Efficient Region Embedding with Multi-View Spatial Networks: A Perspective of Locality-Constrained Spatial Autocorrelations

Yanjie Fu,[1] Pengyang Wang,[1] Jiadi Du,[1] Le Wu,[2] Xiaolin Li[3]
[1]Missouri University of Science and Technology, [2]HeFei University of Technology, [3]Nanjing University
fuyan@mst.edu, pwqt3@mst.edu, jddgh@mst.edu, lewu@hfut.edu.cn, lixl@nju.edu.cn

Abstract

Urban regions are places where people live, work, consume, and entertain. In this study, we investigate the problem of learning an embedding space for regions. Studying the representations of regions can help us to better understand the patterns, structures, and dynamics of cities, support urban planning, and, ultimately, to make our cities more livable and sustainable. While some efforts have been made for learning the embeddings of regions, existing methods can be improved by incorporating locality-constrained spatial autocorrelations into an encode-decode framework. Such embedding strategy is capable of taking into account both intra-region structural information and inter-region spatial autocorrelations. To this end, we propose to learn the representations of regions via a new embedding strategy with awareness of locality-constrained spatial autocorrelations. Specifically, we first construct multi-view (i.e., distance and mobility connectivity) POI-POI networks to represent regions. In addition, we introduce two properties into region embedding: (i) spatial autocorrelations: a global similarity between regions; (ii) top-k locality: spatial autocorrelations locally and approximately reside on top k most autocorrelated regions. We propose a new encoder-decoder based formulation that preserves the two properties while remaining efficient. As an application, we exploit the learned embeddings to predict the mobile checkin popularity of regions. Finally, extensive experiments with real-world urban region data demonstrate the effectiveness and efficiency of our method.

Introduction

An urban region (region for short) is a spatial area along with the agglomeration of high density population, infrastructures of built environments, and socioeconomic activities. As fundamental elements of a city, regions are places where people consume, live, work, and entertain. In this study, we investigate the problem of Learning an Embedding Space for Regions (LESR). The LESR problem refers to the efforts of finding a mapping function to represent regions into vectors in a latent embedding feature space, while preserving certain geospatial characteristics and regularizations. Studying the quantitative representations of regions can help us to better understand the patterns, structures, and dynamics of cities, to support urban planning, to boost commercial activities, and, ultimately, to make our cities more livable and sustainable.

All of the above evidence suggests it is highly appealing to investigate the LESR problem. With the pervasiveness of mobile sensing and location based services, Big Crowd-sourced Geo-tagged Data (BCGD), such as Point-Of-Interests, road networks, trajectories of vehicles, GPS traces from location based Apps, are increasingly available from diverse sources (e.g., buildings, vehicles, sensors, devices) in urban space, and provide great opportunities for understanding and quantifying the configurations, structures, connections of regions. Meanwhile, deep network and word embedding techniques (e.g., Skipgram (Cheng, Greaves, and Warren 2006), LINE (Tang et al. 2015), DeepWalk (Perozzi, Al-Rfou, and Skiena 2014a), AutoEncoder (Liou et al. 2014)) have been widely used to learn an embedding space for social networks and natural languages.

These studies have inspired several preliminary studies (Zhang et al. 2017; Wang and Li 2017; Yao et al. 2018; Wang et al. 2018a; Liu et al. 2019) for region embeddings. Several strategies have been studied. The first strategy learns embeddings from coarse-grained inter-region correlations. For example, the study in (Wang and Li 2017) learns embeddings from a homogeneous network, where vertexes are regions and node embeddings are embeddings of regions, via a skip-gram based method. The second strategy learns embeddings from coarse-grained cross-geo-type (region and other spatiotempral items) correlations. The studies in (Yao et al. 2018; Zhang et al. 2017) learn embeddings from a heterogeneous graph, where vertexes are mixtures of regions, time, mobility arrival events, and/or texts, and the vertex embeddings are not just embeddings of regions, but also embeddings of time, texts, or mobility events, via matrix factorization or skipgram based methods. The third strategy learns embeddings from fine-grained intra-region structural information (e.g., configurations, structures, connections). The study in (Wang et al. 2018a) learns region embeddings from a set of networks, where a network is a region, vertexes are critical infrastructures (i.e., POIs) of a region, and the embedding of a region indeed is the embedding of an entire graph, via an encode-decode framework.

However, after exploiting and exploring the three strategies with BCGD, we find that these strategies have both strengths and weaknesses. For example, the first and second strategy respectively captures inter-region and cross-geo-type correlations, but both of them ignore intra-region structural information; the third strategy captures intra-region structural information, but ignores inter-region correlations.

Thus, there is a compelling need to develop a comprehensive method that can answer the following questions: (i) how can both the inter-region correlations and the intra-region structural information be modeled simultaneously in a unified model? (ii) how can the correlations between regions and other geo-types (e.g., distances, mobility events) be incorporated into the modeling? Specifically, we propose to combine and improve the three strategies from the following perspectives.

