Learning Urban Community Structures: A Collective Embedding Perspective with Periodic Spatial-temporal Mobility Graphs

Pengyang Wang, Yanjie Fu, Jiawei Zhang,

Xiaolin Li, Dan Lin

Outline

- **Background and Motivation**
- Definition and Problem Statement
- Methodology
- Application
- Evaluation
- Conclusion

Background and Motivation

- Urban life is getting more diverse and vibrant

Urban community

Why we study urban communities?

- **Spatial Imbalance**

  ----vibrancy differences between communities

Challenges & Insights

- **Challenge I – Graph construction**

How to unify and represent the POIs and human periodic mobility records as a set of mobility graphs?

- **Insight I**

a set of periodic spatial-temporal mobility graphs

Challenges & Insights

- **Challenge II – Collective embedding**

How to collectively learn the embeddings of POIs from multiple periodic mobility graphs?

- **Insight II**

Collective deep auto-encoder

Challenges & Insights

- ## Challenge III - Embedding aggregation

How to align and aggregate POI embeddings for community structure representation learning?

- ## Insight III

unsupervised graph-based weighting method

Outline

- Background and Motivation
- **Definition and Problem Statement**
- Methodology
- Application
- Evaluation
- Conclusion

Definition I

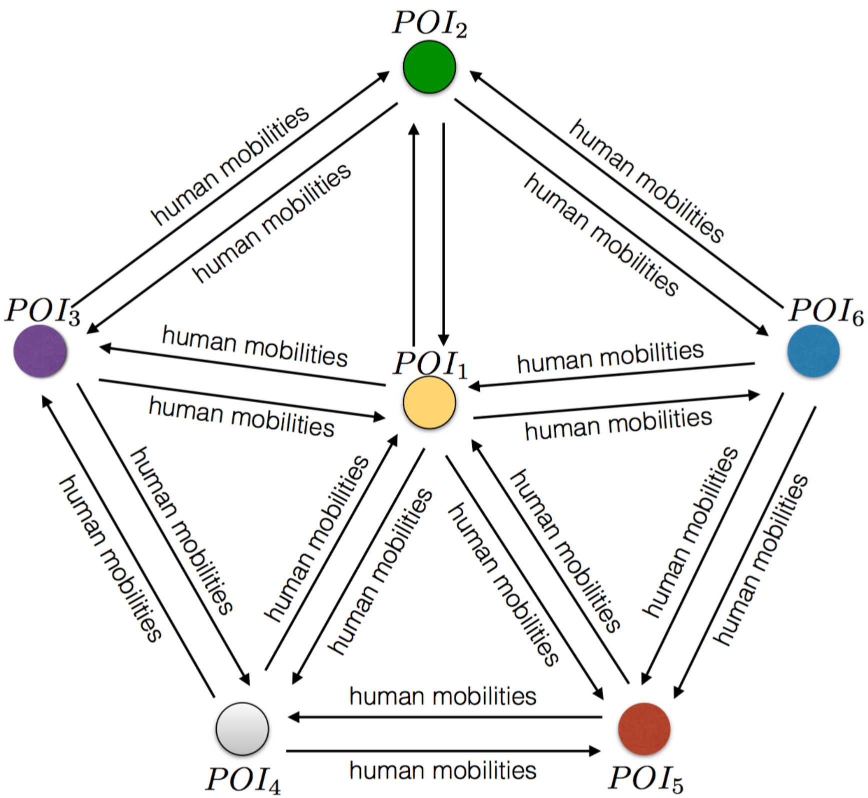
- Urban communities

residential complex

neighborhood area

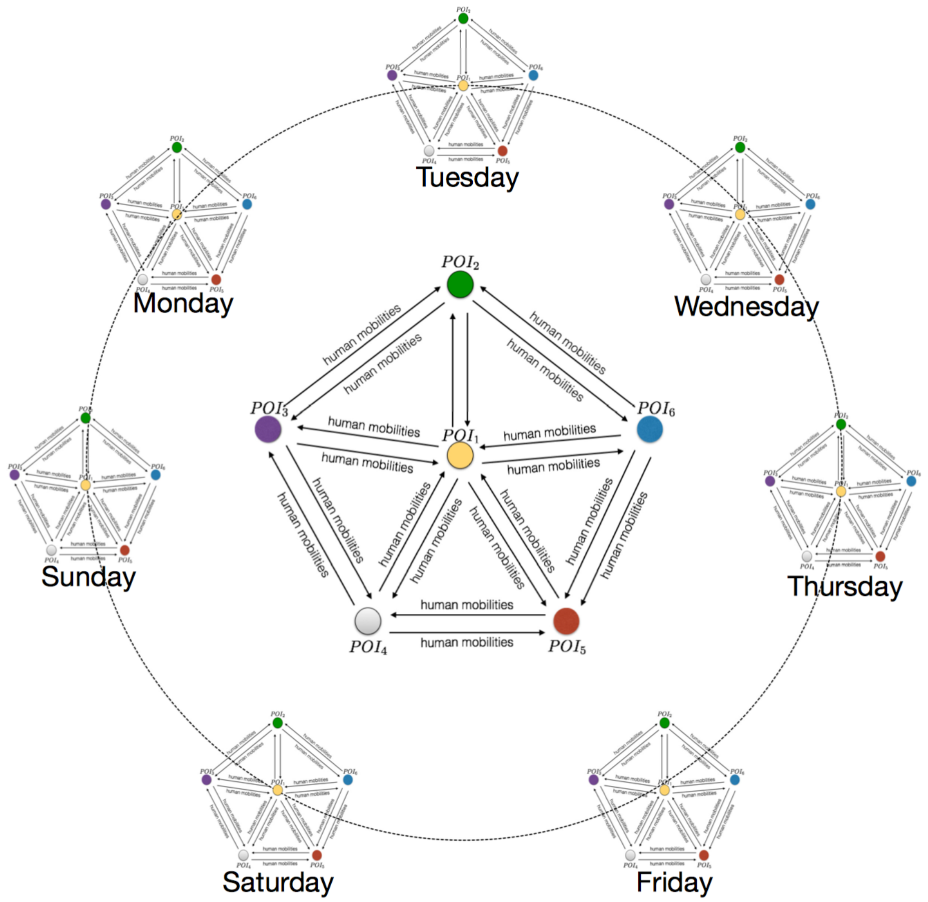
Definition II

- Mobility Graph



Definition III

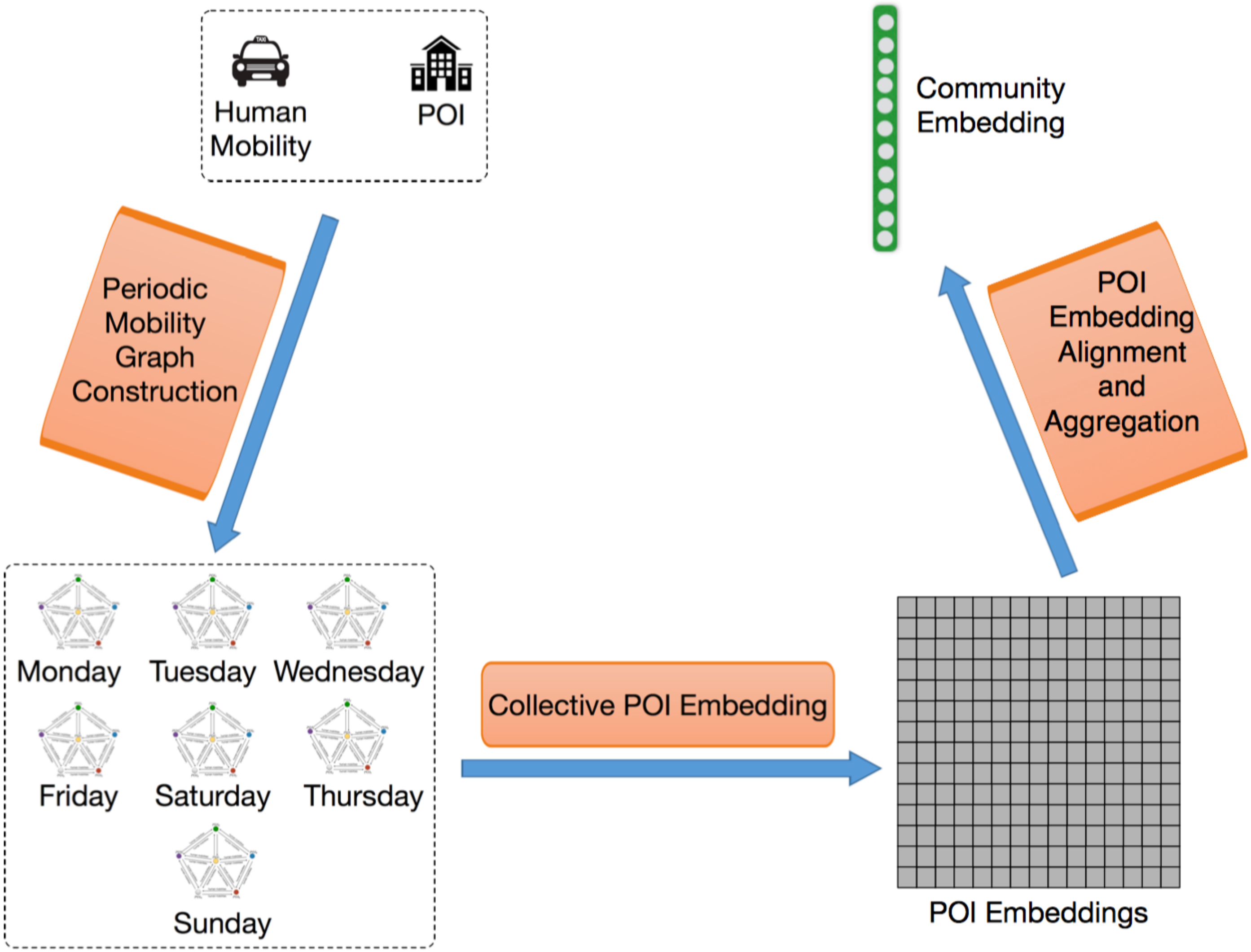
- Periodic Mobility Graphs



Problem Statement

- **Given**
  - Residential communities (locations, POIs)
  - Human mobility (e.g., taxi GPS traces)
- **Objective**
  - Learning representations about static spatial configurations
  - Learning representations about dynamic human mobility connectivity of POIs in the community
- **Core tasks**
  - Construction of the periodic mobility graph set for a community
  - Collectively embedding
  - Aggregating and aligning POI embedding into community embedding.

