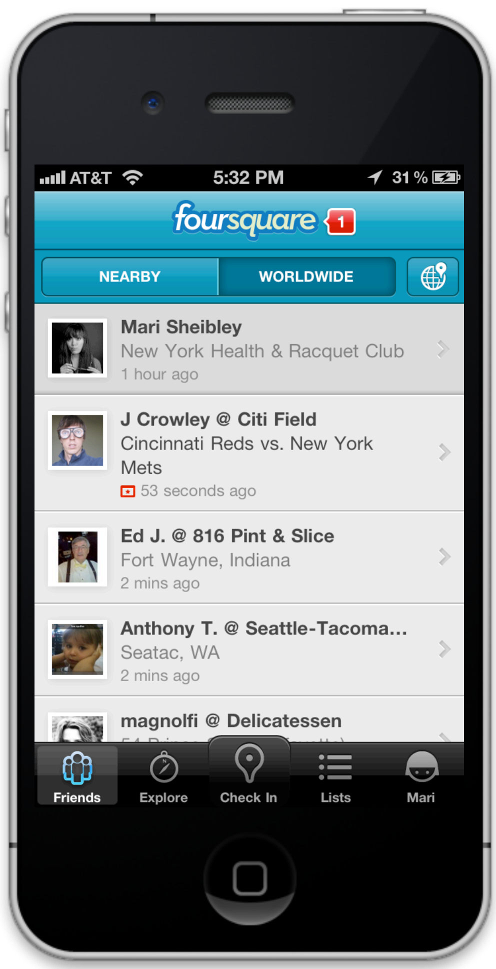
Machine Learning with Large Networks of People and Places

Blake Shaw, PhD
Data Scientist @ Foursquare
@metablake

foursquare®

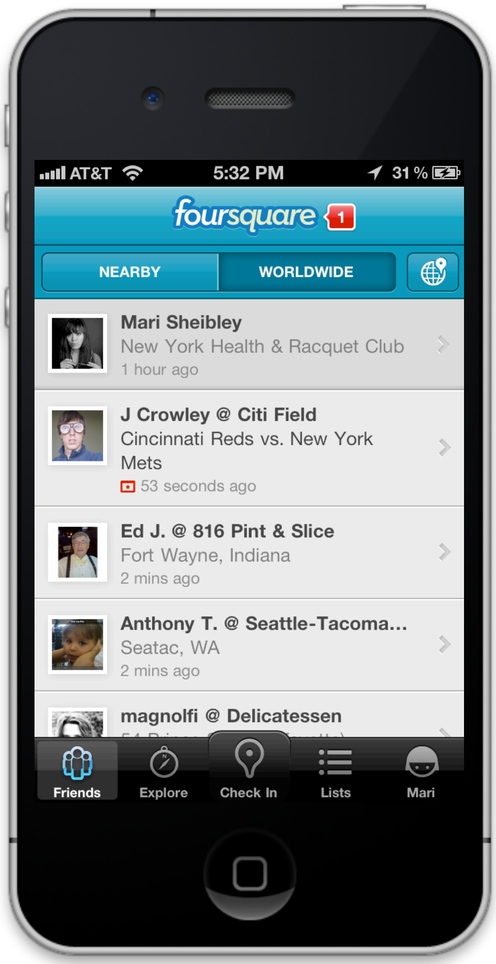
What is foursquare?



An app that helps you explore your city and connect with friends

A platform for location based services and data

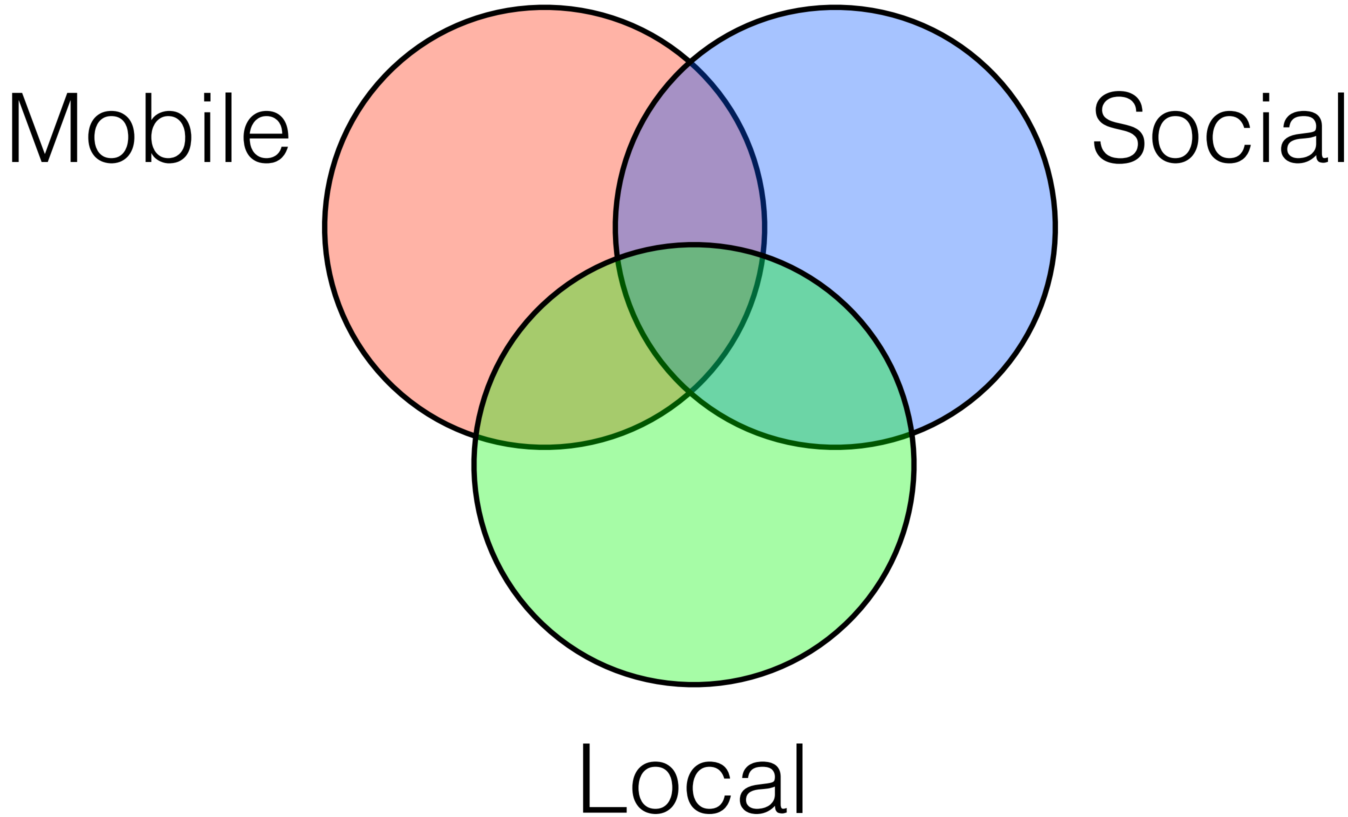
What is foursquare?



People use foursquare to:

- share with friends
- discover new places
- get tips
- get deals
- earn points and badges
- keep track of visits

What is foursquare?