First, it is challenging to simultaneously take into account both the inter-region correlations and the intra-region structural information. Inspired by the third strategy, we propose to use a set of POI networks, where an entire network is a region, to capture the intra-region structural information. We find that, compared with skipgram and its variants, an encode-decode framework is more appropriate to encode an entire POI network into a region embedding. Meanwhile, we propose to use a global pairwise similarity matrix of regions to parameterize inter-region spatial autocorrelations. The parameter matrix will be integrated into the optimization to serve as a graph similarity guarantee, in order to regularize the encode-decode process. In this way, we encode both inter-region spatial correlations and intra-region structural information. Second, we propose to use multi-view POI networks at the intra-region level to capture the cross-geo-type correlations that is modeled by the second strategy. We construct two types of networks: (i) distance based POI-POI networks and (ii) mobility connectivity based POI-POI networks, as inputs of our method.

Aside from improving the effectiveness, we improve the efficiency of the model. Be sure to notice that the pairwise spatial autocorrelation matrix is big and the computation cost will exponentially increase with respect to the number of regions. Consequently, it is computationally expensive. We propose a more efficient embedding method by introducing the locality property of the inter-region spatial autocorrelations. The hypothesis of the locality property is that, a region can be described by its top-k most autocorrelated regions. In this way, we maintain efficient since it is performed on a small number of autocorrelated regions, compared to all of the other regions. In some sense, the locality property is a reduced form of the inter-region correlations that result in reduced computation costs. To accurately select top-k most autocorrelated regions, we propose a fused inter-region proximity to fuse (i) distance based proximity, in which if two regions are geographically close, the two regions tend to have higher distance based correlation; and (ii) functionality based proximity, in which if two regions provide similar and/or complementary urban functions and services to residents, the two regions tend to have higher functionality based correlation. The fused inter-region proximity preserves the consistency of locality embedding both in the geographic distance and region functionality domains.

Along these lines, in this paper, we propose a method for learning an embedding space for regions, by combining the strengths of the three existing strategies while reducing computational costs. Specifically, our contributions are: (1) The proposed method can simultaneously capture both inter-region and intra-region structural information by integrating an encode-decode framework with global spatial autocorrelations; (2) The proposed method can incorporate cross-geo-type correlations (i.e., regions, distances, human mobility connectivities) via multi-view networks as inputs; (3) The proposed method can introduce a fused inter-region proximity to preserve the consistency of locality embedding both in the geographic distance and region functionality domains, while maintaining efficient. (4) Extensive experimental results with real-world urban region data demonstrate the improved effectiveness and efficiency of the proposed method.

Preliminaries

We first introduce some definitions and the problem statement, and then present an overview of the proposed method.

Definitions and Problem Statement

Definition 1 (Urban Region) *An urban region is defined as an area that consists of a location (i.e., latitude and longitude) of a residential complex and a neighborhood area (e.g., a circle with radius of 1 kilometer). There are many POIs located in the neighborhood area that provide a variety of urban functions and living services to residents.*

Definition 2 (Intra-region Multi-view POI-POI Networks) *The intra-region multi-view POI-POI networks are defined to represent each region and are the inputs of the proposed model. In this study, for each region, we construct two POI-POI networks from two viewpoints: (i) distance based POI-POI networks, where vertexes are POI categories, and the weight of an edge represents the distance between two associated POI categories; (ii) mobility connectivity based POI-POI networks, where vertexes are POI categories, and the weight of an edge represents the mobility connectivity between two associated POI categories.*

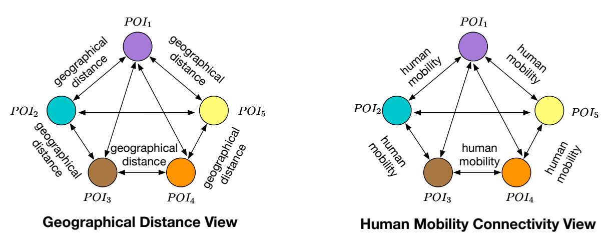


Figure 1: An example of intra-region multi-view POI-POI networks.

Be sure to notice that we regard a POI category (not a POI) as a vertex. Since different regions have different numbers of POIs, if we regard a POI as a vertex, the sizes of the POI-POI networks vary over regions. However, many existing neural network based embedding methods require inputs of fixed size. Since we will exploit an encode-decode framework, we

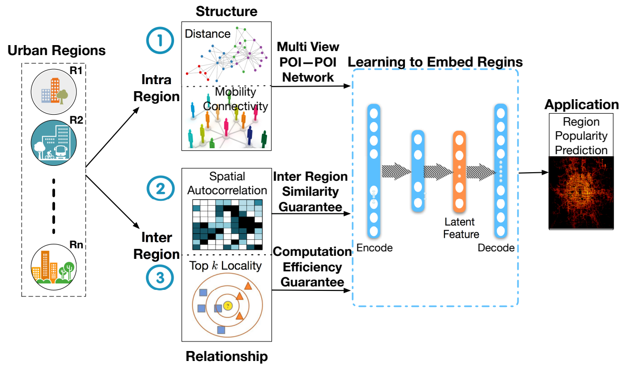


Figure 2: An overview of the proposed region embedding framework.

propose to regard POI categories as vertexes. It is obvious that the number of POI categories is fixed. Consequently, the intra-region POI-POI networks are of fixed sizes as well.