Framework Overview



Outline

- Background and Motivation
- Problem Statement
- **Methodology**
- Application
- Evaluation
- Conclusion

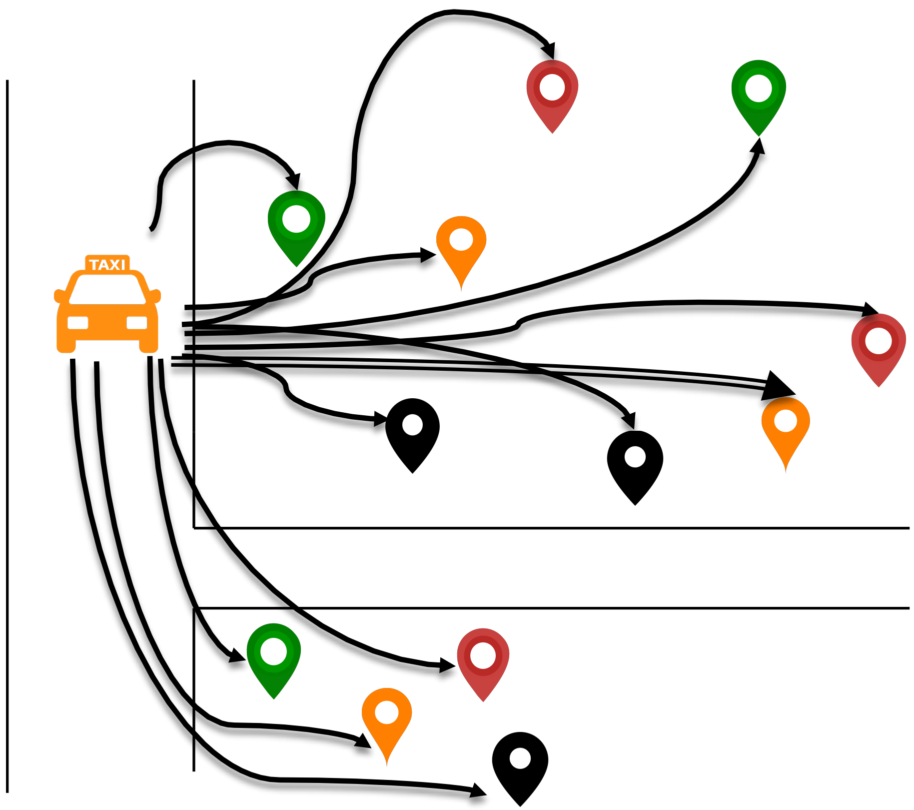
Methodology

- Periodic Mobility Graph Construction
- Collective POI Embedding
- Aligning and Aggregating POI Embeddings to Community Embeddings

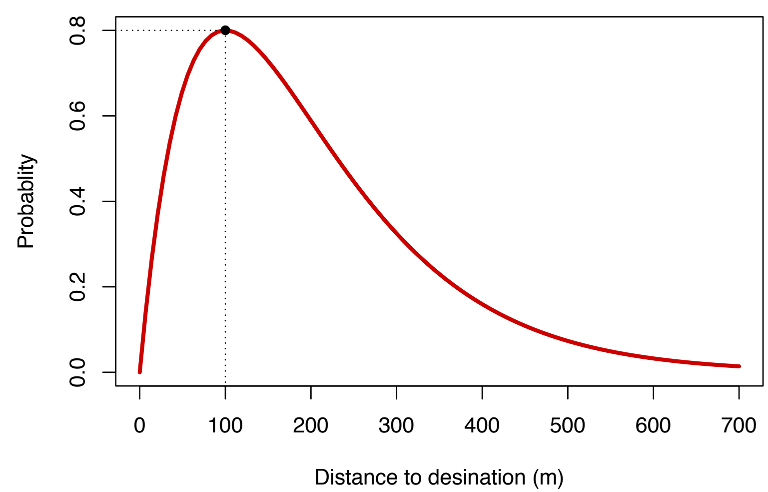
Periodic Mobility Graph Construction

Propagate visit probability

$$P(x) = \frac{\beta_1}{\beta_2} \cdot x \cdot exp(1 - \frac{x}{\beta_2}),$$
$$\beta_1 = \max_x P(x) \text{ and } \beta_2 = \arg\max_x P(x)$$