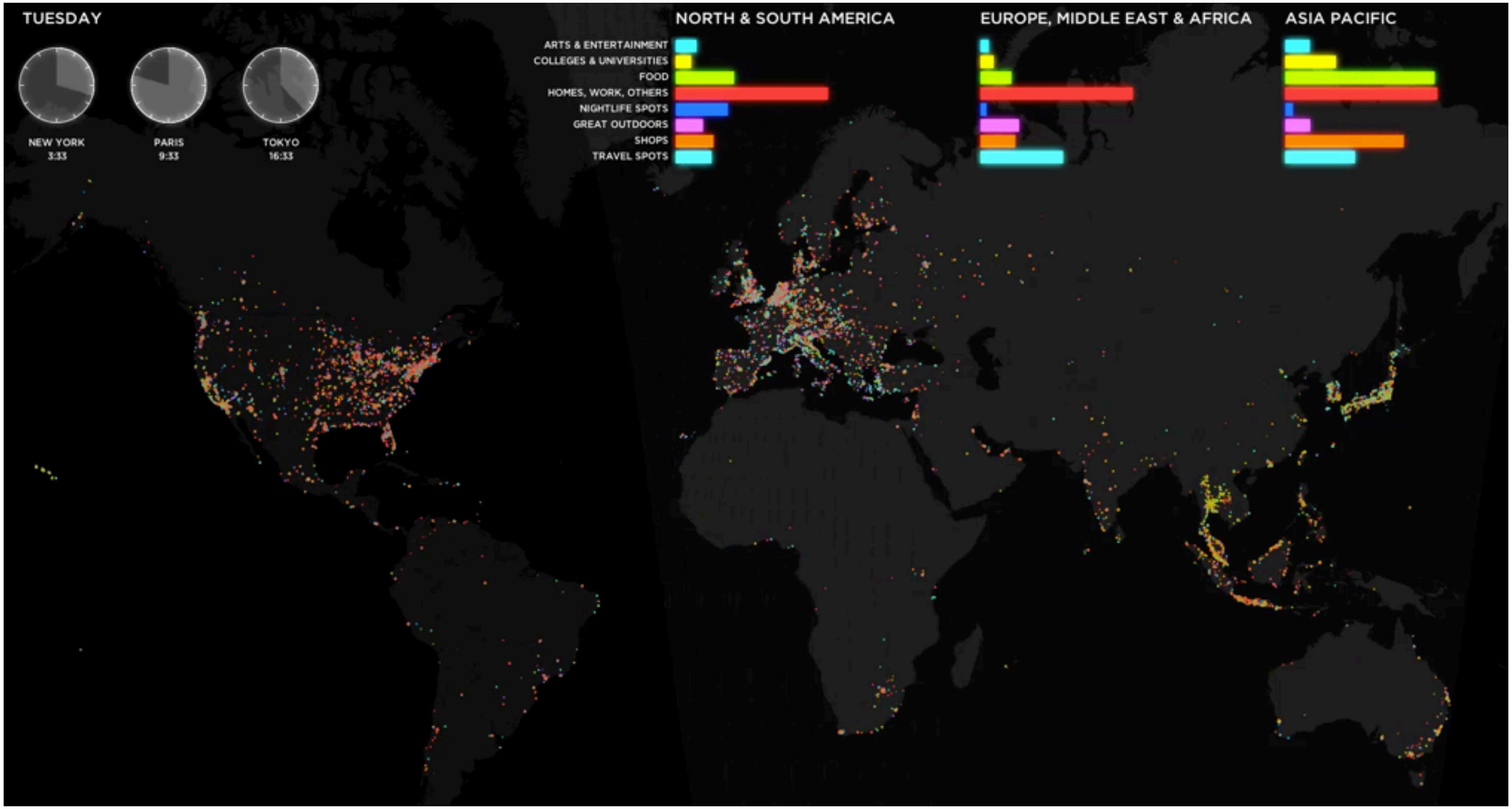
Stats

15,000,000+ people

30,000,000+ places

1,500,000,000+ check-ins

1500+ actions/second



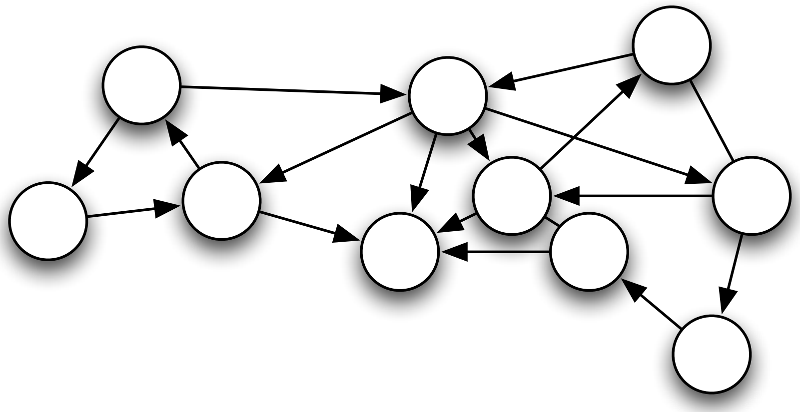
Video: http://vimeo.com/29323612

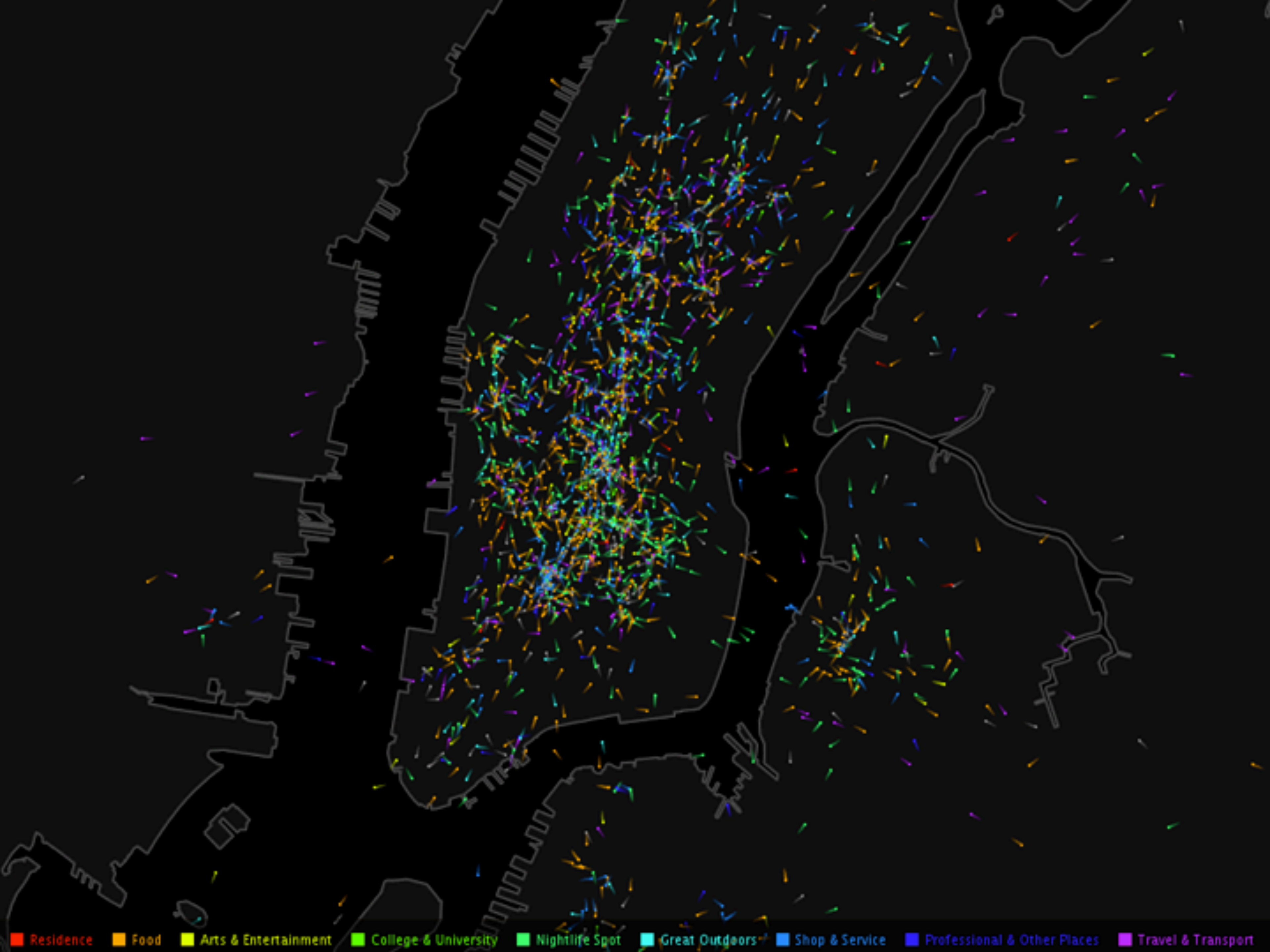
Overview

- Intro to Foursquare Data
- Place Graph
- Social Graph
- Explore
- Conclusions

The Place Graph

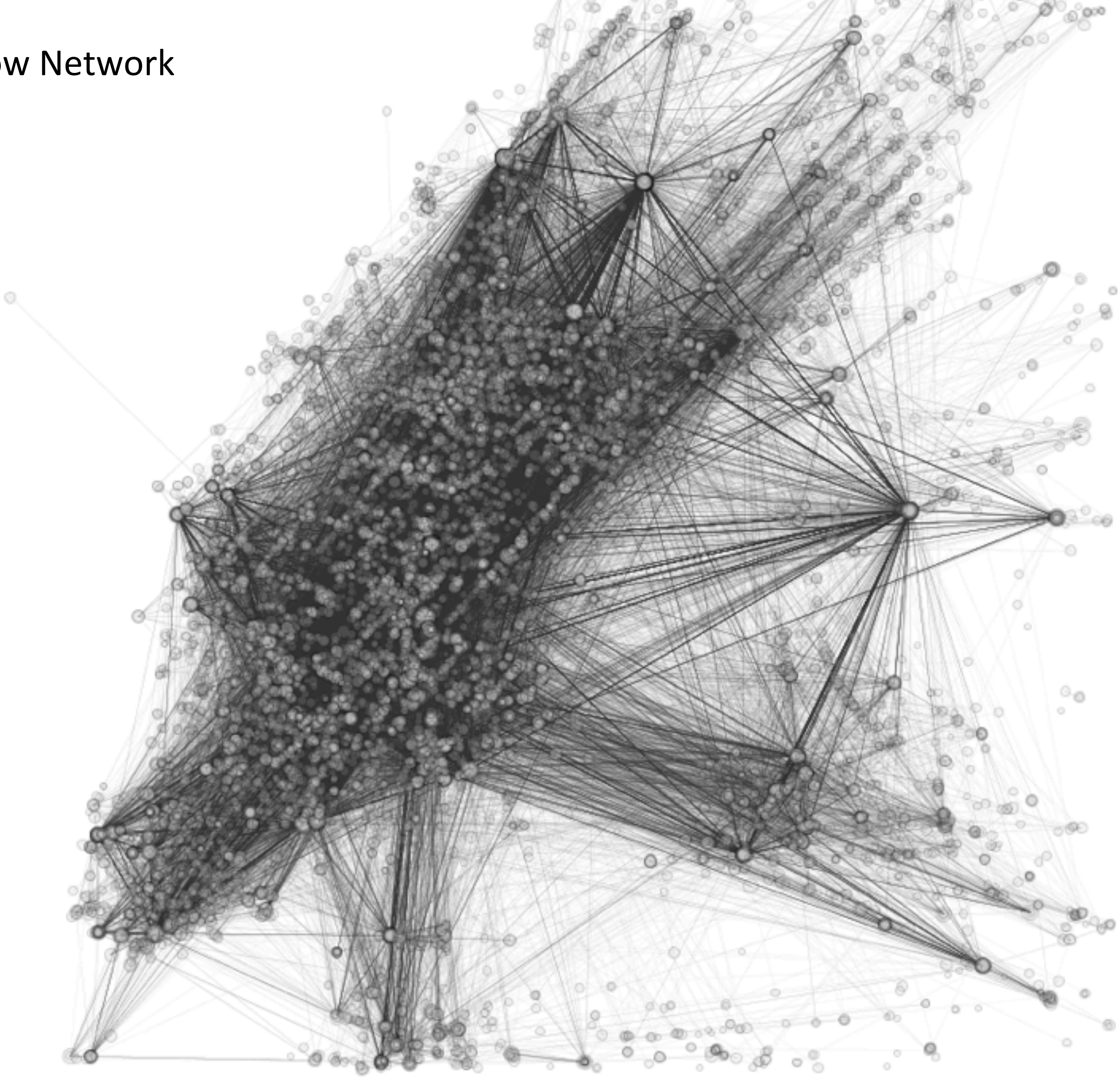
- 30m places interconnected w/ different signals:
  - flow
  - co-visitation
  - categories
  - menus
  - tips and shouts





Residence
Food
Arts & Entertainment
College & University
Nightlife Spot
Great Outdoors
Shop & Service
Professional & Other Places
Travel & Transport

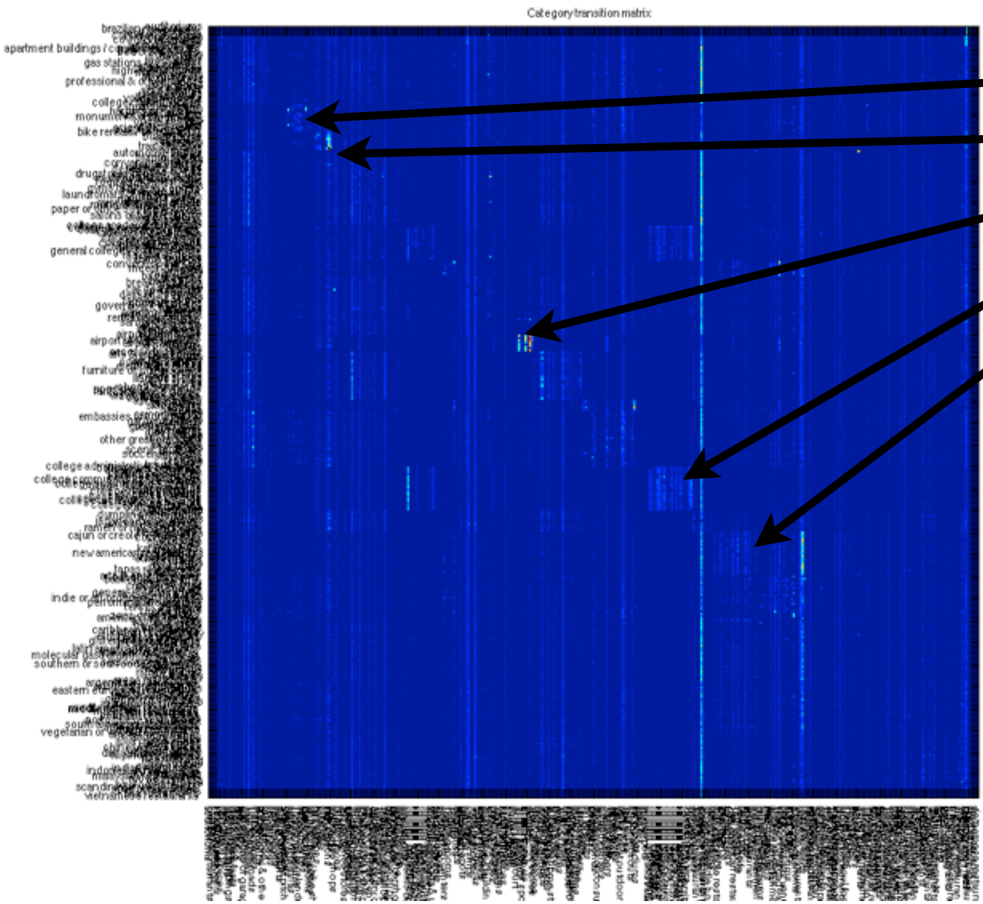
NY Flow Network



People connect places over time

- Places people go after the Museum of Modern Art (MOMA):
