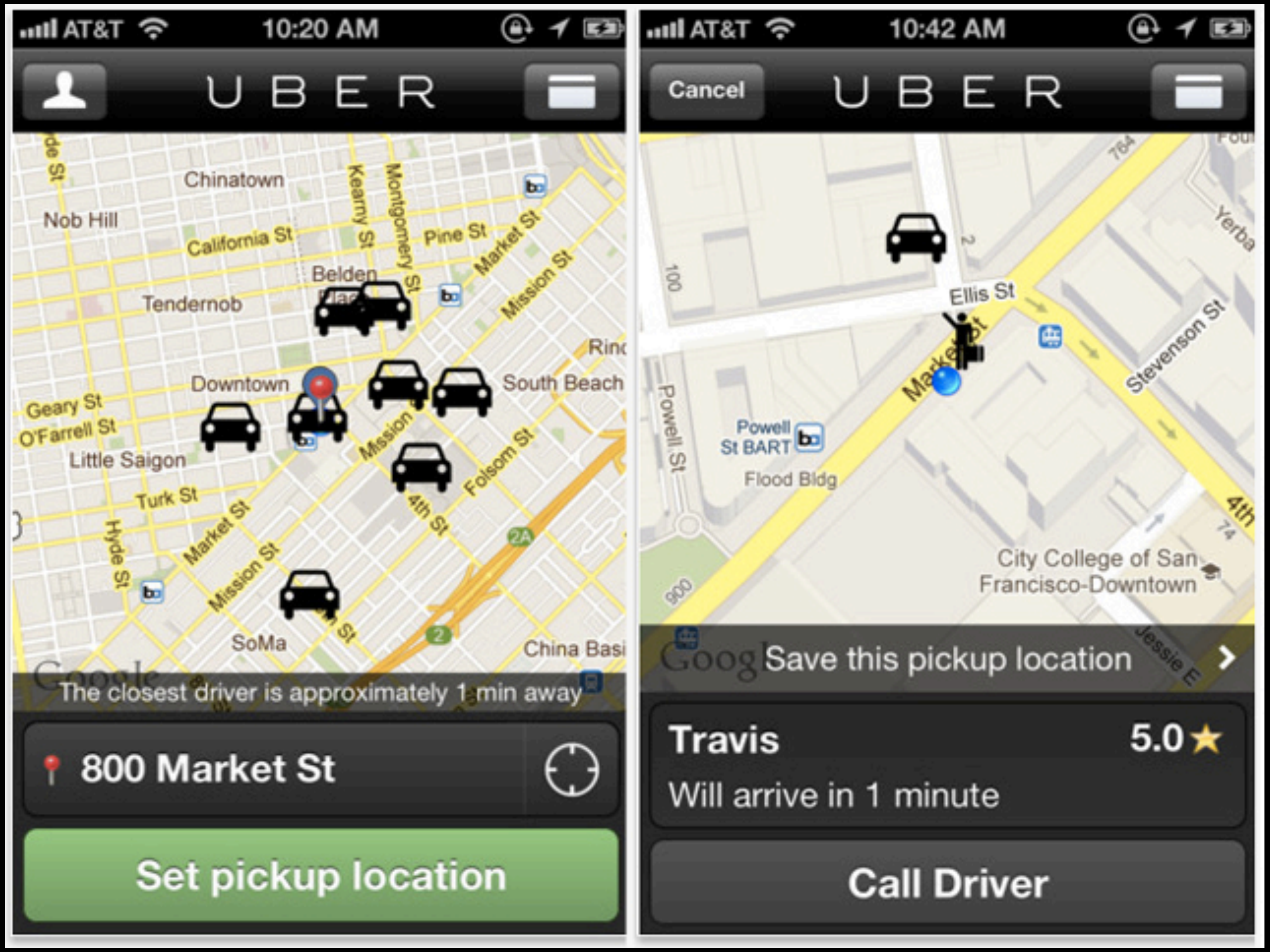
industry: transport

funding: 1.5 Billion $ in 6 rounds

funding: 1.2 Billion $ in June 2014

named company of the year
by USA Today in 2013

THE APP: "EVERYONE'S PRIVATE DRIVER"



NOT EVERYONE IS HAPPY ABOUT UBER

june 11, 2014 "In a concerted action, taxis blocked roads in major European cities in protest against what they perceive as a threat to their livehoods by companies such as uber. The cabbies contended that Uber and similar smartphone app-based services have an unfair advantage because they are not subject to price control and regulations.

july 2014 "Seoul city government said it will ban uber, joining the battle by municipalities and traditional taxi services..."

january 2014 "Cab drivers in Paris attacked uber driver in Charles de Gaulle Airport..."

