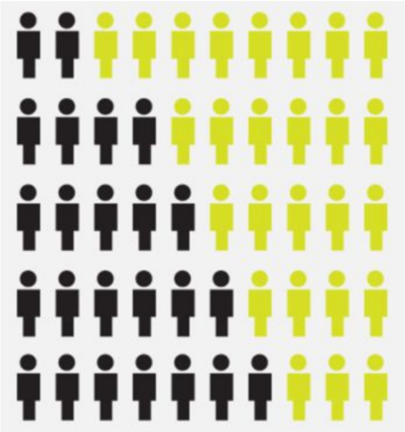
Regions, Periods, Activities: Uncovering Urban Dynamics via Cross-Modal Representation Learning

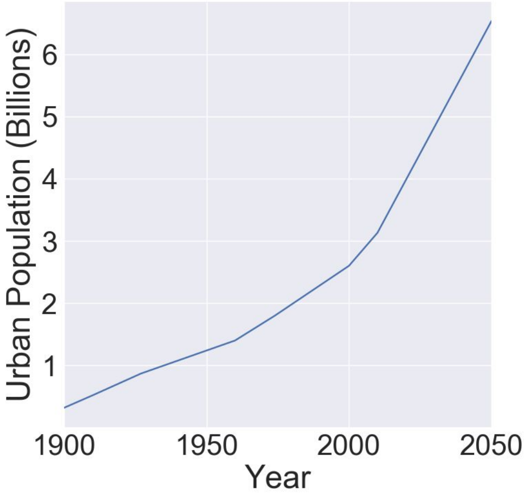
Chao Zhang
University of Illinois at Urbana-Champaign
czhang82@illinois.edu

with Keyang Zhang, Quan Yuan, Haoruo Peng, Yu Zheng, Tim Hanratty, Shaowen Wang, and Jiawei Han

Small World, Big Cities, Big Challenges

- The ever-increasing urbanization process:
  - 1900: 20% people live in cities
  - 1910: 40% people live in cities
  - 2010: 50% people live in cities
  - 2030: 60% people live in cities
  - 2050: 70% people live in cities





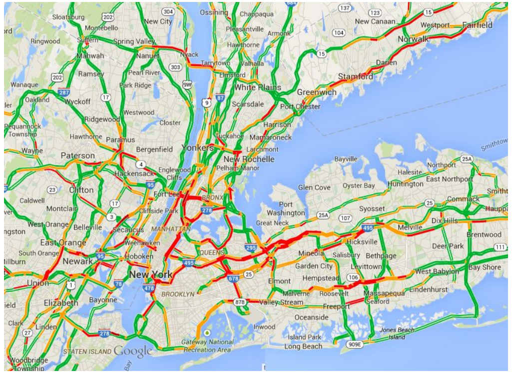
City	**Shanghai**	**Tokyo**	**New York**	**London**
Population	24.3M	13.5M	8.4M	8.7M

Small World, Big Cities, Big Challenges

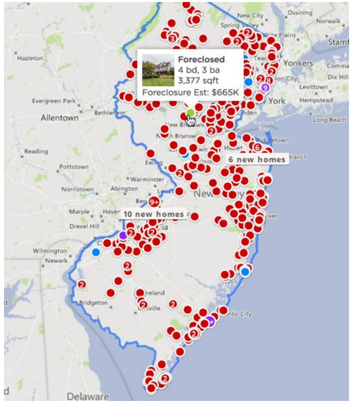
Life is not that easy in big cities when we try:

Finding a place

Avoiding traffic jams



Buying a house



Modeling people's activities in the city is key to tackling these challenges!

Why Geo-Tagged Social Media (GTSM)?

An invaluable resource as a result of ***human sensing.***

Millions of human sensors probe the city and return semantics-rich data.

Problem Description

Input: a collection of GTSM records

- Each record has: a location, a timestamp, a bag of keywords.

Task: model people's typical activities in different regions and periods

- Multiple retrieval schemas:
  - region + time -> keywords
  - region + keyword -> time
  - time + keyword -> region

"What are happening around my hotel now?"

"Where should I go to hang out with my friends at 9pm?"

Challenges

- Data Sparsity

  GTSM messages are very short

  Many regions have few records

- Data Variation

  Spatiotemporal variation: people may share the same activity with slightly different locations and timestamps.

  Textual variation: people often describe the same activity with different keywords.