Definition 3 (Fused Inter-region Proximity) *The spatial autocorrelation matrix is a special version of the covariance matrix of regions that describes the pairwise proximity between every pair of two regions. To provide a robust estimation of the inter-region spatial autocorrelations, we propose an empirical metric, which we call fused inter-region proximity, that fuses (i) distance based proximity, which is related to the geographic distance between two regions; and (ii) functionality based proximity, which is measured by the similarity between the POI distributions of two regions.*

Definition 4 (The LESR Problem) *Considering the existence of a set of regions, given the multi-view POI-POI networks for each region, we aim to learn an effective low-rank vector in an embedding space for each region in order to preserve spatial autocorrelations and top-k locality, while maintaining efficient. Essentially, to learn the embeddings of regions, the proposed method should be capable of: (i) describing coarse-grained inter-region spatial autocorrelations; (ii) describing fine-grained intra-region structural information; (iii) exploiting the top-k locality property to reduce computation costs.*

Framework Overview

Figure 2 shows an overview of the proposed framework that includes the following tasks: (i) construction of intra-region multi-view POI-POI networks; (ii) development of a region embedding model that takes into account inter-region spatial autocorrelations and intra-region structural information; (iii) exploiting the top-k locality to reduce computational complexity. Specifically, in the first task, for each region, we construct a distance based POI-POI network and a mobility connectivity based POI-POI network, where vertexes are POI categories and edges denote the relationships (i.e., average distances and human mobility connectivities) between POI categories. In the second task, we develop a new encode-decode framework that takes into account both the inter-region spatial autocorrelations and the intra-region multi-view POI-POI networks for the optimal representation learning. In the third task, we propose an approximate method that incorporates top-k autocorrelated region selection during the learning procedure.

Methodology

We first propose an auto-encoder based formulation by introducing multi-view POI networks as inputs and spatial autocorrelations. In addition, we propose a further improved method by preserving the locality encoding while maintaining efficient.

Model Intuitions

We aim to model the inter-region spatial autocorrelations and intra-region structural information efficiently. The proposed method will be based on following intuitions.

Intuition 1: Intra-region Structural Guarantee. To represent the intra-region structural information, we extract multi-view POI-POI networks (i.e., a distance based POI-POI network and a mobility connectivity based POI-POI network) to describe the intra-region structural information for each region. We need a representation learning based method that can take multiple networks as inputs and encode such fine-grain intra-region profiles in a latent embedding space. Consequently, the method should be able to project multi-view networks into lower-dimensional vectors while preserving corresponding characteristics and structures.

Intuition 2: Inter-region Autocorrelations Guarantee. We introduce the property of spatial autocorrelations into the embedding formulation to help improve region representation learning. Specifically, if two regions show a high autocorrelation score, the embeddings of the two regions should be close to each other. Consequently, the method should be able to learn an embedding space for regions with inter-region autocorrelations guarantee.

Intuition 3: Computation-efficient Guarantee. Suppose that there are N regions in a city, for each region, we need to consider the autocorrelations among this region and other $N-1$ regions. Therefore, the computational complexity for calculating autocorrelations can be estimated as $O(N^2)$. If N is large, the computational cost increases exponentially. We introduce the locality property, which indicates that the impacts of the autocorrelations in a region's embedding can be approximated by its top-k nearest neighbors, instead of all of the other regions, for the purpose of reducing computational costs. We further develop a fused inter-region proximity to provide a robust estimation of top-k autocorrelated regions.

Constructing Intra-region Multi-view POI-POI Networks

We propose to construct a distance based POI-POI network and a mobility connectivity based POI-POI network to describe intra-region structural information for each region.

A View of Geographical Distance. The relative geographical distances among POIs in a region can reflect the spatial structure and configuration of this region via the spatial allocation of buildings (Wang et al. 2018b). Therefore, we construct distance based POI-POI networks from the perspective of geographical distances. In the distance based POI-POI network, vertexes are POI categories, the weights of edges are the average distances among POI categories.

A View of Human Mobility Connectivity. People's outdoor activities include the transitions from one POI to another POI, and ultimately form mobility flows in a region. As a result, human mobility can indicate the connectivity among POIs (Wang et al. 2018b). Therefore, we construct mobility connectivity based POI-POI networks from the perspective of human mobility connectivities. In a mobility connectivity based POI-POI network, vertexes are POI categories, and the weights of edges are human mobility connectivities. We adopt the four-step method proposed in (Wang et al. 2018b) to estimate the human mobility connectivities among POI categories using a probabilistic method. Specifically, we first estimate POI visited probabilities as follows: Given the drop-off point d of a taxi trace, we model the probability of a POI p visited by a passenger as a parametric function, whose input x is the road network distance between the drop-off point d and the POI p: $P(x) = \frac{\beta_1}{\beta_2} \cdot x \cdot exp(1 - \frac{x}{\beta_2})$, where $\beta_1 = \max_x P(x)$ and $\beta_2 = \arg\max_x P(x)$; Second, we aggregate all probabilities from all drop-off points in taxi traces: $\tau(p) = \sum_{d \in \mathcal{D}} P(dis(d,p))$, where $\mathcal{D}$ is the drop-off point set of taxi traces in the region; Then, for each POI category i in a region, the POI category-level aggregated visit probability is given by: $\phi_i = \sum_{p \in i} \tau(p)$, where $p \in i$ denotes the POI p belongs to the i^{th} POI category; Finally, we calculate the flow probabilities from the i^{th} POI category to the j^{th} POI category: $\phi_{\vec{ij}} = \begin{cases} \phi_i \cdot \phi_j, & if\ i \neq j \\ 0, & if\ i = j \end{cases}$. We use the flow probabilities to represent the human mobility connectivities among POI categories.