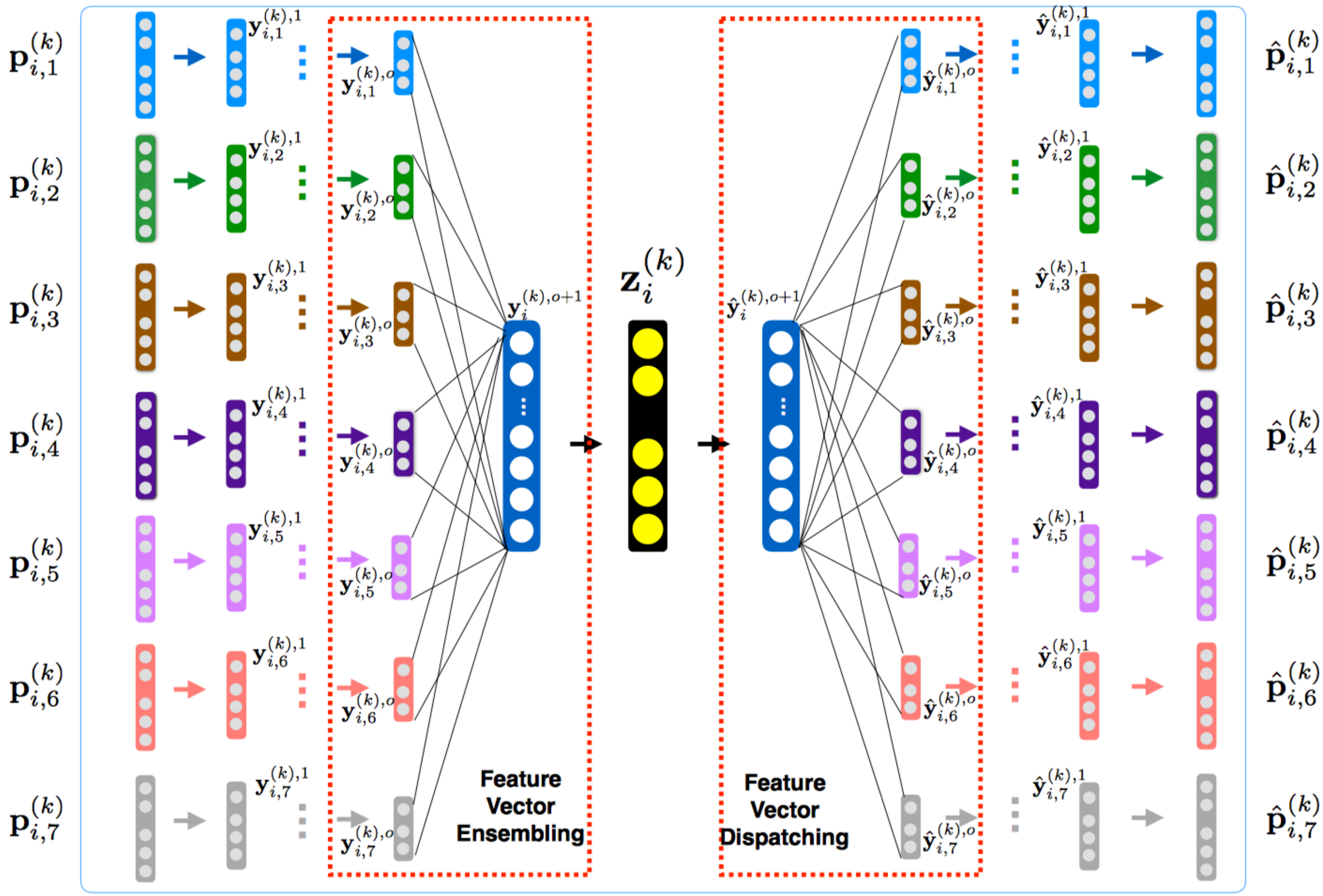


the closer, the e likely to visit?



Probability distribution w.r.t $\beta_1 = 0.8, \beta_2 = 100.$

Collective POI Embedding



Collective POI Embedding

Encoder

$$\begin{cases} \mathbf{y}_{i,t}^{(k),1} & = \sigma(\mathbf{W}_{i,t}^{(k),1}\mathbf{p}_{i,t}^{(k)} + \mathbf{b}_{i,t}^{(k),1}), \forall t \in \{1,2,\cdots,7\}, \\ \mathbf{y}_{i,t}^{(k),r} & = \sigma(\mathbf{W}_{i,t}^{(k),r}\mathbf{p}_{i,t}^{(k)} + \mathbf{b}_{i,t}^{(k),r}), \forall r \in \{2,3,\cdots,o\}, \\ \mathbf{y}_{i}^{(k),o+1} & = \sigma(\sum_t \mathbf{W}_{t}^{(k),o+1}\mathbf{y}_{i,t}^{(k),o} + \mathbf{b}_{t}^{(k),o+1}), \\ \mathbf{z}_{i}^{(k)} & = \sigma(\mathbf{W}^{(k),o+2}\mathbf{y}_{i}^{(k),o+1} + \mathbf{b}^{(k),o+2}), \end{cases}$$

Decoder

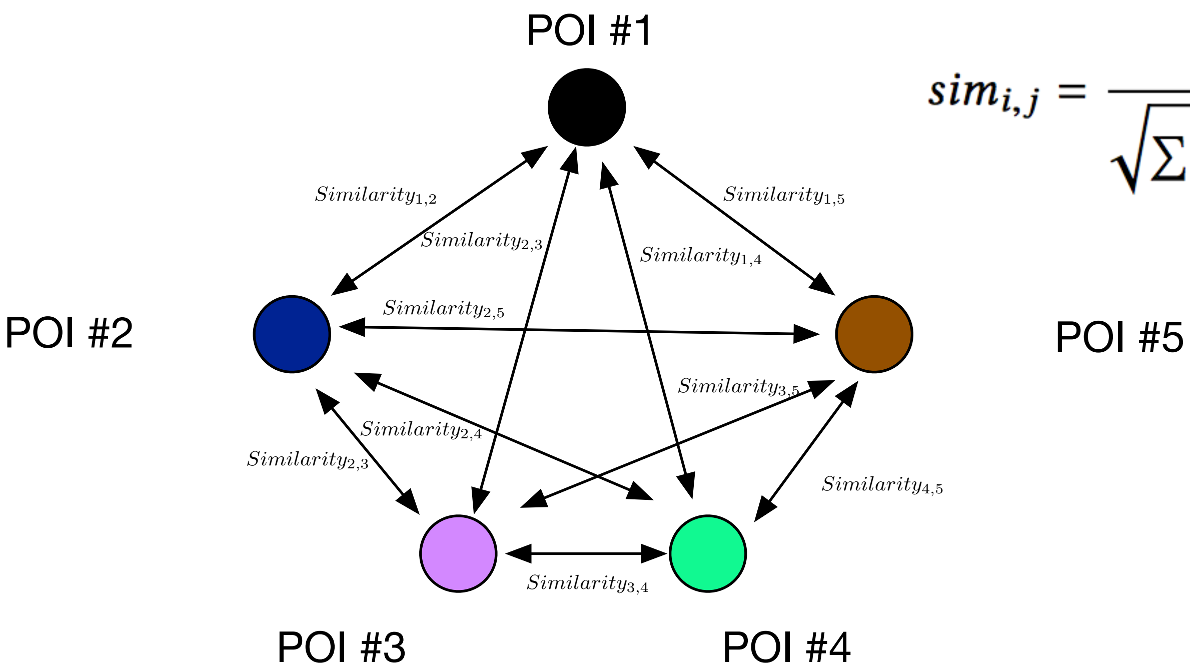
$$\begin{cases} \hat{\mathbf{y}}_{i}^{(k),o+1} & = \sigma(\hat{\mathbf{W}}^{(k),o+2}\mathbf{z}_{i}^{(k)} + \hat{\mathbf{b}}^{(k),o+2}), \\ \hat{\mathbf{y}}_{i,t}^{(k),o} & = \sigma(\hat{\mathbf{W}}_{t}^{(k),o+1}\hat{\mathbf{y}}_{i}^{(k),o+1} + \hat{\mathbf{b}}_{t}^{(k),o+1}), \\ \hat{\mathbf{y}}_{i,t}^{(k),r-1} & = \sigma(\hat{\mathbf{W}}_{i,t}^{(k),r}\hat{\mathbf{y}}_{i,t}^{(k),r} + \hat{\mathbf{b}}_{i,t}^{(k),r}), \forall r \in \{2,3,\cdots,o\}, \\ \hat{\mathbf{p}}_{i,t}^{(k)} & = \sigma(\hat{\mathbf{W}}_{i,t}^{(k),1}\hat{\mathbf{y}}_{i,t}^{(k),1} + \hat{\mathbf{b}}_{i,t}^{(k),1}), \end{cases}$$