uberX

NOW CHEAPER THAN A NEW YORK CITY TAXI

HOW THESE PRICES COMPARE

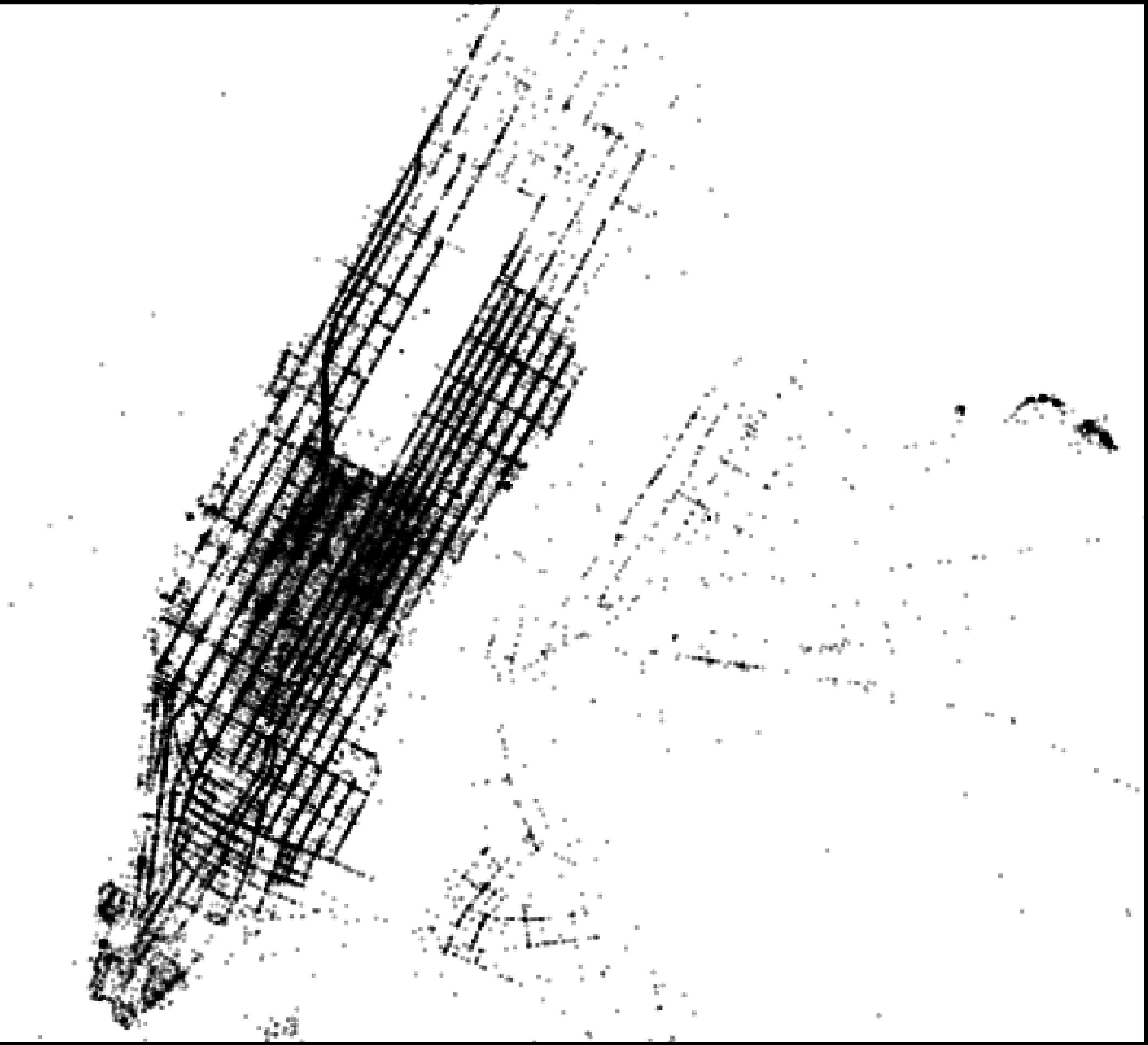
Williamsburg to East Village

KEEP IN MIND

These prices are only in effect for a limited time. The more you ride, the more likely we can keep them this low!

We know you may be asking yourself how this affects our partner drivers. What we've seen in cities across the country is that lower fares mean greater demand, lower pickup times and more trips per hour – increasing earning potential and creating better economics for drivers. What does what mean in the long run? They'll be making more than ever!