Formally, let G_i^ϕ denote the original graph representation of the i-th region r_i at the ϕ^{th} view.

Base Model

We adopt Auto-encoder as the base model (Bengio et al. 2007). Auto-encoder is an unsupervised neural network model, which projects the instances in original feature representations into a lower-dimensional feature space via a series of non-linear mappings. The Auto-encoder model involves two steps: encode and decode. The encode part projects the original feature vector to the objective feature space, while the decode step recovers the latent feature representation to a reconstruction space. In the auto-encoder model, we need to ensure that the original feature representation of instances should be as similar to the reconstructed feature representation as possible.

Formally, we flatten each graph into a vector to obtain the original vector representation of regions, denoted by $\mathbf{x}_i^\phi$, where $\mathbf{x}_i^\phi$ is the original vector representation of G_i^ϕ. And $(\mathbf{y}_i^\phi)^1, (\mathbf{y}_i^\phi)^2, \cdots, (\mathbf{y}_i^\phi)^o$ be the latent feature representations of the region at hidden layers $1, 2, \cdots, o$ in the encode step respectively, the encoding result in the objective lower-dimension feature space can be represented as $\mathbf{z}_i^\phi \in \mathbb{R}^d$ with dimension d. Formally, the relationship between these vector variables is denoted by:

$$\begin{cases} (\mathbf{y}_i^\phi)^1 & = \sigma((\mathbf{W}^\phi)^1 \mathbf{x}_i^\phi + (\mathbf{b}^\phi)^1), \\ (\mathbf{y}_i^\phi)^k & = \sigma((\mathbf{W}^\phi)^k (\mathbf{y}_i^\phi)^{k-1} + (\mathbf{b}^\phi)^k), \forall k \in \{2, 3, \cdots, o\}, \\ \mathbf{z}_i^\phi & = \sigma((\mathbf{W}^\phi)^{o+1} (\mathbf{y}_i^\phi)^o + (\mathbf{b}^\phi)^{o+1}). \end{cases} \tag{1}$$

Meanwhile, in the decode step, the input will be the latent feature vector $\mathbf{z}_i^\phi$ (i.e., the output of the encode step), and the final output will be the reconstructed vector $\hat{\mathbf{x}}_i^\phi$. The latent feature vectors at each hidden layers can be represented as $(\hat{\mathbf{y}}_i^\phi)^o, (\hat{\mathbf{y}}_i^\phi)^{o-1}, \cdots, (\hat{\mathbf{y}}_i^\phi)^1$. The relationship between these vector variables is denoted by:

$$\begin{cases} (\hat{\mathbf{y}}_i^\phi)^o & = \sigma((\hat{\mathbf{W}}^\phi)^{o+1} \mathbf{z}_i^\phi + (\hat{\mathbf{b}}^\phi)^{o+1}), \\ (\hat{\mathbf{y}}_i^\phi)^{k-1} & = \sigma((\hat{\mathbf{W}}^\phi)^k (\hat{\mathbf{y}}_i^\phi)^k + (\hat{\mathbf{b}}^\phi)^k), \forall k \in \{2, 3, \cdots, o\}, \\ \hat{\mathbf{x}}_i^\phi & = \sigma((\hat{\mathbf{W}}^\phi)^1 (\hat{\mathbf{y}}_i^\phi)^1 + (\hat{\mathbf{b}}^\phi)^1). \end{cases} \tag{2}$$

where $\mathbf{W}$s and $\mathbf{b}$s are the weight matrices and bias terms to be learned in the model.

The objective of the auto-encoder model is to minimize the loss between the original feature vector $\mathbf{x}$ and the reconstructed feature vector $\hat{\mathbf{x}}$. Formally, the loss function is

$$\mathcal{H}(\mathcal{R}) = \frac{1}{2} \sum_{\phi \in \Phi} \sum_{r_i \in \mathcal{R}} \left\| (\mathbf{x}_i^\phi - \hat{\mathbf{x}}_i^\phi) \right\|_2^2 \tag{3}$$

where r_i denotes the i^{th} region and $\mathcal{R}$ denotes the region set, Φ denotes the view set.

Incorporating Spatial Autocorrelations

The basic Auto-encoder is not able to preserve the property of spatial autocorrelations in representation learning. Therefore, we propose to use a matrix to parameterize the spatial autocorrelations between regions and incorporate the learning of the parameter matrix into the Auto-encoder.