Loss Function:

$$\mathcal{L}^{(k)} = \sum_{t \in \{1,2,\ldots,7\}} \sum_{i} \|(\mathbf{p}_{i,t}^{(k)} - \hat{\mathbf{p}}_{i,t}^{(k)}) \odot \mathbf{v}_{i,t}^{(k)}\|_2^2$$

Aligning and Aggregating POI Embeddings to Community Embeddings

- Graph based weighting method



$$sim_{i,j} = \frac{\sum_l \tilde{\mathcal{G}}^{(k)}[i,l] \times \tilde{\mathcal{G}}^{(k)}[j,l]}{\sqrt{\sum_l \tilde{\mathcal{G}}^{(k)}[i,l]^2} \times \sqrt{\sum_l \tilde{\mathcal{G}}^{(k)}[j,l]^2}}$$

POI similarity graph

Graph based weighting method

- **Weight Calculation**

$$w_l^{(k)} = \frac{\sum_{i \in c_k} \sum_{j \in c_k} sim_{i,j} \times |\tilde{\mathcal{G}}^{(k)}[i,l] - \tilde{\mathcal{G}}^{(k)}[j,l]|}{M}$$

if the l-th dimension of the latent feature makes more sense, when POI p_i and p_j are very similar, the difference of p_i and p_j on the l-th dimension $|\tilde{\mathcal{G}}^{(k)}[i,l] - \tilde{\mathcal{G}}^{(k)}[j,l]|$ should be very small. Therefore, if the l-th dimension of the latent feature does not make much sense, $|g[i,l] - g[j,l]|$will increase; if p_i and p_j are very similar, $Sim_{i,j}$will further penalize $|g[i,l] - g[j,l]|$

$$\hat{\mathcal{G}}^{(k)}[s,l] = \sum_{p_i \in \Phi_s} \tilde{\mathcal{G}}^{(k)}[i,l] \times w_l^{(k)}$$

Outline

- Background and Motivation
- Definition and Problem Statement
- Methodology
- **Application**
- Evaluation
- Conclusion

Application I

- Predicting Willing to Pay (WTP)

Final Price

$$r = \frac{P_f - P_i}{P_i}$$

Initial Price

Application II

- Spotting vibrant urban communities

Density of Consumer Activities

$$u_k = \frac{2 \times freq^{(k)} \times div^{(k)}}{freq^{(k)} \times div^{(k)}}$$

Urban Vibrancy Value

Diversity of Consumer Activities

Outline

- Background and Motivation
- Definition and Problem Statement
- Methodology
- Application
- **Evaluation**
- Conclusion and Future Work

Evaluation

Data Description

From Beijing City

Data Sources	Properties	Statistics
Taxi Traces	Number of taxis	13,597
	Effective days	92
	Time period	Apr. - Aug. 2012
	Number of trips	8,202,012
	Number of GPS points	111,602
	Total distance(km)	61,269,029
Residential Communities	Number of residential communities	2,990
	Latitude and Longitude	
	Time period of transactions	04/2011 - 09/2012
POIs	Number of POIs	328668
	Number of POI categories	20
	Latitude and Longitude	
Check-Ins	Number of check-in events	2,762,128
	Number of POI categories	20
	Time Period	01/2012-12/2012

The Application of WTP Prediction

Baselines

- Explicit Features (EF): (i) POI numbers per category; (ii) Average commute distance; (iii) Average commute speed; (iiii) Average commute time; (v) Number of mobilities; (vi) Average distance between POIs.
- Latent Features (LF): Specifically, the latent features are learned from the proposed collective embedding method.
- The combination of EF and LF (ELF).
- Variation of step1 (V-1): using distance-based matching of the records.
- Variation of step2 (V-2): computing the POI embedding as an average of the embeddings.
- Variation of step3 (V-3): averaging over the POI embeddings.

Evaluation Metric

- Root-Mean-Square Error (RMSE)

The Application of WTP Prediction

- Results

Feature set	ELF	LF	EF	V-1	V-2	V-3
RMSE	0.0036	0.0057	0.0422	0.0273	0.0350	0.0193

Spotting vibrant urban communities

Baselines

Learning to Rank

(1)MART: it is a boosted tree model, specifically, a linear combination of the outputs of a set of regression trees.

(2)RankBoost (RB): it is a boosted pairwise ranking method, which trains multiple weak rankers and combines their outputs as final ranking.

(3)LambdaMART (LM): it is the boosted tree version of LambdaRank.

(4)ListNet (LN): It is a listwise ranking model with permutation top-k ranking likelihood as objective function.

(5) RankNet (RN): it uses a neural network to model the underlying probabilistic cost function.