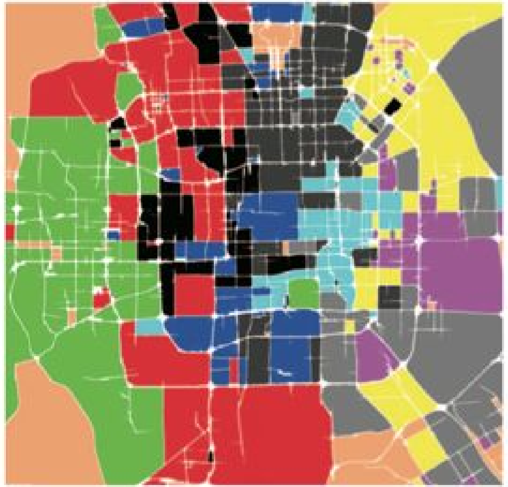
  E.g., consider two Lakers' fans watching the NBA game at Staples center.

Previous Works

Urban function discovery

- Find regions of similar functions in the city

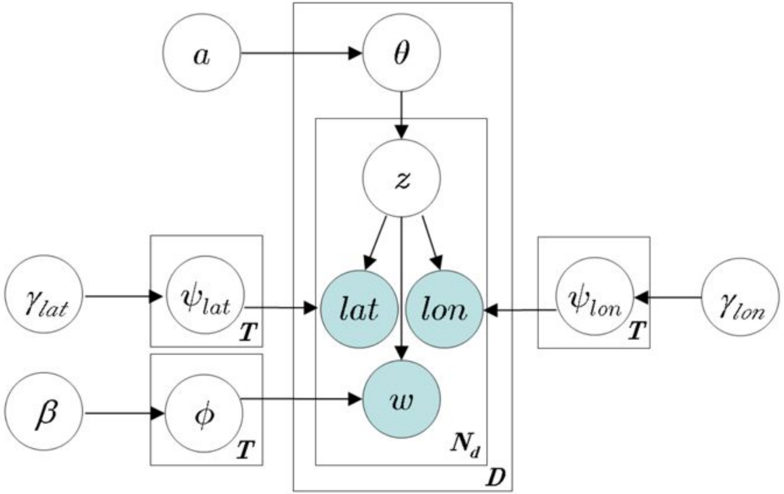
(Cranshaw et. al. AAAI 2010, Yuan et. al. KDD 2012)



Geographical topic discovery

- Incorporate location into topic models

(Sizov et al. WSDM 2010, Yin et al. WWW 2011)

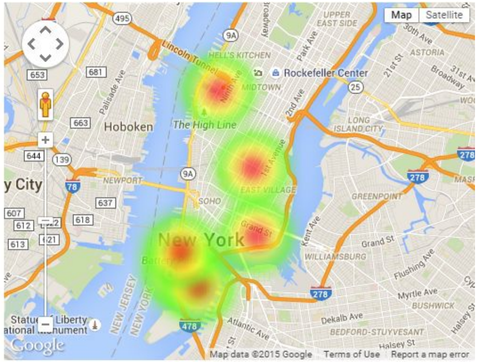


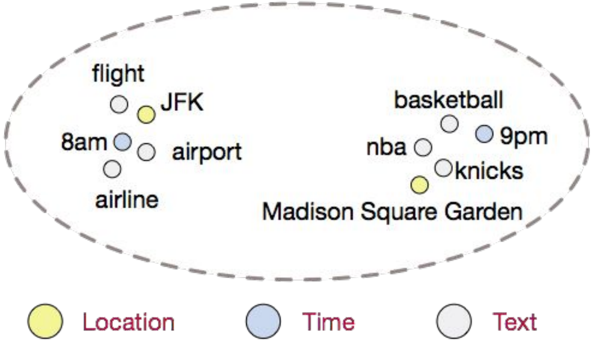
An Overview of Our Method: CrossMap

Hotspot Detection → **Joint Embedding**

Detect regions and periods where people's activities burst

Map regions, periods, and keywords into the same space



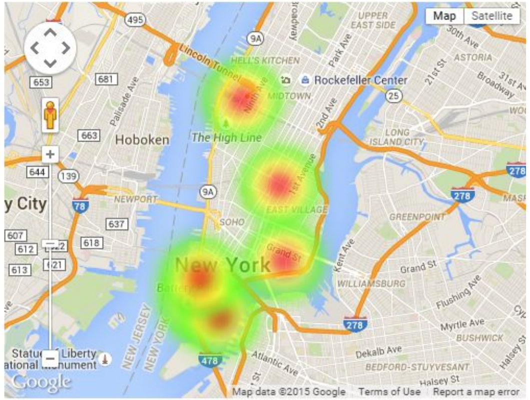


Hotspot Detection: A Mode-Seeking Procedure

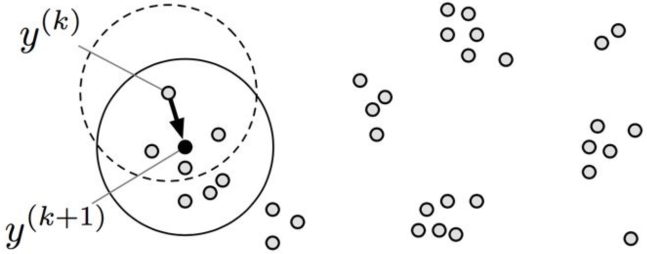
- A spatial (temporal) hotspot is a density maxima in the 2D (1D) space
- We design a fast mode seeking procedure to find the hotspots.