Formally, at the view ϕ, let $Q_{j,i}^\phi$ denote the autocorrelation factor of j^{th} region over i^{th} region. Then, for the i^{th} region, we formulate the autocorrelations from other regions as

$$(\mathbf{Q}_i^\phi)^* = \sum_{j \neq i, r_j \in \mathcal{R}} \mathbf{Q}_{i,j}^\phi \mathbf{z}_j^\phi. \tag{4}$$

Therefore, the updated encode procedure can be represented as:

$$\begin{cases} (\mathbf{y}_i^\phi)^1 & = \sigma((\mathbf{W}^\phi)^1 \mathbf{x}_i^\phi + (\mathbf{b}^\phi)^1), \\ (\mathbf{y}_i^\phi)^k & = \sigma((\mathbf{W}^\phi)^k (\mathbf{y}_i^\phi)^{k-1} + (\mathbf{b}^\phi)^k), \forall k \in \{2, 3, \cdots, o\}, \\ (\mathbf{y}_i^\phi)^{o+1} & = \sigma((\mathbf{W}^\phi)^{o+1} (\mathbf{y}_i^\phi)^o + (\mathbf{b}^\phi)^{o+1}), \\ \mathbf{z}_i^\phi & = (\mathbf{y}_i^\phi)^{o+1} + (\mathbf{Q}_i^\phi)^*. \end{cases} \tag{5}$$

The decode procedure will be the same.

Incorporating Top-k Locality for Efficiency

We further improve the method from the perspective of computational efficiency. Remind that we use a big and exponentially-increased pairwise region-region matrix to parameterize the inter-region spatial autocorrelation. The calculation and search of the parameter matrix is computationally expensive. We propose to develop a more efficient method by introducing the top-k locality property of the spatial autocorrelations. The underlying hypothesis is that, the impacts of autocorrelations in a region's embedding can be approximated by the top-k most autocorrelated regions of this region, which provides a great potential for us to perform locality embedding on a small number of autocorrelated regions, instead of all of the other regions.

To accurately select top-k most autocorrelated regions, we propose a fused inter-region proximity, which is a combination of two perspectives: (i) distance based proximity and (ii) functionality based proximity. In particular, we first calculate the normalized geographical distance $D_{i,j}$ between the i-th region r_i and the j-th region r_j, where $i \neq j, r_i, r_j \in \mathcal{R}$. Later, we adopt the method proposed in (Yuan, Zheng, and Xie 2012) to learn the urban function distributions of r_i and r_j, denoted by $\mathbf{f}_i$ and $\mathbf{f}_j$, respectively. After that, we calculate the urban function similarity of r_i and r_j: $F_{i,j} = cosine(\mathbf{f}_i, \mathbf{f}_j)$ With the extracted distance and functionality based scores, the fused inter-region proximity score $S_{i,j}$ between r_i and r_j is given by:

$$S_{i,j} = \frac{1}{D_{i,j}} \times F_{i,j}. \tag{6}$$

And we rank the regions in terms of the fused proximity scores in a descending order, in order to select top-k regions as the candidate regions to carry out locality embedding.

Finally, the learned region representations that consider the top-k most autocorrelated regions can be formulated as

$$\mathbf{z}_i^{\phi} = (\mathbf{y}_i^{\phi})^{o+1} + \sum_{j \neq i, r_j \in (\mathcal{R}_i)^{TopK}} \mathbf{Q}_{i,j}^{\phi} \mathbf{z}_j^{\phi}, \tag{7}$$

where $(\mathcal{R}_i)^{TopK}$ denotes the top-k most correlated candidate regions for the region r_i.

Experimental Results

We provide an empirical evaluation of the performances of the proposed method on real-world data.

Evaluation Scenario

In order to evaluate and interpret our embedding results, we exploit both linear regression models and regional embeddings to predict regional mobility popularity, which is mobile check-in counts per region. For the region r_i, let $\mathbf{u}_i$ denote the features, P_i denote the ground truth of regional mobility popularity, and $\hat{P}_i$ denote the predicted regional mobility popularity. Then, the application is to learn a linear regression model:

$$P_i = \mathbf{W}^T \mathbf{u}_i + \mathbf{b}, \tag{8}$$

where $\mathbf{W}$ and $\mathbf{b}$ are the corresponding weights and biases. We use different methods to generate features, including our proposed embedding method and other baseline algorithms. We compare the prediction performances of the linear regression model using different features generated by various methods, to evaluate the effectiveness and efficiency of our proposed region embedding method.

Data Description

Table 1 shows the statistics of four data sources used in the experiment. The taxi GPS traces are collected from a Beijing taxi company. Each trajectory contains trip id, distance(m), travel time(s), average speed(km/h), pick-up time and drop-off time, pick-up point and drop-off point. Also, we extract POIs related data from www.dianping.com which is a business review site in China. Moreover, we crawl the Beijing residential region data from www.soufun.com which is the largest real-estate online system in China. Furthermore, the check-in data of Beijing is crawled from www.jiepang.com which is a Chinese version of Fourquare. Each check-in event includes name, category, address, longitude and latitude of POIs.

