Hierarchical Knowledge Graph Learning Enabled Socioeconomic Indicator Prediction in Location-Based Social Network

Zhilun Zhou
Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University
Beijing, China
zzl22@mails.tsinghua.edu.cn

Yu Liu*
Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University
Beijing, China
liuyu2419@126.com

Jingtao Ding
Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University
Beijing, China
dingjt15@tsinghua.org.cn

Depeng Jin
BNRist, Department of Electronic Engineering, Tsinghua University
Beijing, China
jindp@tsinghua.edu.cn

Yong Li
BNRist, Department of Electronic Engineering, Tsinghua University
Beijing, China
liyong07@tsinghua.edu.cn

ABSTRACT

Socioeconomic indicators reflect location status from various aspects such as demographics, economy, crime and land usage, which play an important role in the understanding of location-based social networks (LBSNs). Especially, several existing works leverage multi-source data for socioeconomic indicator prediction in LBSNs, which however fail to capture semantic information as well as distil comprehensive knowledge therein. On the other hand, knowledge graph (KG), which distils semantic knowledge from multi-source data, has been popular in recent LBSN research, which inspires us to introduce KG for socioeconomic indicator prediction in LBSNs. Specifically, we first construct a location-based KG (LBKG) to integrate various kinds of knowledge from heterogeneous LBSN data, including locations and other related elements like point of interests (POIs), business areas as well as various relationships between them, such as spatial proximity and functional similarity. Then we propose a hierarchical KG learning model to capture both global knowledge from LBKG and domain knowledge from several sub-KGs. Extensive experiments on three datasets demonstrate our model's superiority over state-of-the-art methods in socioeconomic indicators prediction. Our code is released at: https://github.com/tsinghua-fib-lab/KG-socioeconomic-indicator-prediction.

CCS CONCEPTS

• **Human-centered computing** → *Social network analysis*; • **Computing methodologies** → *Knowledge representation and reasoning*.

KEYWORDS

Location-based social network, knowledge graph, graph representation learning

ACM Reference Format:
Zhilun Zhou, Yu Liu, Jingtao Ding, Depeng Jin, and Yong Li. 2023. Hierarchical Knowledge Graph Learning Enabled Socioeconomic Indicator Prediction in Location-Based Social Network. In *Proceedings of the ACM Web Conference 2023 (WWW '23), April 30–May 04, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3543507.3583239

*Corresponding Author.

WWW '23, April 30–May 04, 2023, Austin, TX, USA

ACM ISBN 978-1-4503-9416-1/23/04…$15.00
https://doi.org/10.1145/3543507.3583239

1 INTRODUCTION

Socioeconomic indicators of a location, such as population, income level and education level, characterize the attributes of location-based social network (LBSN) from various aspects, and thus are important to the study of LBSNs. A novel and promising way for socioeconomic indicator prediction in LBSNs is location representation learning, which aims to learn low-dimensional embedding vectors for locations. Effective embeddings can characterize various properties of locations to help socioeconomic indicator prediction.

The proliferation of LBSNs has generated a large amount of LBSN data, which provides a comprehensive description of locations. For example, locations with a large mobility outflow on weekday mornings and a large inflow in the evening may serve as residential areas in the city, which indicates that mobility flow data generated by location-based devices can reflect the function of locations. Therefore, LBSN data is essential to location representation learning and socioeconomic indicator prediction.

LBSN data is often of complex structure with multiple types of interconnections therein. As a result, graph structure is usually used to model LBSN data, and thus location representation learning can be formulated as a network representation learning problem. Moreover, most existing works use multi-view graph embedding techniques to model multiple factors such as mobility, spatial vicinity and function of LBSNs [13, 32, 38]. However, such works only consider locations during the final information aggregation process and ignore other elements in LBSNs, which leads to the lack of semantic information. For example, the function of a location is largely reflected by POIs and categories therein, while existing

works do not incorporate such elements into the graphs. Moreover, existing works fail to consider knowledge in LBSNs from a global view, because they model different kinds of knowledge in different graphs, and use simple attention mechanism to fuse them together. However, various kinds of knowledge in LBSNs are deeply entangled with each other. For example, there may be a large mobility flow from residential areas to work areas on weekday mornings, which indicates that mobility knowledge of LBSNs is correlated with their function knowledge. As a result, it is insufficient to model mobility knowledge and function knowledge in different graphs.

Therefore, better socioeconomic indicators prediction faces two key challenges: (1) First, how to model rich semantic information in LBSNs? The semantic information in LBSNs is determined not only by locations, but also by various other elements like POIs and business areas, and complex relationships between them. It is non-trivial to model such information in the graphs. (2) Second, how to capture different LBSN knowledge in a comprehensive way? As mentioned above, existing works mostly consider different knowledge in different graphs, while they fail to consider all the knowledge at a global level.