THE NEW YORK CITY TAXI DATASET



FOILing NYC's Taxi Trip Data

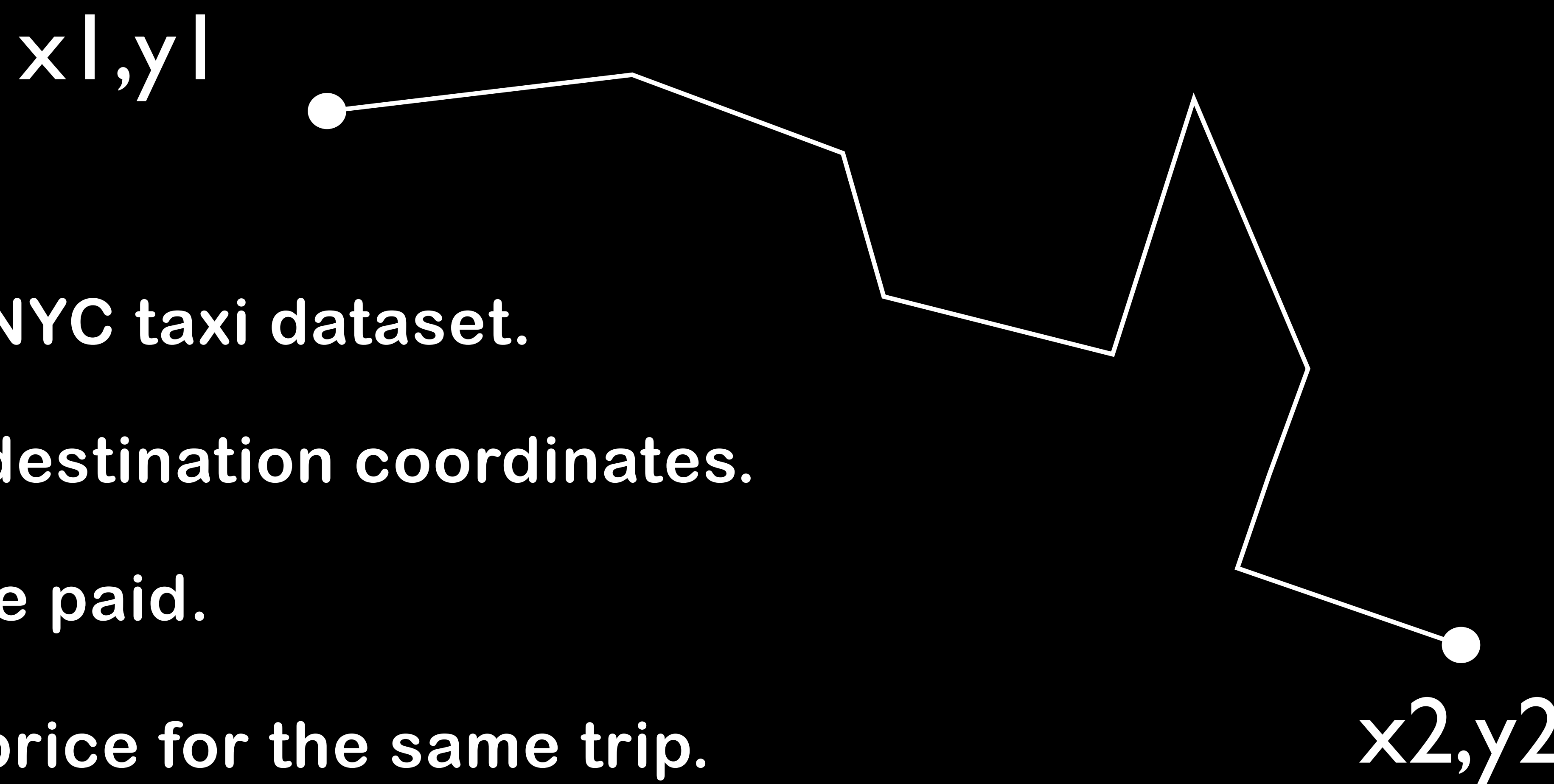
Freedom of Information Law