Benefits:

- Fast
- No distributional assumptions



Kernel density: $f(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)$



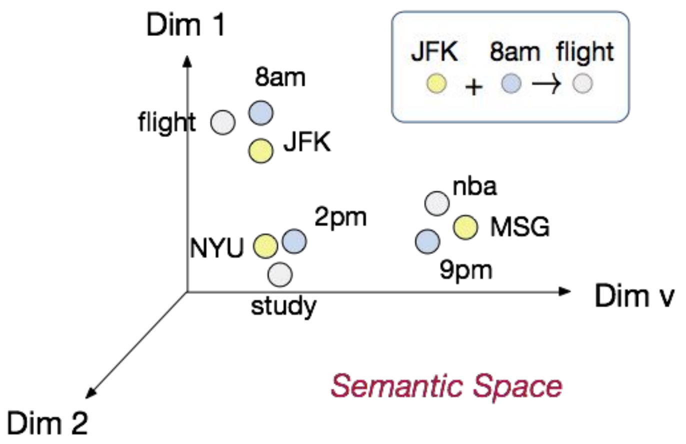
Cross-Modal Embedding: Designing Philosophy

Map regions, periods, and keywords into the same space:

- **Region**: a spatial hotspot
- **Period**: a temporal hotspot

Aim to preserve two types of correlations:

1. **Co-occurrence**: two units are correlated if there co-occur frequently
2. **Neighborhood**: two regions (periods) are correlated if they are adjacent



Cross-Modal Embedding: Two Strategies

Reconstruction-based embedding

1. Consider each record as a relation
2. Mark off one unit i and try to predict it from the observed units

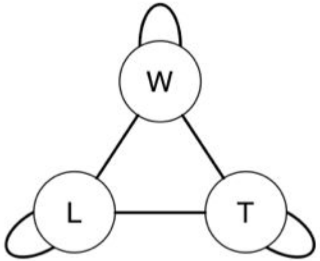
$$p(i|r_{-i}) = \exp(s(i, r_{-i})) / \sum_{j \in X} \exp(s(j, r_{-i}))$$

Overall objective function:

$$O = -\sum_{r \in C} \sum_{i \in r} \log p(i|r_{-i})$$

Graph-based embedding

1. Use a graph to encode the correlations between regions, periods, and activities
2. Learn graph node embeddings to preserve the graph structure



$$O = O_{WW} + O_{LL} + O_{TT} + O_{WL} + O_{WT} + O_{LT}$$

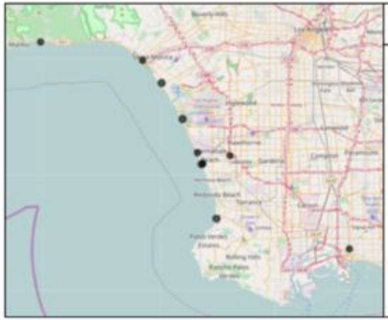
$$O_{XY} = \sum_{i \in X} d_i \mathrm{KL}(p'(\cdot|i)||p(\cdot|i)) + \sum_{j \in Y} d_j \mathrm{KL}(p'(\cdot|j)||p(\cdot|j))$$

Experimental Settings

- Data sets
  - LA: ~1 million geo-tagged tweets in Los Angeles
  - NY: ~0.6 million Foursquare checkins in New York City
- Compared methods
  - LGTA [1]
  - MGTM [2]
  - TF-IDF
  - SVD
  - Tensor Factorization

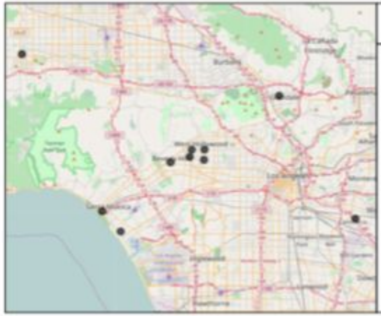
[1] Geographical topic discovery and comparison. 2011 WWW.
[2] Detecting non-gaussian geographical topics in tagged photo collections. 2014 WSDM.