  - MOMA Design Store, Metropolitan Museum of Art, Rockefeller Center, The Modern, Abby Aldrich Rockefeller, Sculpture Garden, Whitney Museum of American Art, FAO Schwarz
- Places people go after the Statue of Liberty:
  - Ellis Island Immigration Museum, Battery Park, Liberty Island, National September 11 Memorial, New York Stock Exchange, Empire State Building

Predicting where people will go next



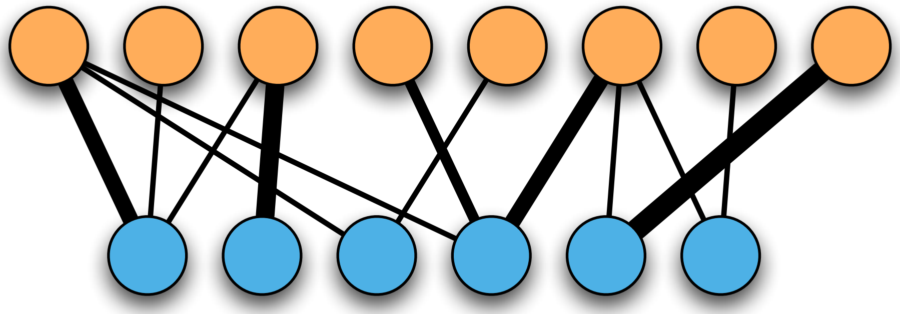
- Cultural places (landmarks etc.)
- Bus stops, subways, train stations
- Airports
- College Places
- Nightlife

After “bars”: american restaurants, nightclubs, pubs, lounges, cafes, hotels, pizza places

After “coffee shops”: offices, cafes, grocery stores, dept. stores, malls

Collaborative filtering

How do we connect people to new places they'll like?