2013 Trip Data, 11GB, zipped!

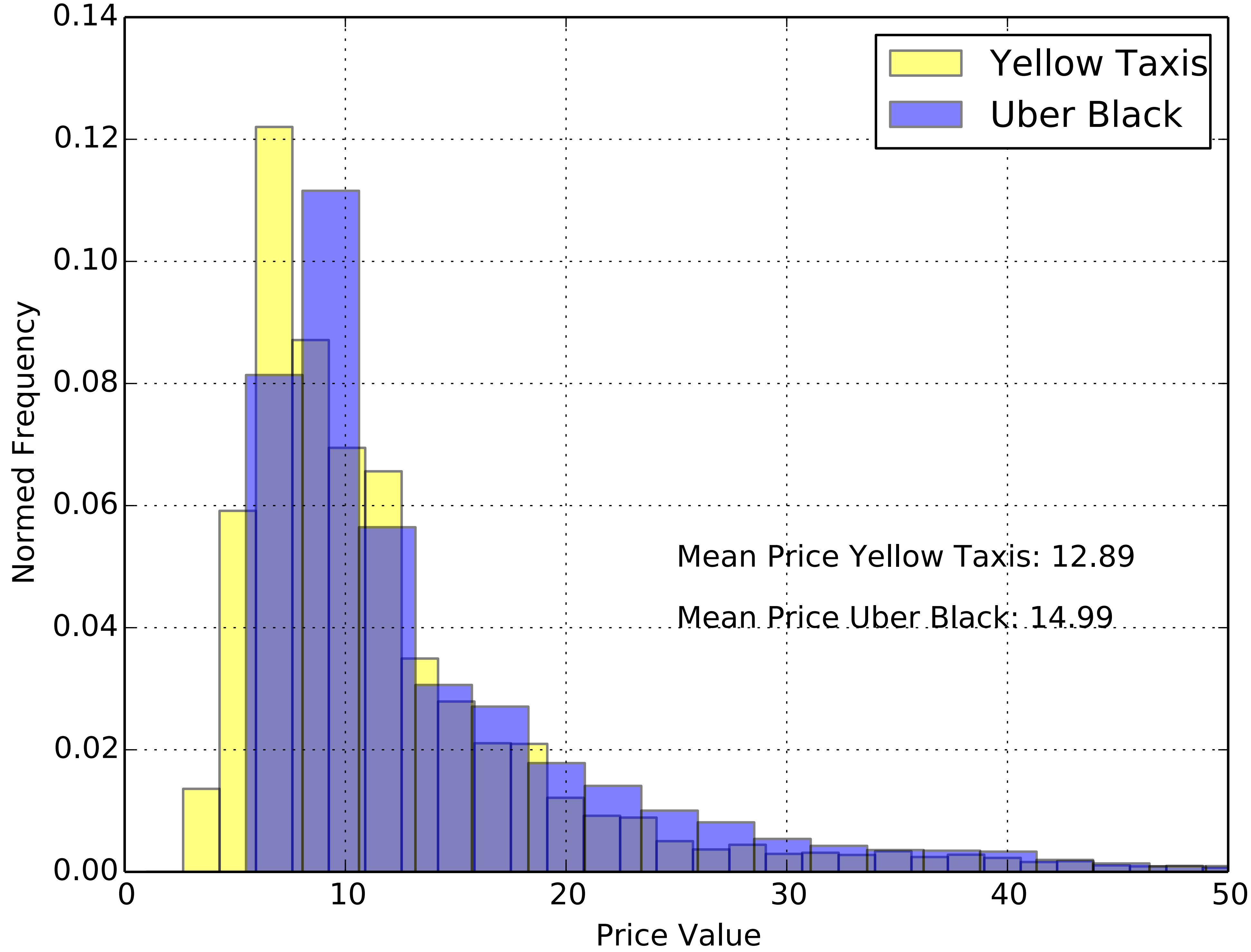
2013 Fare Data, 7.7GB

Idea: Uber Vs Yellow Taxi Price Comparison.