Feature Set

(1)Explicit Features

(2)Latent features

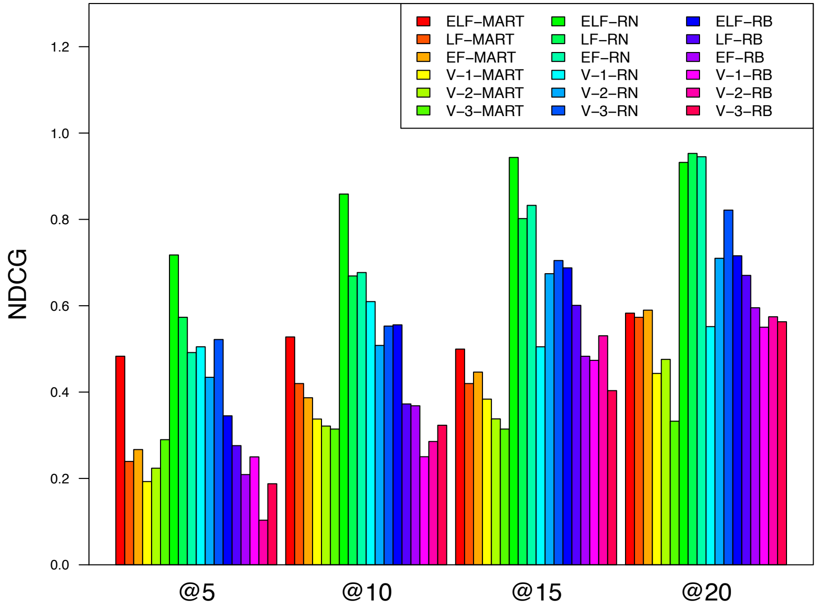
(3)Explicit&Latent features

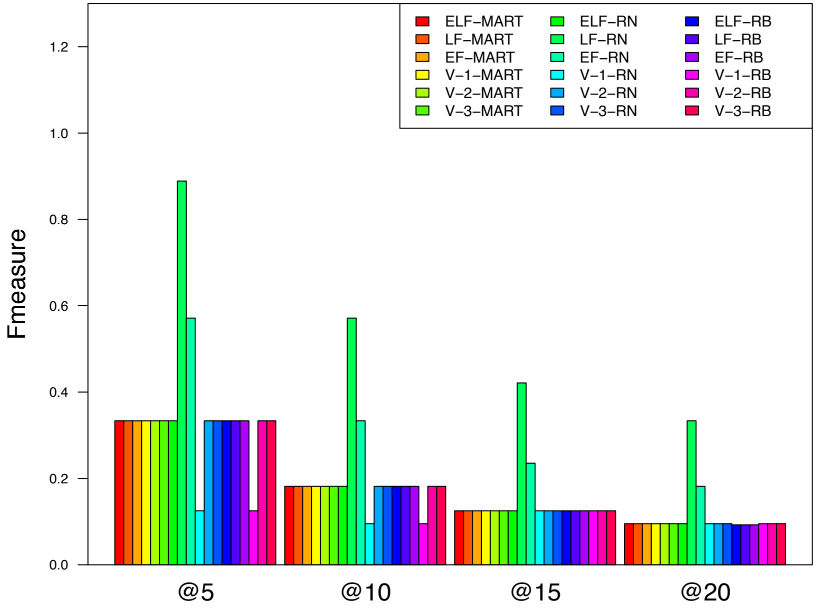
Evaluation

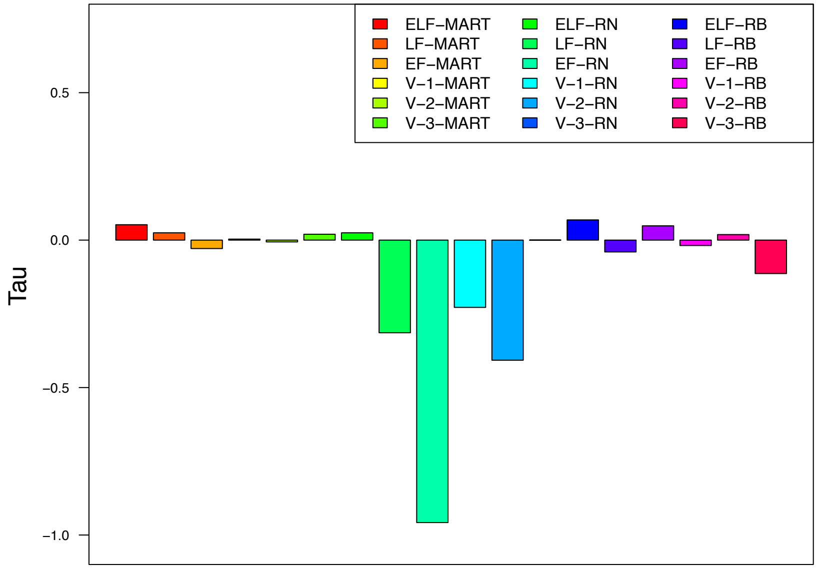
Evaluation Metrics

- **Root-Mean-Square Error (RMSE)**
- **Normalized Discounted Cumulative Gain(NDCG@N)**
  - Evaluate the ranking performance at TopN
- **Kendall's Tau Coefficient(Tau)**
  - Measure the overall ranking accuracy.
- **F-measure@N**
  - "high-vibrancy" and the rating > 3
  - "low-vibrancy" and the rating < 3
  - measure the ranking precision and recall @ TopN