Collaborative filtering

[Koren, Bell '08]

- Item-Item similarity
  - Find items which are similar to items that a user has already liked
- User-User similarity
  - Find items from users similar to the current user
- Low-rank matrix factorization
  - First find latent low-dimensional coordinates of users and items, then find the nearest items in this space to a user

Collaborative filtering

- Item-Item similarity
  - Pro: can easily update w/ new data for a user
  - Pro: explainable e.g “people who like Joe’s pizza, also like Lombardi’s”
  - Con: not as performant as richer global models
- User-User similarity
  - Pro: can leverage social signals here as well... similar can mean people you are friends with, whom you’ve colocated with, whom you follow, etc...

Finding similar items

- Large sparse k-nearest neighbor problem
  - Items can be places, people, brands
  - Different distance metrics
  - Need to exploit sparsity otherwise intractable

Finding similar items

- Metrics we find work best for recommending:
  - Places: cosine similarity

$$\mathrm{sim}(\mathbf{x}_i, \mathbf{x}_j) = \frac{\mathbf{x}_i \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|}$$

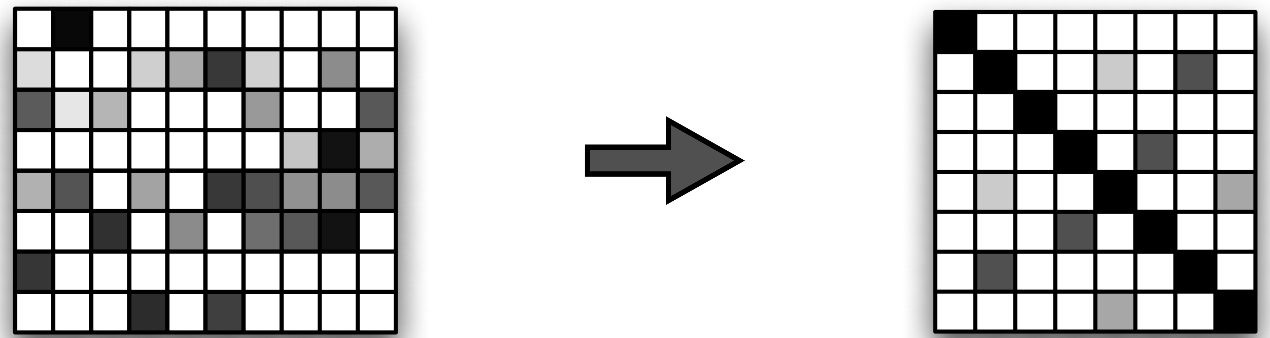
  - Friends: intersection

$$\mathrm{sim}(A, B) = |A \cap B|$$

  - Brands: Jaccard similarity

$$\mathrm{sim}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Computing venue similarity



$$\mathbf{X} \in \mathbb{R}^{n \times d}$$

each entry is the log(# of checkins at place i by user j)

one row for every 30m venues...

$$\mathbf{K} \in \mathbb{R}^{n \times n}$$

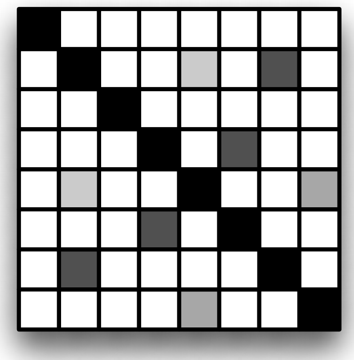
$$K_{ij} = \mathrm{sim}(\mathbf{x}_i, \mathbf{x}_j) = \frac{\mathbf{x}_i \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|}$$

Computing venue similarity

- Naive solution for computing $\mathbf{K}$:

$$O(n^2 d)$$

- Requires ~4.5m machines to compute in < 24 hours!!! and 3.6PB to store!



$$\mathbf{K} \in \mathbb{R}^{n \times n}$$

$$K_{ij} = \mathrm{sim}(\mathbf{x}_i, \mathbf{x}_j) = \frac{\mathbf{x}_i \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|}$$

Venue similarity w/ map reduce

key

user | visited venues — **map**

v_i, v_j score

v_i, v_j score

...

emit "all" pairs of visited venues
for each user

---

key

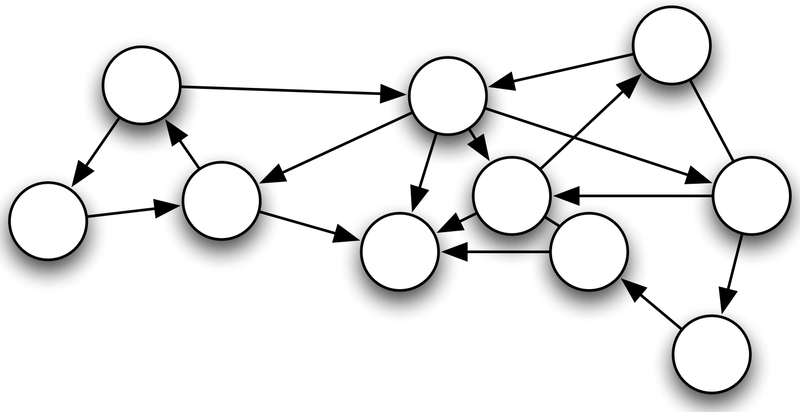
v_i, v_j score score ... score — **reduce**

final score

Sum up each user's score
contribution to this pair of venues

The Social Graph

- 15m person social network w/ lots of different interaction types:
  - friends
  - follows
  - dones
  - comments
  - colocation

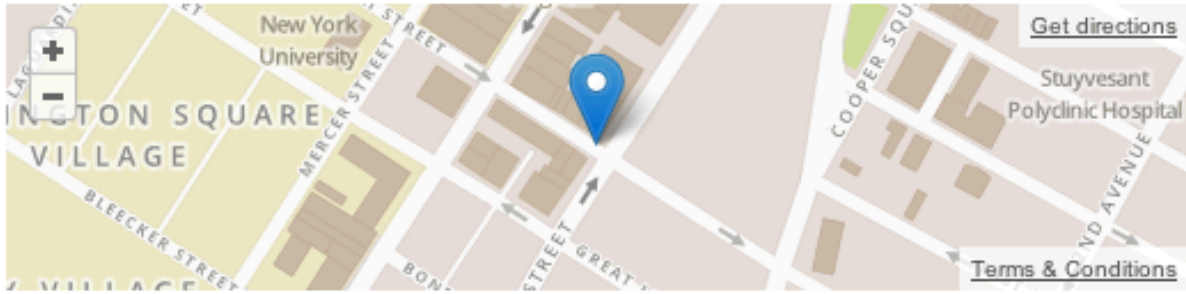


What happens when a new coffee shop opens in the East Village?