Table 1: Statistics of the Experimental Data

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 Regions	Number of residential regions	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

Baseline Algorithms

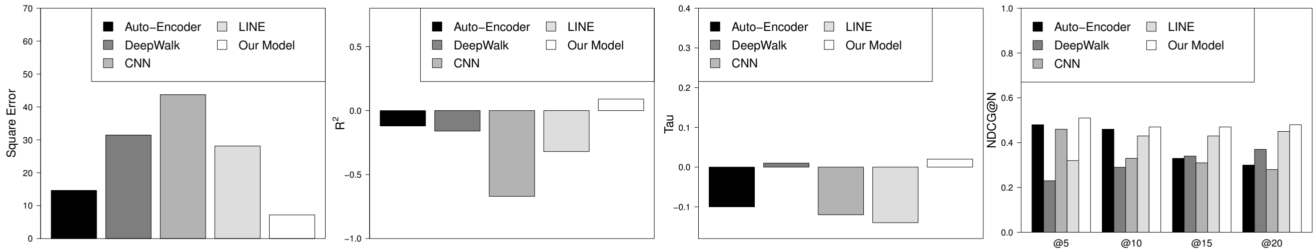
To demonstrate the performances of the proposed method, we compared our method against the following algorithms.

(1) Auto-Encoder with Multi-View Graphs. In the work (Wang et al. 2018b), the authors propose a region embedding method based on Auto-Encoder. Multi-view graphs are built to represent regions and then fed into Auto-Encoder to learn region embeddings. In the experiments, we set the number of hidden layers $= 4$, the size of middle layer $= 50$.

(2) DeepWalk. The DeepWalk model (Perozzi, Al-Rfou, and Skiena 2014a) extends the word2vec model (Mikolov et al. 2013) to the scenario of network embedding. DeepWalk uses local information obtained from truncated random walks to learn latent representations. In the experiments, we set the number of walks $= 70$, the size of representation $= 50$, the walk length $= 30$, and the window size $= 10$.

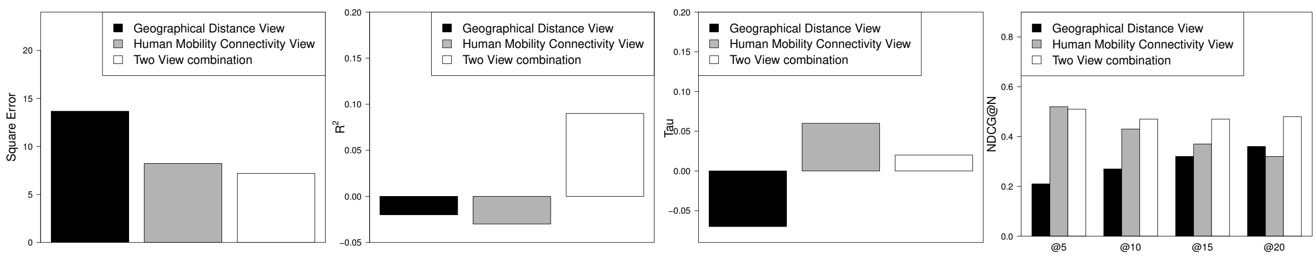
(3) LINE. The LINE model optimizes the objective function that preserves both the local and global network structures with an edge-sampling algorithm (Tang et al. 2015). In the experiments, we set the size of representation $= 50$, the number of negative samples $= 5$, and the starting value of the learning rate $= 0.001$.

(4) CNN. The CNN model refers to Convolutional Neural Network, which projects original feature space into a new



(a) SE (b) R^2 (c) Tau (d) NDCG@N

Figure 3: Overall comparison.



(a) SE (b) R^2 (c) Tau (d) NDCG@N

Figure 4: Performance in different views.

space via a variation of multilayer perceptrons (Lawrence et al. 1997). In the experiments, we set the number of convolutional layer = 2.

For other methods mentioned in the Introduction, the first strategy is for coarse-grained inter-region correlations, and the second strategy is designed for heterogeneous graph, while our proposed method is for fine-grained inter-region and intra-region correlations. Therefore, the first two strategies are both not suitable for the comparison.

Evaluation Metrics

We use the following evaluation metrics to evaluate the performances of our method and baseline algorithms.

Square Error. We utilize *Square Error* (SE) to measure regression errors. $SE = \sum_i (y_i - f_i)^2$, where N is the number of regions. The lower the SE is, the better the learned representation is.

Coefficient of Determination. We utilize the *coefficient of determination* (or R^2 for short) to measure the regression accuracy. R^2 can be represented as $R^2 = 1 - \sum_i (y_i - f_i)^2 / \sum_i (y_i - \bar{y})^2$, where $\bar{y}$ is the mean of benchmarks. The higher the SE is, the better the learned representation is.

Kendall's Tau Coefficient. We utilize *Kendall's Tau Coefficient (Tau)* to measure the overall ranking accuracy. For a region pair $< i, j >$, $< i, j >$ is said to be concordant, if both $y_i > y_j$ and $f_i > f_j$ or if both $y_i < y_j$ and $f_i < f_j$. Also, $< i, j >$ is said to be discordant, if both $y_i < y_j$ and $f_i > f_j$ or if both $y_i < y_j$ and $f_i > f_j$. Tau is given by Tau $= \frac{\#_{conc} - \#_{disc}}{\#_{conc} + \#_{disc}}$.

Normalized Discounted Cumulative Gain. We utilize *Normalized Discounted Cumulative Gain* (NDCG@N) to measure the ranking accuracy at the top-N cases. The discounted cumulative gain (DCG@N) is given by $NDCG[N] = \sum_{i'=1}^{N} \frac{y_{i'}}{log_2(1+i')} / \sum_{i=1}^{N} \frac{y_i}{log_2(1+i)}$, where i denotes the original ranking order of the benchmark and $i^{'}$ denotes the ranking order of the prediction. The larger NDCG@N is, the higher top-N ranking accuracy is.