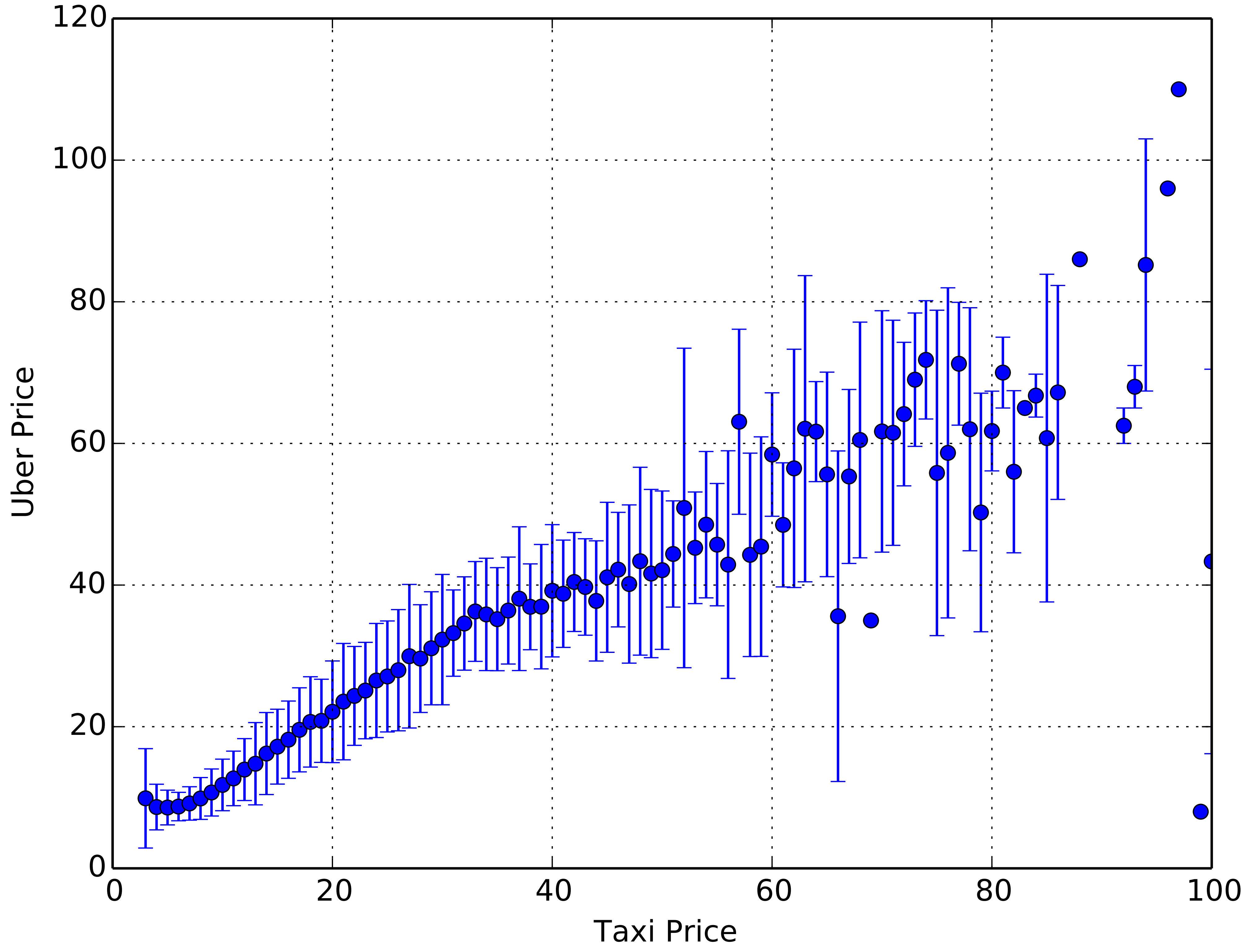
THE EXPERIMENT

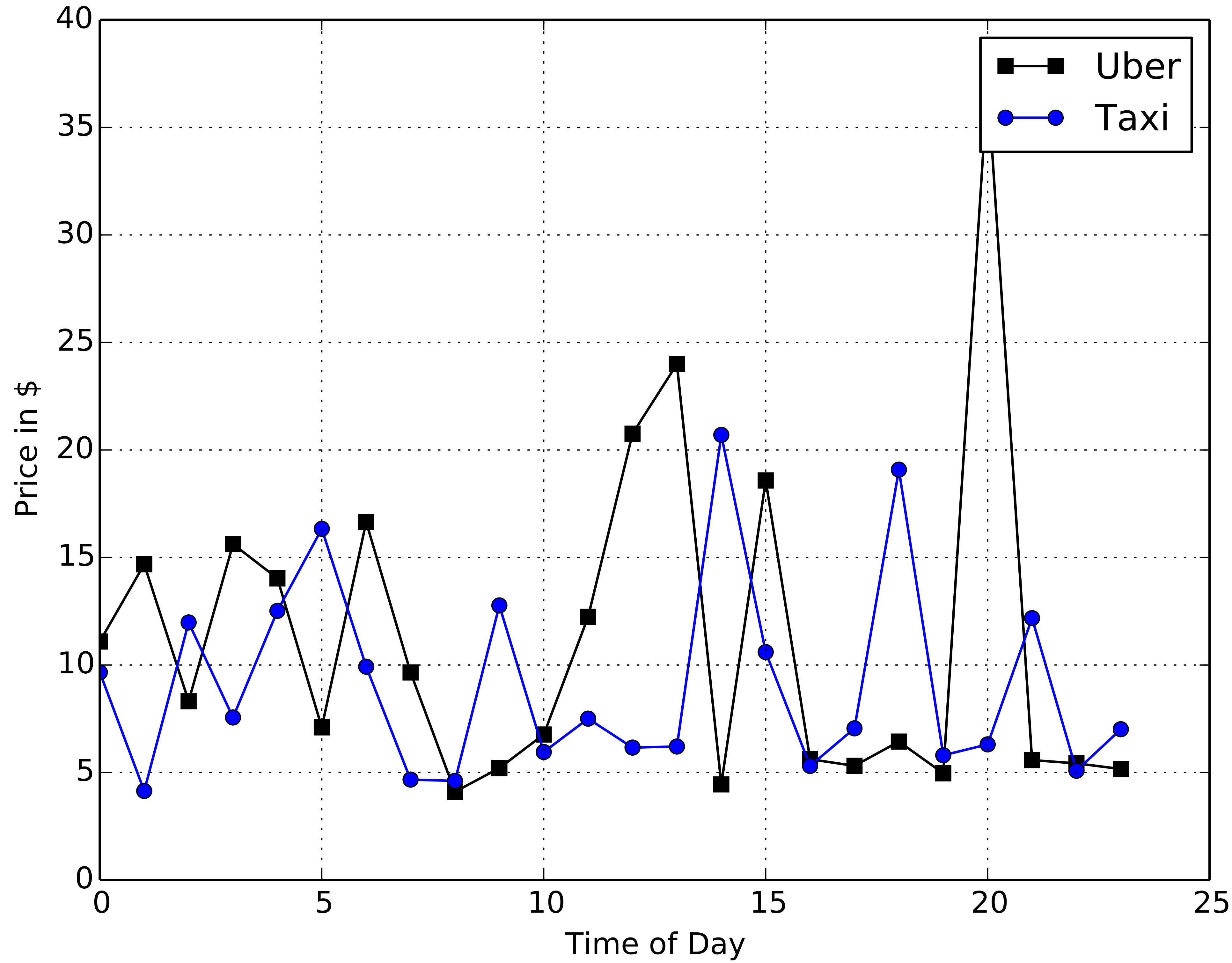


1. For every trip in NYC taxi dataset.
2. Record origin & destination coordinates.
3. Retrieve total fare paid.
4. Query Uber API price for the same trip.
5. Compare yellow taxi VS uber prices.