La Colombe Torrefaction

400 Lafayette St. (at E 4th St.), New York, NY 10003

Coffee Shop

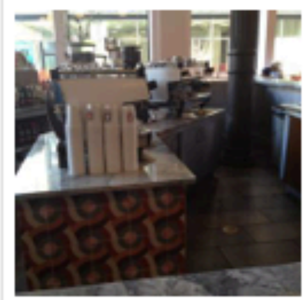


(800) 563-0860 @lacolombecoffee lacolombe.com

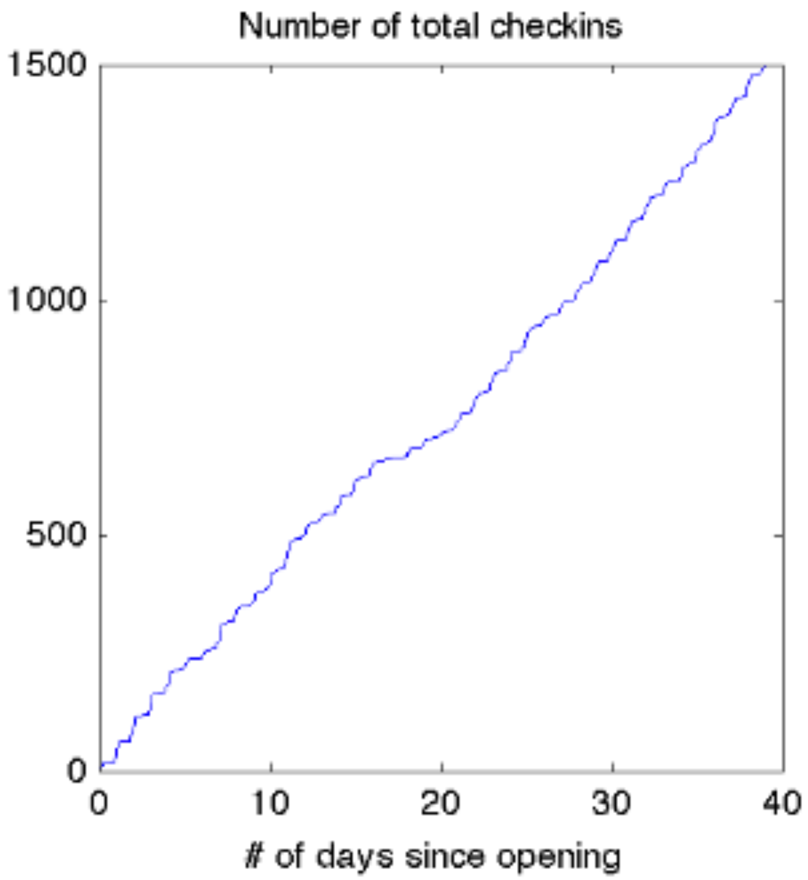
Taste always trumps novelty.

Photos

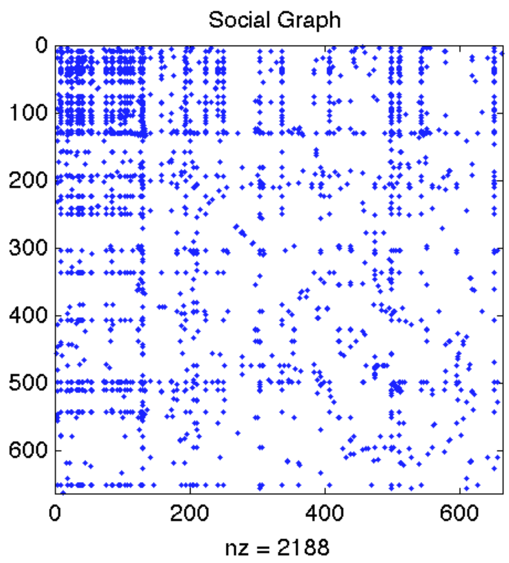
See all 79 photos

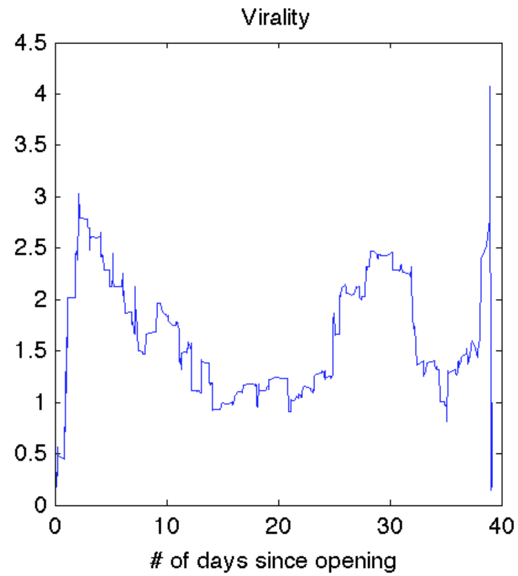


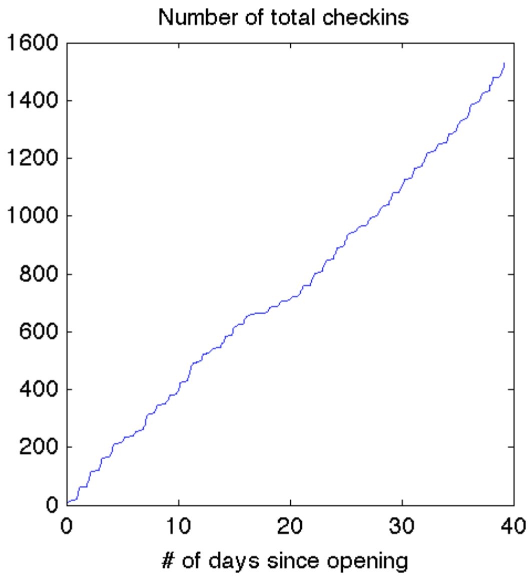
A new coffee shop opens...



The Social Graph

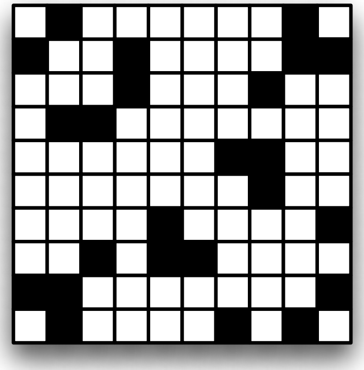




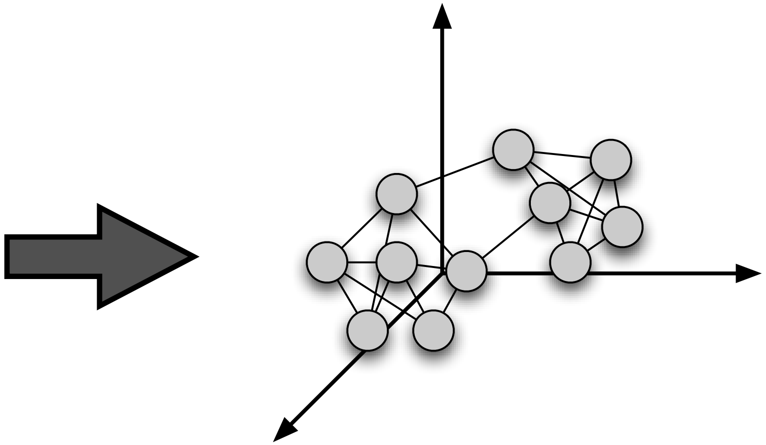


The Social Graph

How can we better visualize this network?