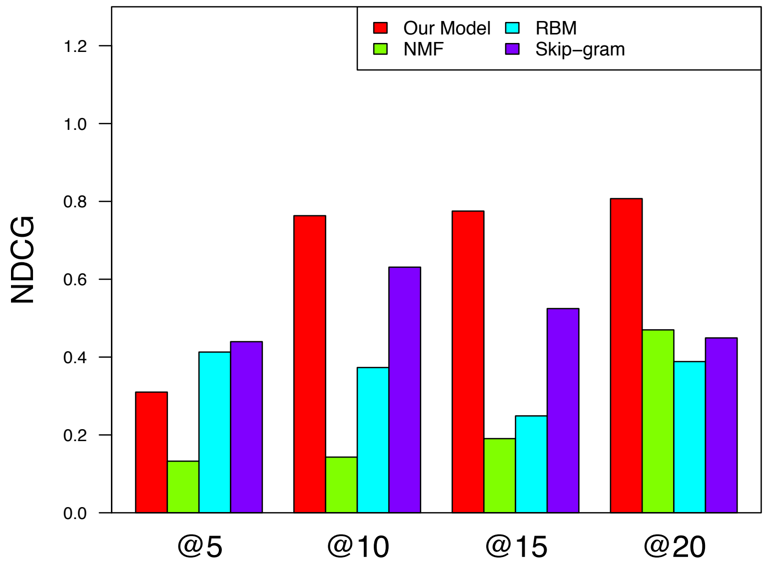
Overall performance

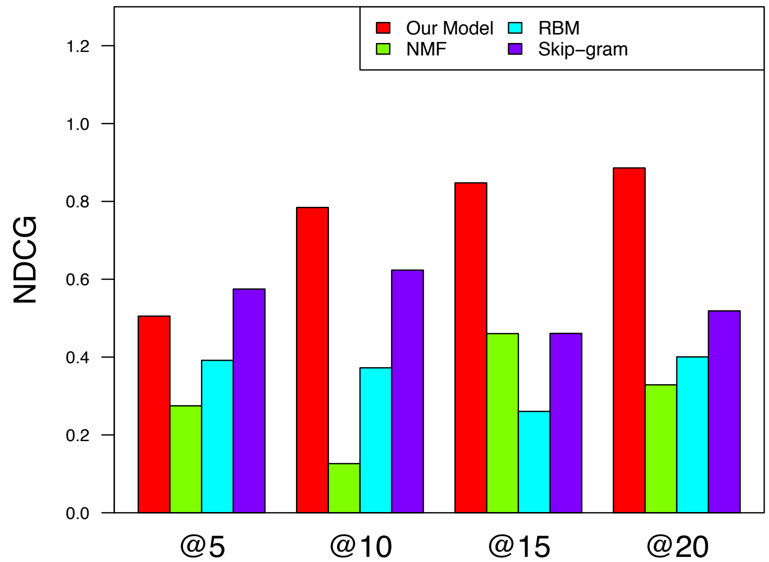


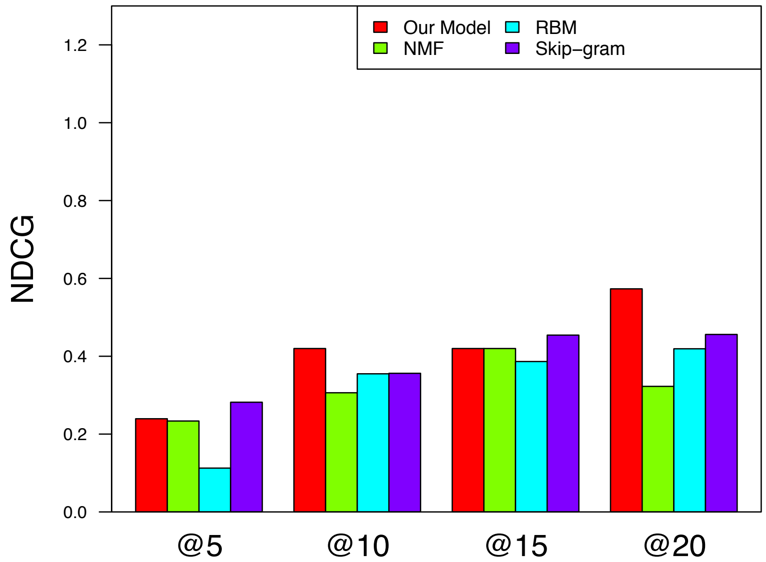


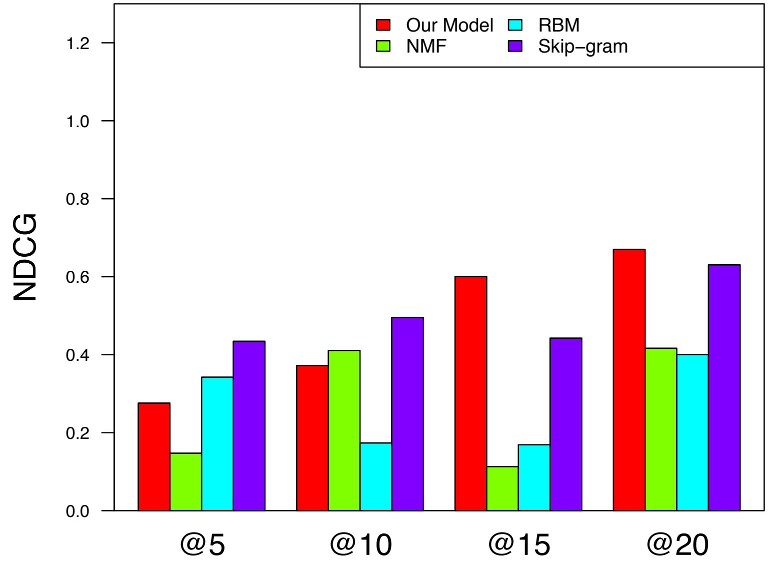


Comparison with Representation Learning Algorithms



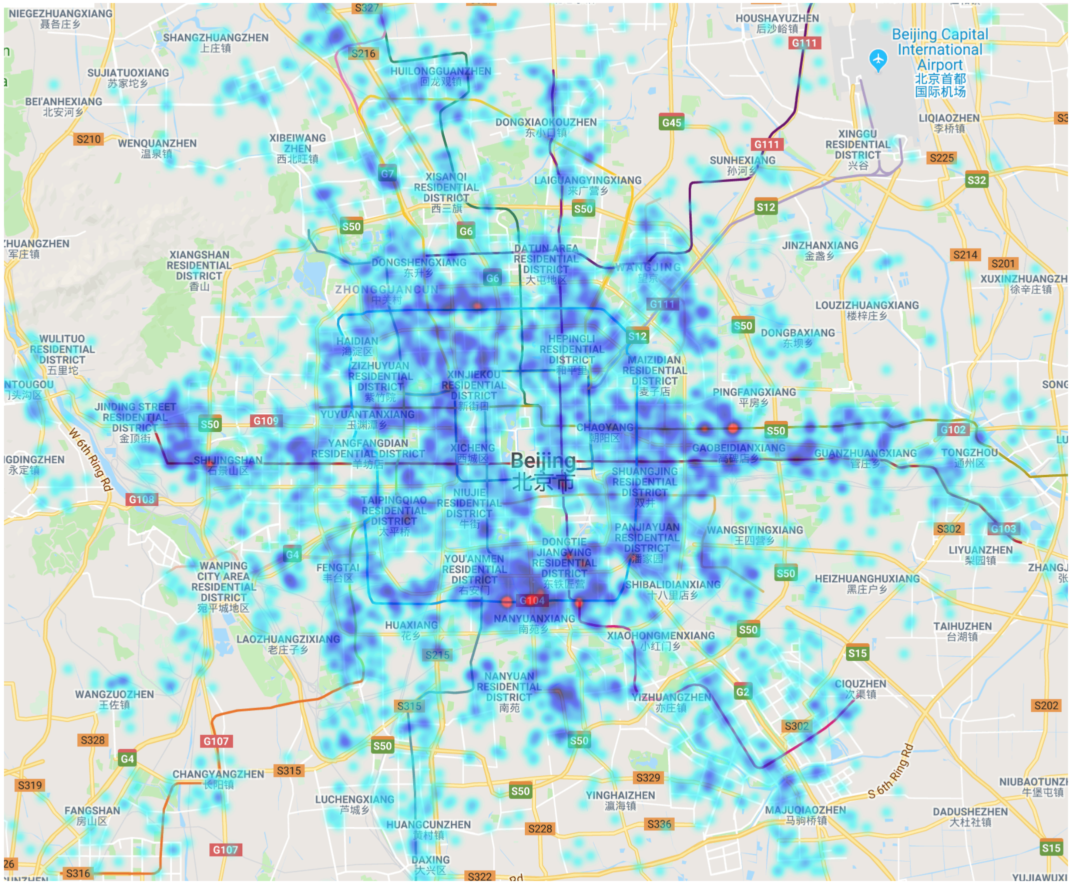






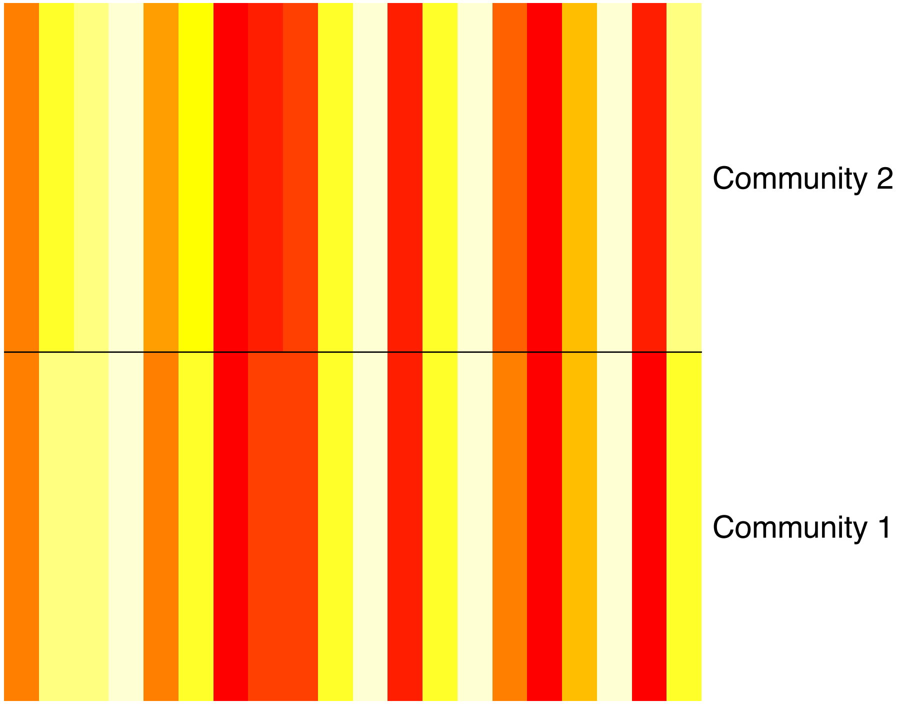
Investigation of Community Structure Properties

- Community Connectivities.



Investigation of Community Structure Properties

- The Learned Representation of the Community Structure



Visualization of the learned structure representations of two similar communities

Outline

- Background and Motivation
- Definition and Problem Statement
- Methodology
- Application
- Evaluation
- **Conclusion**

Conclusion

- We formulate the problem as a learning task over multiple mobility graphs of POIs and propose a novel collective embedding framework.
- We started with a probabilistic propagation method to unify and represent static POIs and dynamic human mobility records as periodic spatial-temporal mobility graphs.
- We then developed a collective embedding method to learn the embeddings of POIs from the obtained mobility graphs.
- Based on the POIs embeddings, we further proposed an unsupervised graph based weighted aggregation method to identify community embeddings.
- The method is effective.

Thanks!

Questions?