Overall Performances

We compare our method with the baseline methods in terms of SE, R^2, Tau and NDCG@N. Figure 3 shows our model achieves the best performance.

Because Auto-Encoder with multi-view graphs, DeepWalk, CNN, and LINE are not able to model spatial autocorrelations among regions, the bad performance of baselines validate the essence of spatial autocorrelations from the other side.

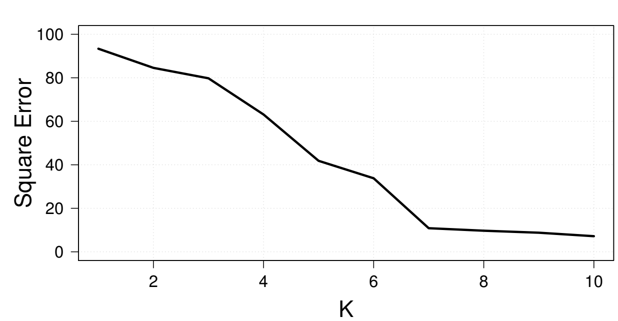


Figure 5: SE sensitivity over k.

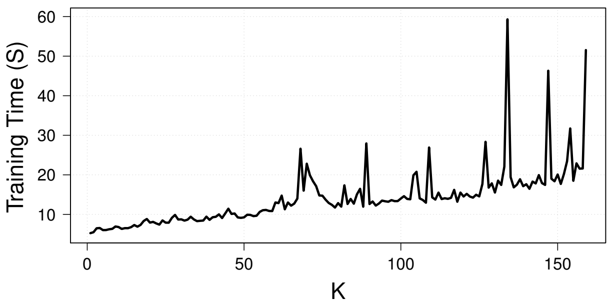


Figure 6: Training time over k.

Study of Performances in Different Views

From Figure 4, it is interesting to observe that the performances of one single view are worse than the combination of these two views. And, the performance of the human mobility connectivity view is slightly better than the geographical distance view.

A potential explanation is that human mobilities have richer semantic information than static geographical distance information. On the other hand, one single view cannot capture complete patterns of regions. The combination of these two views systematically makes up the deficiency of each single view.

The Sensitivity of Regression Errors over k

We investigate the sensitivity of regression errors over k. Figure 5 shows that the error drops significantly with k getting larger, while the decreasing trend will slow down and remain relatively stable after $k = 7$. The observation validates the effectiveness of the proposed top-k locality method that the top-k locality selection can help us reduce the computational complexity, while barely sacrificing correctness as long as we can select a proper small k.

The Model Training Efficiency over k

We also study the model training efficiency over k in Figure 6. We set the training epoch as 100, learning rate as 0.001, numbers of Auto-Encoder layer as 4. We calculate the training time regarding k varying from 1 to 160. One interesting observation is that the training time increases greatly with k getting larger, while there are some fluctuations. This evidence indicates that a small k can help to save a lot of training time, preserving good prediction performance.

Related Work

Network Embedding. Network embedding (Cui et al. 2017) has become an emerging topic since the first seminal work DeepWalk (Perozzi, Al-Rfou, and Skiena 2014b). DeepWalk is inspired from word representation learning method word2vec (Mikolov et al. 2013), which uses random walks in networks to stimulate sentences in language models. And after that random walks have been a general tool for learning embeddings, such as node2vec (Grover and Leskovec 2016) and metapath2vec (Dong, Chawla, and Swami 2017), to name a few. Almost at the same time as DeepWalk, LINE (Tang et al. 2015) is proposed to learn network embedding via preserving the first and second order pairwise proximities. Moreover, predictive text embedding (PTE) extends LINE that works on heterogeneous text networks (Tang, Qu, and Mei 2015).

Spatio-temporal Representation Learning. Our work is relevant to spatio-temporal representation learning. Spatio-temporal representation learning is the elevation of graph representation learning in the spatio-temporal contexts. There are some preliminary studies of spatio-temporal representation learning. For example, *Yao et al.* designed a matrix factorization based method with a geographic probabilistic mutual information index to jointly learn the embeddings of regions and mobility events (Yao et al. 2018). *Zhang et al.* proposed a skipgram based method to learn the embeddings of locations, times, and texts (Zhang et al. 2017). *Wang et al.* developed an encode-decode based method to collectively learn the embeddings of regions from multiple periodic mobility graphs that vary over days (Wang et al. 2018a). *Wang et al.* developed a Multi-task Representation Learning (MTRL) model to predict users' demographic attributes (Wang et al. 2016b).

Urban Computing. Our work is related to the research of urban computing that aims to tackle major issues in cities by analyzing and modeling urban data(e.g., traffic flow, human mobility, and geographical data). One of the biggest challenges in urban computing is to compute with heterogeneous data (Zheng et al. 2014). *Yuan et al.* discovered regional functions of a city using POIs and taxi traces (Yuan, Zheng, and Xie 2012). *Zheng et al.* proposed a semi-supervised learning approach based on a co-training framework that consists of two separated classifiers to infer air quality (Zheng, Liu, and Hsieh 2013). *Bao et al.* proposed a data-driven approach to develop bike lane construction plans based on large-scale real world bike trajectory data (Bao et al. 2017). *Wang et al.* utilized large-scale Point-Of-Interest data and taxi flow data in the city of Chicago to infer crime rate (Wang et al. 2016a).