$\mathbf{A} \in \mathbb{B}^{n \times n}$



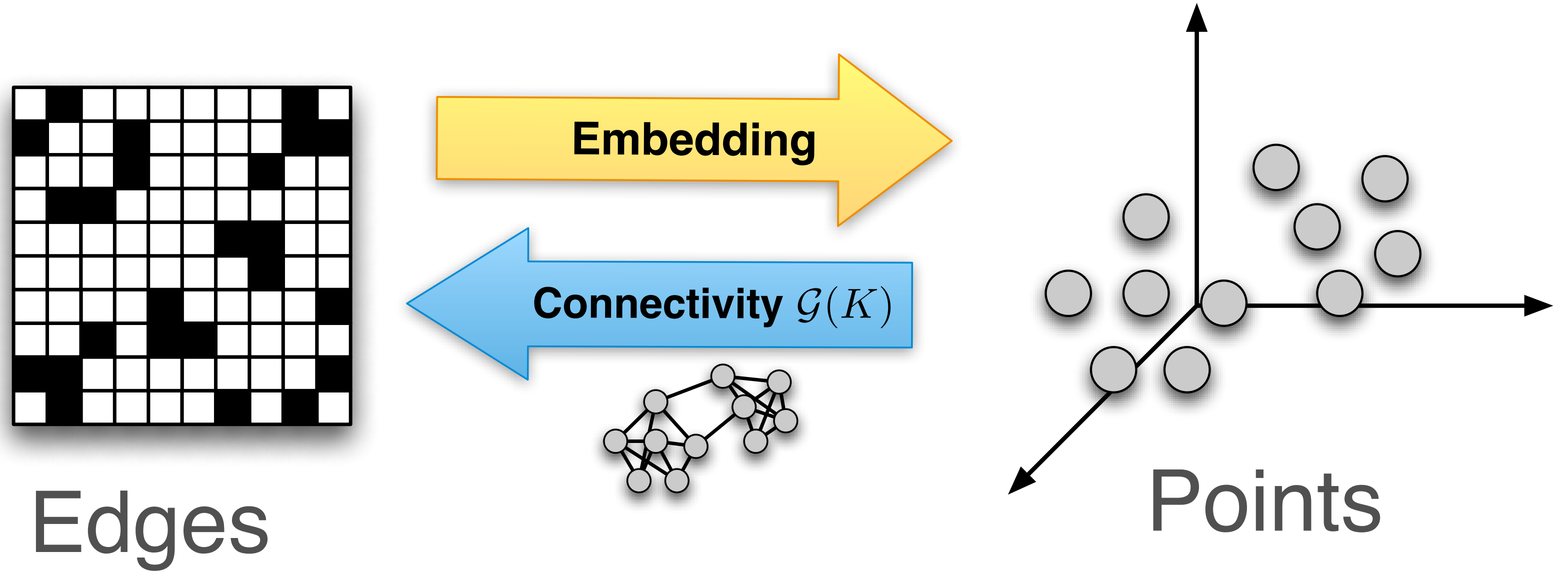
$\mathbf{L} \in \mathbb{R}^{n \times d}$

Graph embedding

- Spring Embedding - Simulate physical system by iterating Hooke's law
- Spectral Embedding - Decompose adjacency matrix $\mathbf{A}$ with an SVD and use eigenvectors with highest eigenvalues for coordinates
- Laplacian eigenmaps [Belkin, Niyogi '02] - form graph laplacian from adjacency matrix, $\mathcal{L} = \mathbf{D} - \mathbf{A}$, apply SVD to $\mathcal{L}$ and use eigenvectors with smallest non-zero eigenvalues for coordinates

Preserving structure

A connectivity algorithm $\mathcal{G}(\mathbf{K})$ such as k-nearest neighbors should be able to recover the edges from the coordinates such that $\mathcal{G}(\mathbf{K}) = \mathbf{A}$



Structure Preserving Embedding

[Shaw, Jebara '09]

- SDP to learn an embedding $\mathbf{K}$ from $\mathbf{A}$
  - Linear constraints on $\mathbf{K}$ preserve the global topology of the input graph
  - Convex objective favors low-rank $\mathbf{K}$ close to the spectral solution, ensuring low-dimensional embedding
- Use eigenvectors of $\mathbf{K}$ with largest eigenvalues as coordinates for each node

Structure Preserving Embedding

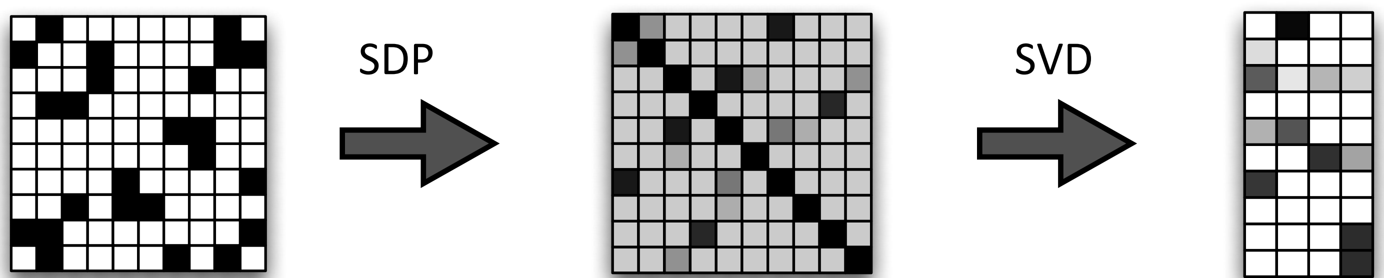
[Shaw, Jebara '09]

$$\max_{\mathbf{K}\in\mathcal{K}} \mathrm{tr}(\mathbf{KA})$$

$$D_{ij} > (1 - A_{ij}) \max_{m}(A_{im} D_{im}) \;\; \forall_{i,j}$$

$$\text{where } \mathcal{K} = \{\mathbf{K} \succeq 0,\ \mathrm{tr}(\mathbf{K}) \leq 1, \textstyle\sum_{ij} K_{ij} = 0\}$$

$$D_{ij} = K_{ii} + K_{jj} - 2K_{ij}$$

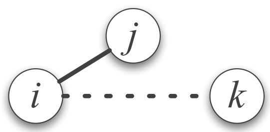


$\mathbf{A} \in \mathbb{B}^{n\times n}$ $\mathbf{K} \in \mathbb{R}^{n\times n}$ $\mathbf{L} \in \mathbb{R}^{n\times d}$

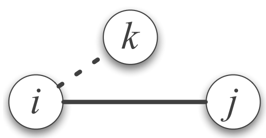
Large-scale SPE

[Shaw, Jebara '11]

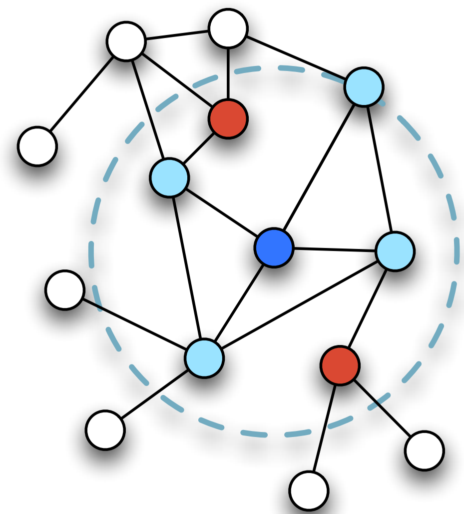
$$\mathrm{tr}(\mathbf{C}_l\mathbf{K}) = K_{jj} - 2K_{ij} + 2K_{ik} - K_{kk}$$