Illustrative Cases: Cross-Modal Retrieval



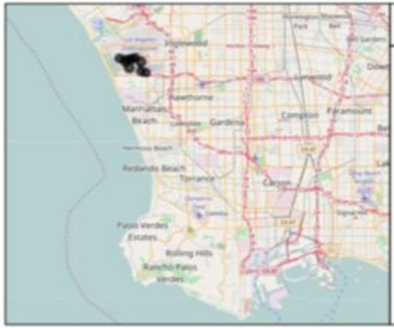
Text	Time
beach	19
beachday	18
beachlife	17
surfing	16
sand	20
boardwalk	14
pacificocean	15
longbeach	13
redondobeach	11
dockweiler	12

Query: “Beach”



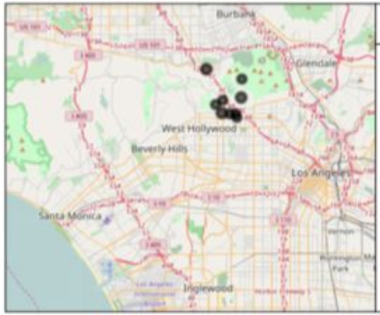
Text	Time
shopping	15
nordstrom	16
mall	14
jambajuice	17
grocery	13
blackfriday	18
sephora	12
ulta	19
michaelkor	20
kmart	21

Query: “Shopping”



Text	Time
airport	7
tsa	10
airline	8
lax	6
southwester	11
americanair	9
delay	5
terminal	12
jfk	16
sfo	14

Query: “33.9424, -118.4137” (LAX Airport)



Text	Time
hollywood	20
photo	21
touring	0
hollywoodhills	23
walkoffame	22
nights	19
kids	13
halloween	1
marilymonroe	16
parishilton	18

Query = “34.0928, -118.3287” (Hollywood)

Quantitative Evaluation: Attribute Recovery

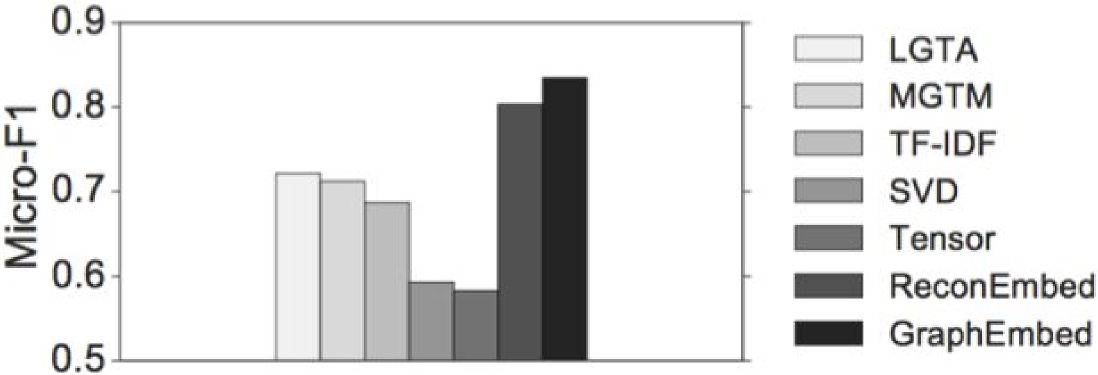
Mark off one attribute (location, time, or text) and predict it based on the observed ones.

Mean reciprocal ranks:

	Text		Location		Time	
Method	Tweet	4SQ	Tweet	4SQ	Tweet	4SQ
LGTA	0.376	0.6107	0.3792	0.6083	-	-
MGTM	0.3874	0.5974	0.4474	0.5753	-	-
TF-IDF	0.62	0.8505	0.4298	0.7097	0.3197	0.3431
SVD	0.4475	0.7137	0.3953	0.646	0.3256	0.3187
Tensor	0.4382	0.6826	0.3871	0.6251	0.3179	0.2983
RECON	0.6877	0.9219	0.6526	0.9044	0.3582	0.3612
GRAPH	**0.7011**	**0.9449**	**0.6758**	**0.9168**	**0.3895**	**0.3716**

Application: Activity Classification

Foursquare checkins belong to nine categories. We predict the category based on the embeddings.



Classification F1 Score.

Embedding Visualization

(a) LGTA (b) CrossMap

Visualizing the feature vectors generated by LGTA and CrossMap for three activity categories: “Food” (cyan), “Travel & Transport” (blue), and “Residence” (orange).

Summary

- A cross-modal embedding framework for modeling urban dynamics from geo-tagged social media.
    - Perform much better than baselines in attribute prediction and downstream applications
- Interesting future directions
    - Can be easily extended to learn user embeddings for user-level applications
    - Using deep architecture for embedding could further improve the performance at the time cost

Thanks!