Conclusion Remarks

In this study, we researched the problem of learning an embedding space for regions. We identified and introduced two properties into region embedding: (i) spatial autocorrelations, which are a global similarity between regions; (ii) top-k locality, which assumes that spatial autocorrelations locally and approximately reside on top k most autocorrelated regions. Inspired by the two properties, we proposed a new embedding formulation that strategically combines an encoder-decoder framework, multi-view POI-POI networks, and locality-constrained spatial autocorrelations into the encoding process, in order to take advantages of various benefits. Specifically, the multi-view POI-POI networks were constructed to capture the cross-geo-types correlations with regions, to represent a region per network, and to serve as inputs of region embedding; the encoder-decoder framework was adapted to address the problem of encoding an entire POI network into a region embedding; the spatial autocorrelations were exploited to capture the inter-region mutual correlations; the top-k locality was utilized to remain efficient.

As an application, we exploited the learned embeddings to predict the mobile check-in popularity of regions. Finally, we presented extensive experiments with real-world urban region data to demonstrate the effectiveness and efficiency of our method.

Acknowledgment

This research was partially supported by the National Science Foundation (NSF) via the grant number: 1755946. This research was partially supported by the University of Missouri Research Board (UMRB) via the proposal number: 4991.

References

Bao, J.; He, T.; Ruan, S.; Li, Y.; and Zheng, Y. 2017. Planning bike lanes based on sharing-bikes' trajectories. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 1377–1386. ACM.

Bengio, Y.; Lamblin, P.; Popovici, D.; and Larochelle, H. 2007. Greedy layer-wise training of deep networks. In *Advances in neural information processing systems*, 153–160.

Cheng, W.; Greaves, C.; and Warren, M. 2006. From n-gram to skipgram to concgram. *International journal of corpus linguistics* 11(4):411–433.

Cui, P.; Wang, X.; Pei, J.; and Zhu, W. 2017. A survey on network embedding. *CoRR* abs/1711.08752.

Dong, Y.; Chawla, N. V.; and Swami, A. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 135–144. ACM.

Grover, A., and Leskovec, J. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Lawrence, S.; Giles, C. L.; Tsoi, A. C.; and Back, A. D. 1997. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks* 8(1):98–113.

Liou, C.-Y.; Cheng, W.-C.; Liou, J.-W.; and Liou, D.-R. 2014. Autoencoder for words. *Neurocomputing* 139:84–96.

Liu, H.; Li, T.; Hu, R.; Fu, Y.; Gu, J.; and Xiong, H. 2019. Joint representation learning for multi-modal transportation recommendation. In *AAAI*, to appear.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.

Perozzi, B.; Al-Rfou, R.; and Skiena, S. 2014a. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 701–710. ACM.

Perozzi, B.; Al-Rfou, R.; and Skiena, S. 2014b. Deepwalk: online learning of social representations. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Tang, J.; Qu, M.; Wang, M.; Zhang, M.; Yan, J.; and Mei, Q. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*, 1067–1077. International World Wide Web Conferences Steering Committee.

Tang, J.; Qu, M.; and Mei, Q. 2015. Pte: Predictive text embedding through large-scale heterogeneous text networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Wang, H., and Li, Z. 2017. Region representation learning via mobility flow. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 237–246. ACM.

Wang, H.; Kifer, D.; Graif, C.; and Li, Z. 2016a. Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 635–644. ACM.

Wang, P.; Guo, J.; Lan, Y.; Xu, J.; and Cheng, X. 2016b. Multi-task representation learning for demographic prediction. In *European Conference on Information Retrieval*, 88–99. Springer.

Wang, P.; Fu, Y.; Zhang, J.; Li, X. L.; and Lin, D. 2018a. Learning urban community structures: A collective embedding perspective with periodic spatial- temporal mobility graphs. *ACM Transactions on Intelligent Systems and Technology (TIST)*.

Wang, P.; Zhang, J.; Liu, G.; Fu, Y.; and Aggarwal, C. 2018b. Ensemble-spotting: Ranking urban vibrancy via poi embedding with multi-view spatial graphs. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, 351–359. SIAM.

Yao, Z.; Fu, Y.; Liu, B.; Hu, W.; and Xiong, H. 2018. Representing urban functions through zone embedding with human mobility patterns. In *IJCAI*, 3919–3925.

Yuan, J.; Zheng, Y.; and Xie, X. 2012. Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 186–194. ACM.

Zhang, C.; Zhang, K.; Yuan, Q.; Peng, H.; Zheng, Y.; Hanratty, T.; Wang, S.; and Han, J. 2017. Regions, periods, activities: Uncovering urban dynamics via cross-modal representation learning. In *Proceedings of the 26th International Conference on World Wide Web*, 361–370. International World Wide Web Conferences Steering Committee.

Zheng, Y.; Capra, L.; Wolfson, O.; and Yang, H. 2014. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5(3):38.

Zheng, Y.; Liu, F.; and Hsieh, H.-P. 2013. U-air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1436–1444. ACM.