$\mathrm{tr}(\mathbf{C}_l\mathbf{K}) < 0$



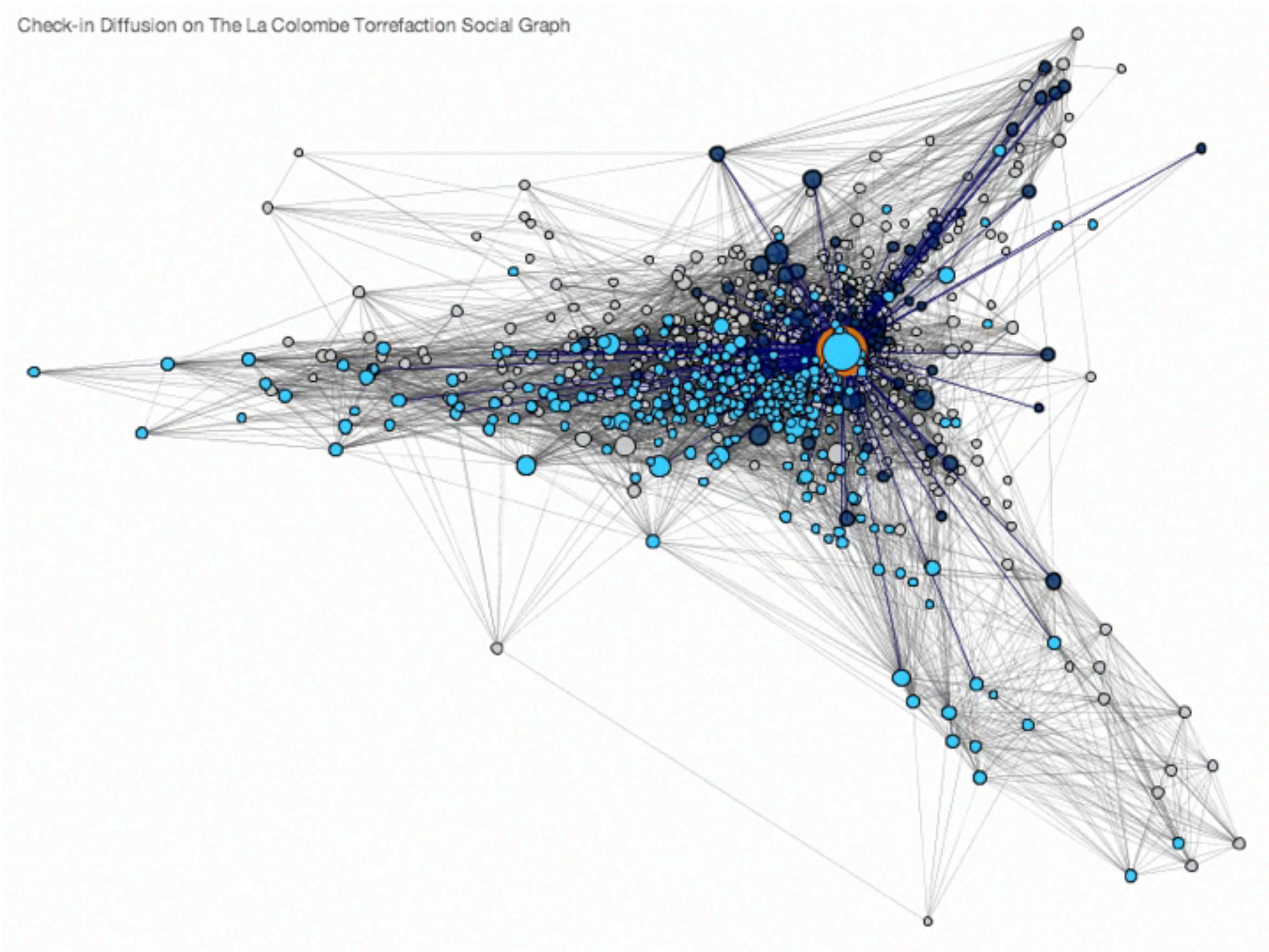
$\mathrm{tr}(\mathbf{C}_l\mathbf{K}) > 0$



$$f(\mathbf{L}) = \lambda \mathrm{tr}(\mathbf{L}^\top\mathbf{L}\mathbf{A}) - \sum_{l \in S} \max(\mathrm{tr}(\mathbf{C}_l\mathbf{L}^\top\mathbf{L}), 0)$$

$$\Delta(f(\mathbf{L}), \mathbf{C}_l) = \begin{cases} 2\mathbf{L}(\lambda\mathbf{A} - \mathbf{C}_l) \text{ if } \mathrm{tr}(\mathbf{C}_l\mathbf{L}^\top\mathbf{L}) > 0 \\ 0 \quad \text{otherwise} \end{cases}$$

$$\mathbf{L}_{t+1} = \mathbf{L}_t + \eta\Delta(f(\mathbf{L}_t), \mathbf{C}_l)$$



Video: http://vimeo.com/39656540

Notes on next slide

Notes for previous slide:
Each node in this network is a person, each edge represents friendship on foursquare. The size of each node is proportional to how many friends that person has. We can see the existence of dense clusters of users, on the right, the top, and on the left. There is a large component in the middle. There are clear hubs.

We can now use this low-dimensional representation of this high-dimensional network, to better track what happens when a new coffee shop opens in the east village.

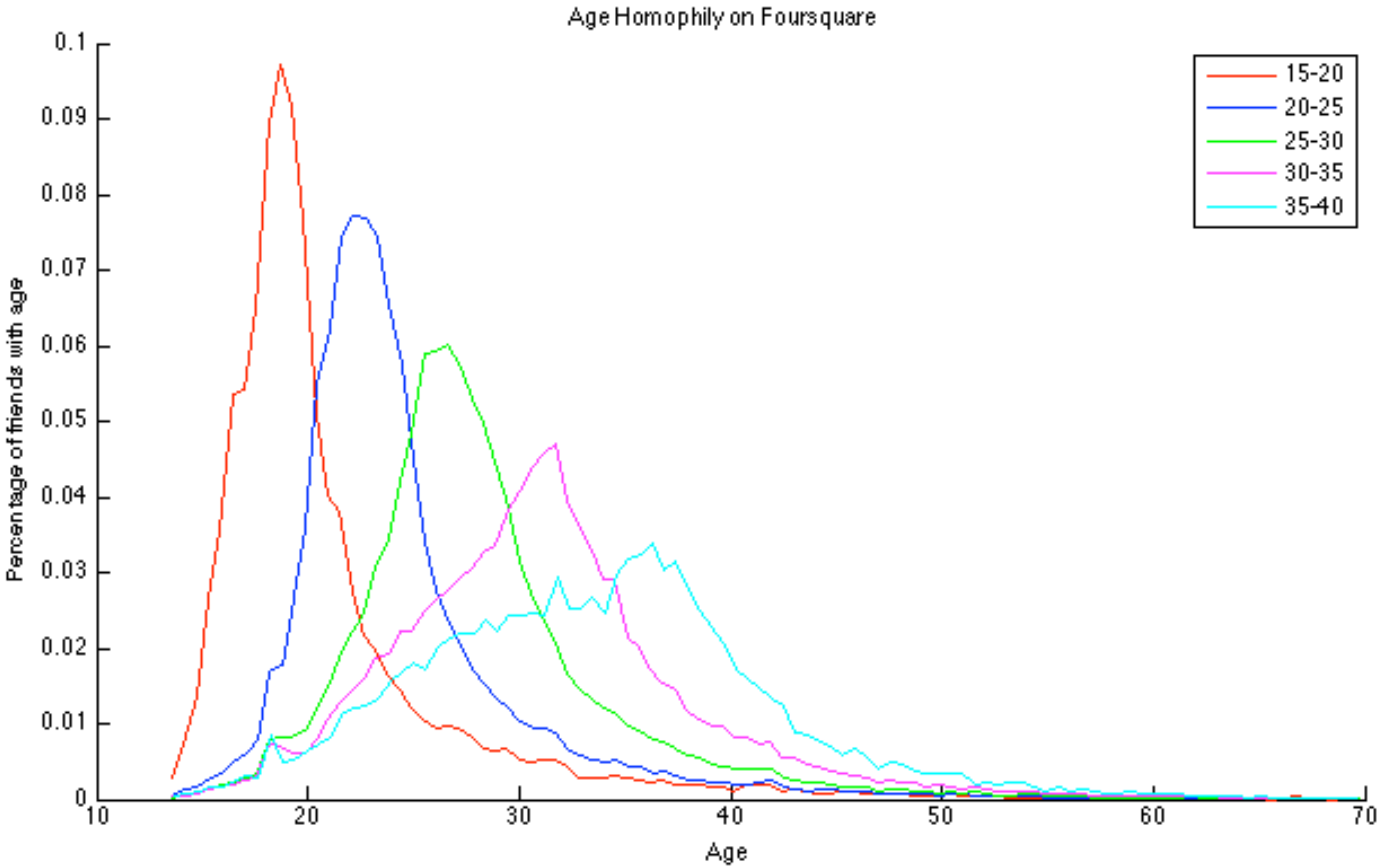
As expected, it spreads ...like a virus, across this social substrate. We see as each person checks in to la colombe, their friends light up. People who have discovered the place are shown in blue. The current checkin is highlighted in orange in orange.