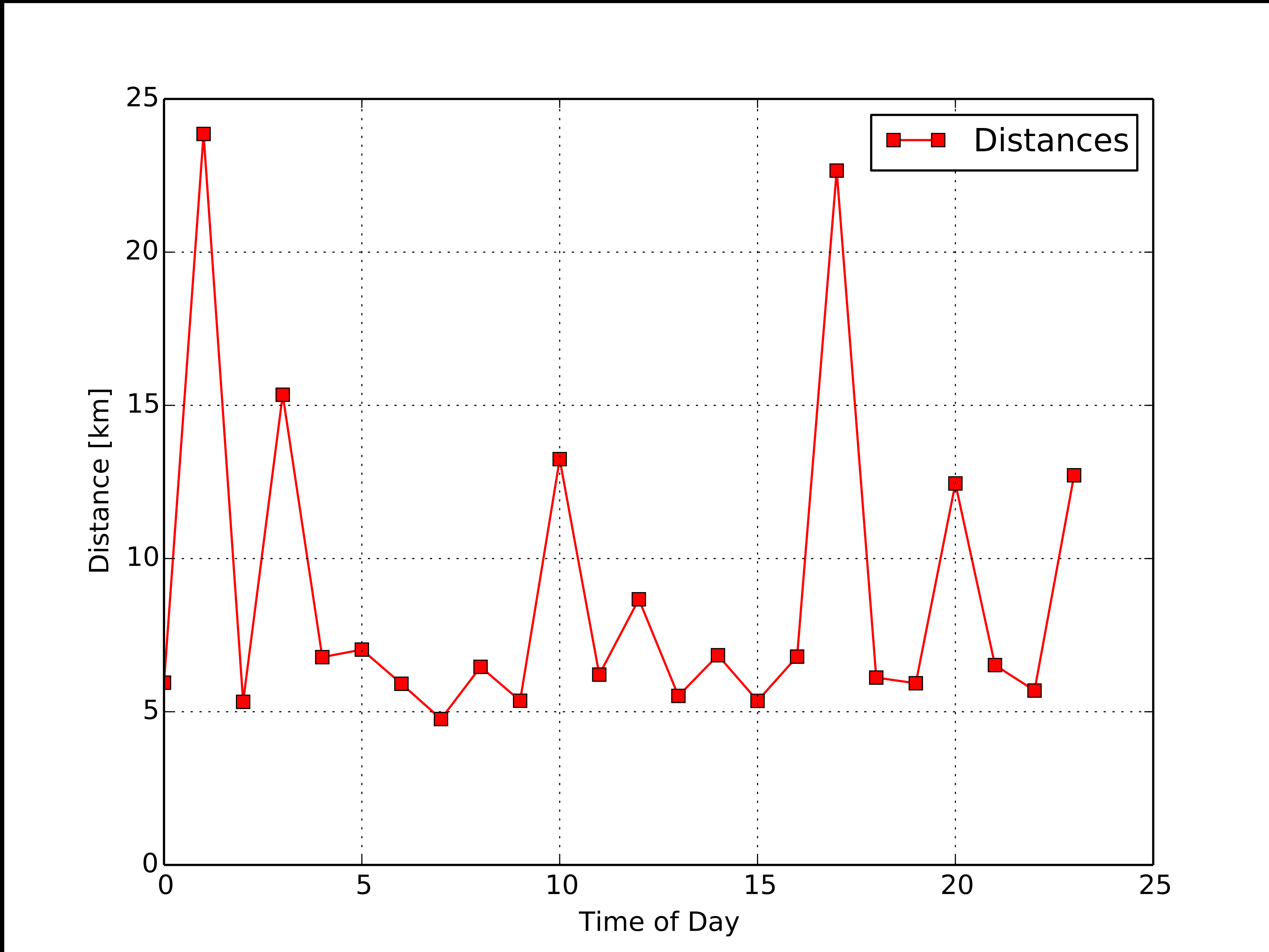


Yellow Taxis
Uber Black
Mean Price Yellow Taxis: 12.89
Mean Price Uber Black: 14.99
Normed Frequency
Price Value
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
10
20
30
40
50

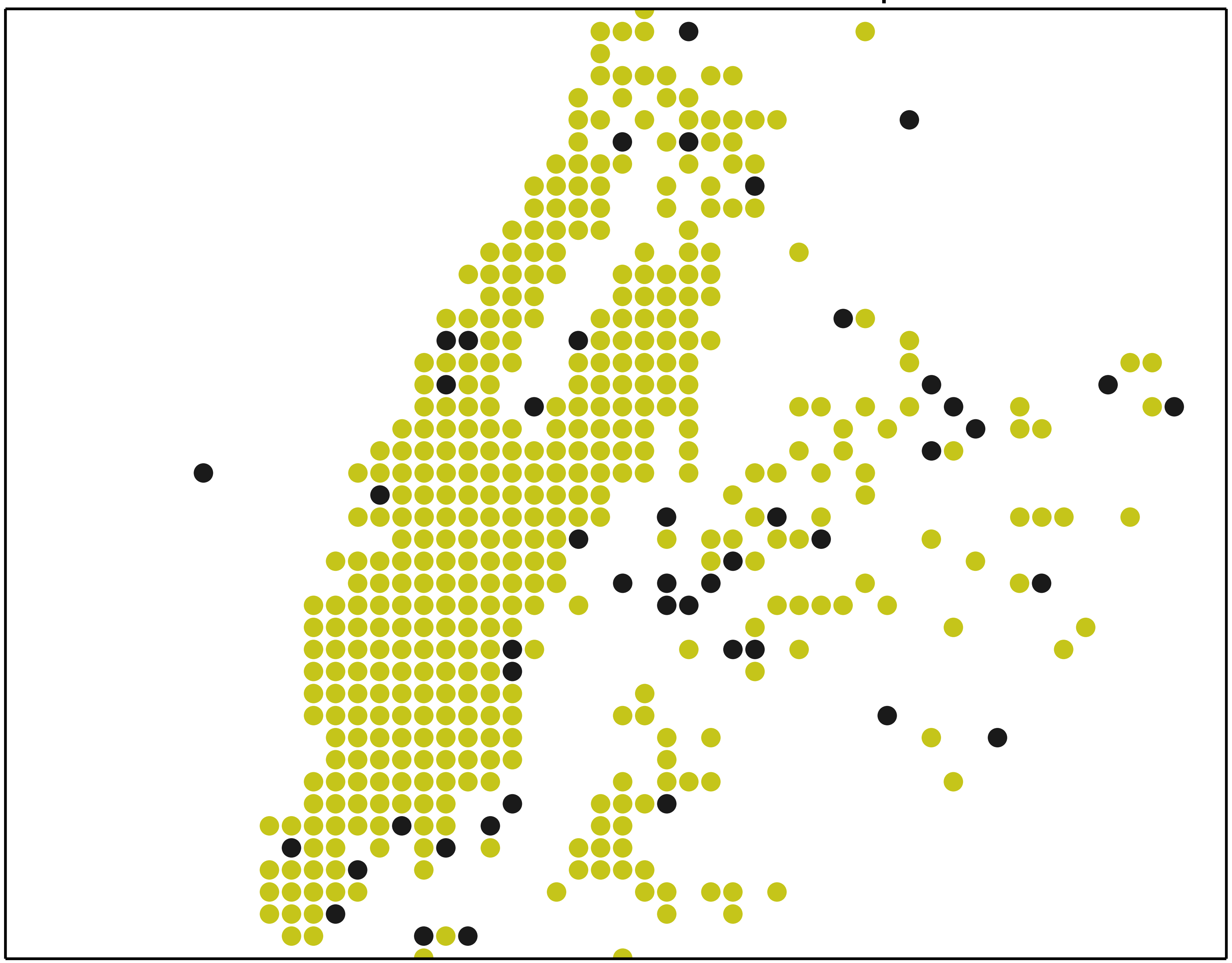

Uber Price
Taxi Price
0
20
40
60
80
100
120



Uber
Taxi
Price in $
Time of Day
0
5
10
15
20
25
30
35
40
0
5
10
15
20
25


Distances
Distance [km]
Time of Day
0
5
10
15
20
25
0
5
10
15
20
25

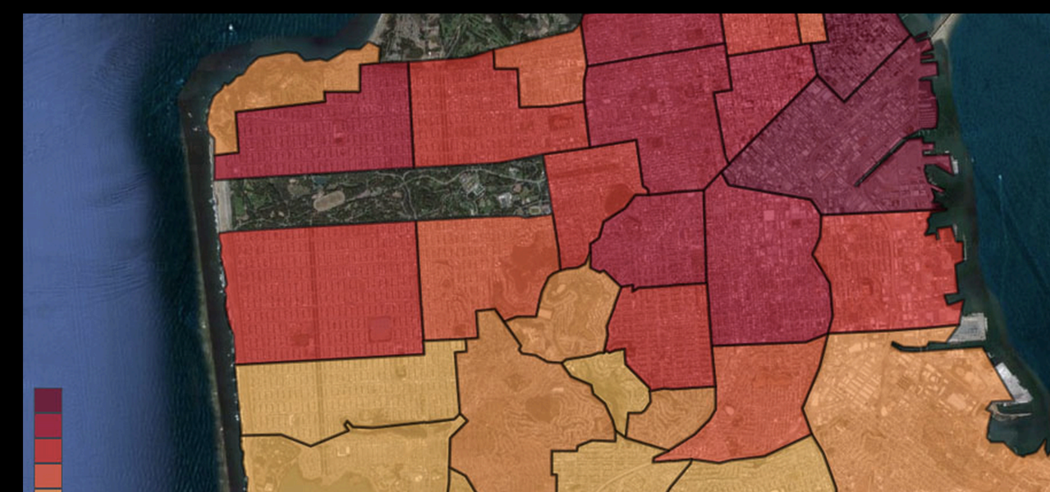
Yellow Taxis VS Uber - Price Comparison



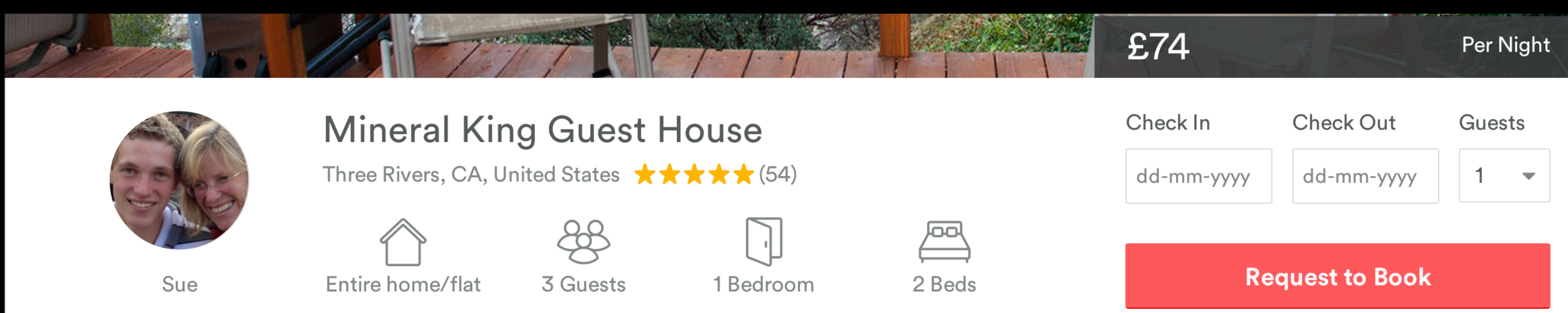
New Research Opportunities

- real-time services demand in the city

2	3	?
?	3	2
1	2	?



- big data predictions in the city



THE END

Where Businesses Thrive: Predicting the Impact of the Olympic Games on Local Retailers through Location-based Services Data.
Petko Georgiev, Anastasios Noulas, Cecilia Mascolo.
In Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM 2014).

The Call of The Crowd: Event Participation in Location-based Social Services.
Petko Georgiev, Anastasios Noulas, Cecilia Mascolo.
In Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM 2014).

Hoodsquare: Modeling and Recommending Neighborhoods in Location-based Social Networks.
Amy X. Zhang, Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo.
In Proceedings of the IEEE/ASE International Conference on Social Computing (SocialCom 2013).

Geo-Spotting: Mining Online Location-based Services for Optimal Retail Store Placement.
Dmytro Karamshuk, Anastasios Noulas, Salvatore Scellato, Vincenzo Nicosia, Cecilia Mascolo.
In Proceedings of 19th ACM International Conference on Knowledge Discovery and Data Mining (KDD 2013)

Mining User Mobility Features for Next Place Prediction in Location-based Services.
Anastasios Noulas, Salvatore Scellato, Neal Lathia and Cecilia Mascolo.
In Proceedings of IEEE International Conference on Data Mining (ICDM 2012).

A Random Walk Around the City: New Venue Recommendation in Location-Based Social Networks.
Anastasios Noulas, Salvatore Scellato, Neal Lathia and Cecilia Mascolo.
In Proceedings of ASE/IEEE International Conference on Social Computing (SocialCom 2012).

A tale of many cities: universal patterns in human urban mobility.
Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, Cecilia Mascolo.
In PLoS ONE. PLoS ONE 7(5): e37027. doi:10.1371/journal.pone.0037027

An Empirical Study of Geographic User Activity Patterns in Foursquare
Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo and Massimiliano Pontil.
In Proceedings of Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011).

Exploiting Semantic Annotations for Clustering Geographic Areas and Users in Location-based Social Networks
Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo and Massimiliano Pontil.
In Proceeding of 3rd Workshop Social Mobile Web (SMW'11). Colocated with ICWSM 2011