It’s amazing to see how la colombe spreads. Many people have been talking about how ideas, tweets, and memes spread across the internet. For the first time we can track how new places opening in the real world spread in a similar way.

The Social Graph

- What does this low-dimensional structure mean?
- Homophily
  - Location, Demographics, etc.

The Social Graph



foursquare®

Influence on foursquare

- Tip network
  - sample of 2.5m people “doing” tips from other people and brands
  - avg. path length 5.15, diameter 22.3
- How can find the authoritative people in this network?

Measuring influence w/ PageRank

[Page et al '99]

- Iterative approach
  - start with random values and iterate
  - works great w/ map-reduce

$$\mathrm{PR}(i) = (1-d) + d \sum_{j \in \{A_{ij}=1\}} \frac{\mathrm{PR}(j)}{\sum_k A_{ik}}$$

Measuring influence w/ PageRank

[Page et al ’99]

- Equivalent to finding the principal eigenvector of the normalized adj. matrix

$$\mathbf{A} \in \mathbb{B}^{n \times n} \qquad P_{ij} = \frac{A_{ij}}{\sum_j A_{ij}}$$

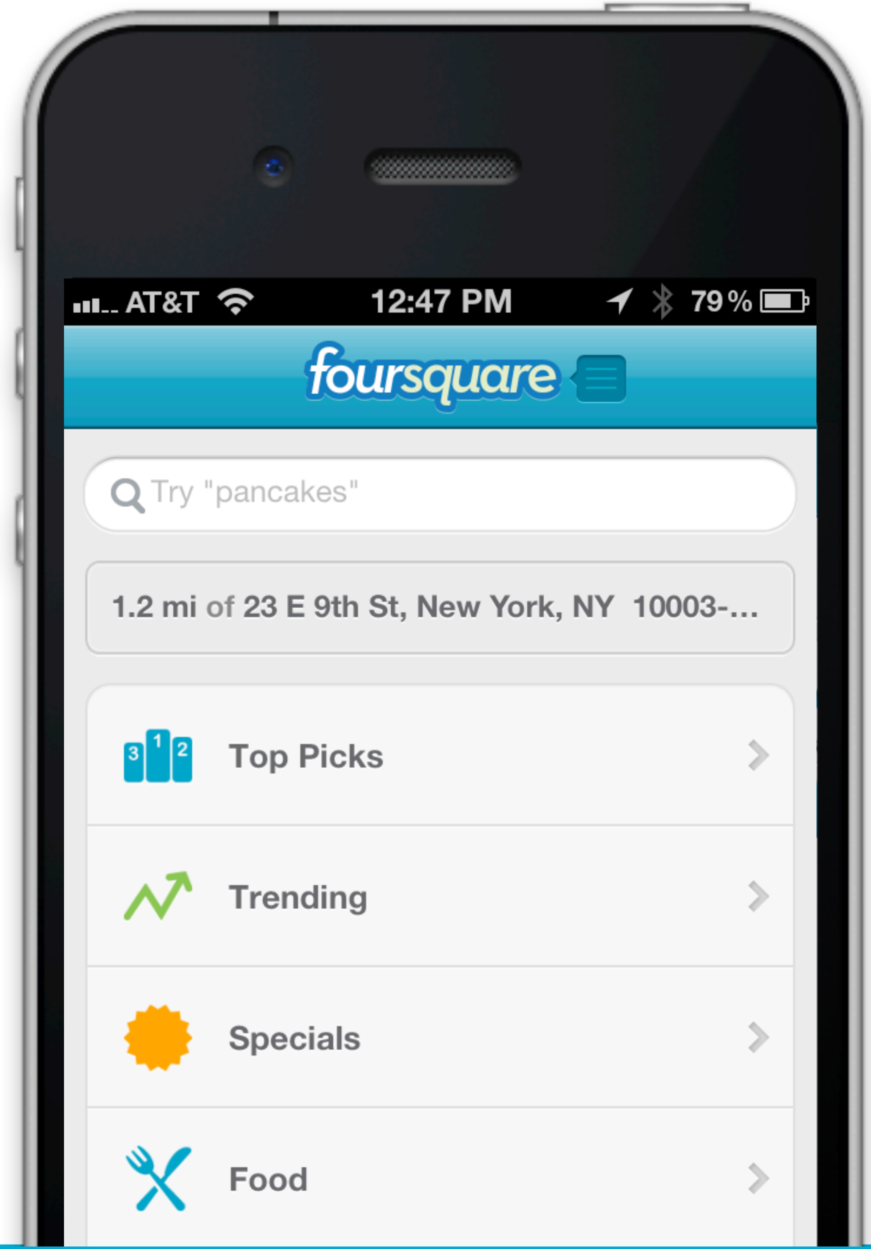
$$\mathrm{PR}(i) \propto v_i \text{ where } \mathbf{P}\mathbf{v} = \lambda_1 \mathbf{v}$$

Influence on foursquare

- Most influential brands:
  - History Channel, Bravo TV, National Post, Eater.com, MTV, Ask Men, WSJ, Zagat, NY Magazine, visitPA, Thrillist, Louis Vuitton
- Most influential users
  - Lockhart S, Jeremy B, Naveen S

Explore

A social recommendation engine built from check-in data



Foursquare Explore

- Realtime recommendations from signals:
  - location
  - time of day
  - check-in history
  - friends preferences
  - venue similarities

Putting it all together

Nearby relevant venues

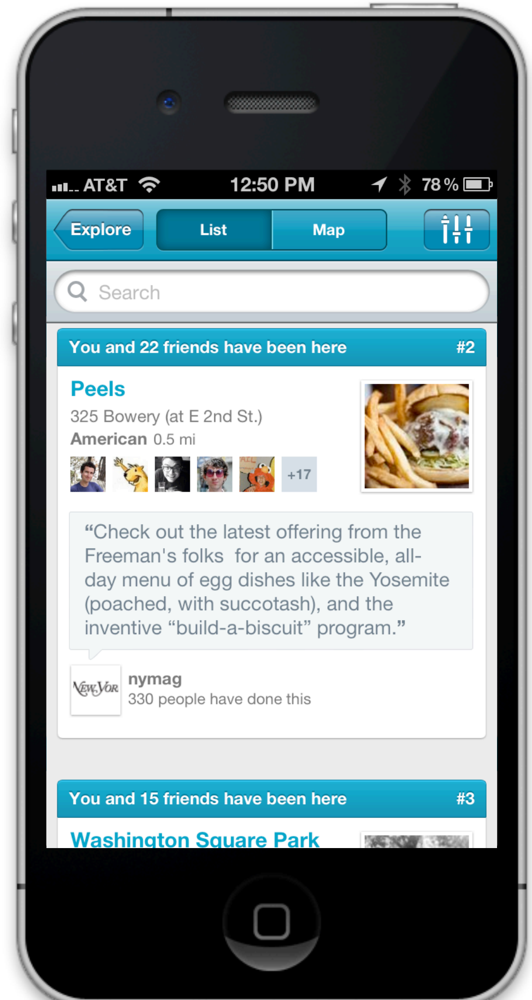
Friends' check-in history, similarity

Similar Venues

User's check-in history

MOAR Signals

< 200 ms



foursquare®

Our data stack

- MongoDB
- Amazon S3, Elastic Mapreduce
- Hadoop
- Hive
- Flume
- R and Matlab

Open questions

- What are the underlying properties and dynamics of these networks?
- How can we predict new connections?
- How do we measure influence?
- Can we infer real-world social networks?

Conclusion

- Unique networks formed by people interacting with each other and with places in the real world
- Massive scale -- today we are working with millions of people and places here at foursquare, but there are over a billion devices in the world constantly emitting this signal of userid, lat, long, timestamp

Join us!

foursquare is hiring!
110+ people and growing

foursquare.com/jobs

Blake Shaw
@metablake
blake@foursquare.com