Inspired by the capability of KG in modeling heterogeneous data [6, 16–18], in this paper, we propose a KG-based model to overcome the challenges. First, according to empirical observation, we construct a location-based KG (LBKG) incorporating various knowledge in LBSNs, i.e., spatiality knowledge, function knowledge, mobility knowledge and business knowledge. The LBKG contains entities including locations, POIs, categories and business areas, as well as complex relationships between them, which can model rich semantic information in LBSN and thus solve the first challenge. As for the second challenge, we propose a hierarchical KG learning framework to integrate both global and domain knowledge in LBSNs. At the higher level, the LBKG incorporates various kinds of knowledge in a single graph, from which we distil global knowledge using KG embedding technique [22]. Besides, at the lower level, we extract a corresponding sub-KG from LBKG for each kind of domain knowledge, and learn embeddings with the assistance of global knowledge. Furthermore, we design a knowledge fusion module to fuse various kinds of domain knowledge. Overall, the LBKG distils global knowledge at the higher level, which also assists the learning of domain knowledge in sub-KGs at the lower level. In this way, the two levels of hierarchical framework together incorporate various kinds of LBSN knowledge comprehensively.

The contributions of this paper can be summarized as follows:

- We introduce KG to socioeconomic indicator prediction problem in LBSN. To the best of our knowledge, we are the first to construct an LBKG to integrate heterogeneous LBSN data in a single multi-relational graph comprehensively. Specifically, we systematically investigate various LBSN knowledge and model it through multiple types of entities and relations in LBKG.
- We propose a hierarchical KG learning model, which leverages LBKG to distil global knowledge from a global view, and use several sub-KGs to capture domain knowledge from different aspects with the assistance of global knowledge.
- We perform extensive experiments on three real-world LBSN datasets, whose results demonstrate our model's superiority over state-of-the-art methods in socioeconomic indicator prediction tasks, outperforming baselines by over 7.5% in terms of R^2. The considerable performance across all indicators on three datasets shows the robustness of our model, and several in-depth experiments further demonstrate the effectiveness of model design.

2 PROBLEM STATEMENT

In this section, we first introduce the preliminary concepts related to our work, and then provide the definition of socioeconomic indicator prediction problem. Specifically, we aim to predict socioeconomic indicators from LBSN data, which is defined as follows.

*Definition 2.1 (**LBSN Data**).* LBSN data $\mathcal{D}_{\text{LBSN}}$ consists of multi-source data including spatial data, attribute data and mobility data. Specifically, spatial data contains spatial information such as the geographic coordinates of POIs and the boundary of locations. Attribute data includes POI brand, category and other attribute information. Mobility data can be the mobility records of mobile devices or taxi trips, reflecting the trajectories of users in LBSNs.

Furthermore, since our model is based on KG, we present the definition of KG here [8, 11, 30].

*Definition 2.2 (**Knowledge Graph**).* A KG is defined as a graph $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{F})$, consisting of the entity set $\mathcal{E}$, relation set $\mathcal{R}$ and fact set $\mathcal{F}$. Each fact in $\mathcal{F}$ is a triplet (h, r, t), where $h, t \in \mathcal{E}$, $r \in \mathcal{R}$, denoting a directional edge from head entity h to tail entity t with relation type r. Besides, each entity can have an entity type determined by a mapping function $\phi : \mathcal{E} \rightarrow \mathcal{A}$, where $\mathcal{A}$ is the set of predefined entity types. The KG schema describes the high-level structure of KG, which shows the types of entities and relations between entity types.

With the concepts defined above, we formulate our research problem as follows.

*Definition 2.3 (**LBSN Socioeconomic Indicator Prediction**).* Given LBSN data $\mathcal{D}_{\text{LBSN}}$ and a set of locations in LBSN $\mathcal{S}_L = \{L_1, L_2, ...L_n\}$, learn a mapping function $f : \mathcal{S}_L \rightarrow \mathcal{V}_I$, where $\mathcal{V}_I$ is the value set of socioeconomic indicator $I \in \mathcal{I}$, i.e., predict the value of socioeconomic indicators for locations in LBSN. Here $\mathcal{I}$ denotes the set of socioeconomic indicators, which are measures of socioeconomic status of a location such as population, income level, education level, etc.

To better solve the problem, two key challenges need to be addressed. First, LBSN data $\mathcal{D}_{\text{LBSN}}$ is heterogeneous, containing multiple types of elements and relationships with rich semantic information. It is difficult to model the semantic information therein. Second, there are various kinds of knowledge in LBSN owing to various sources of LBSN data. Different kinds of knowledge play different roles in LBSN while there are also complex interconnections between them. Therefore, modeling all kinds of knowledge comprehensively is challenging.

3 METHODS

3.1 Framework Overview

To address the aforementioned challenges, we propose a hierarchical KG learning framework, whose architecture is shown in Figure 1. Specifically, to overcome the first challenge, we construct an LBKG

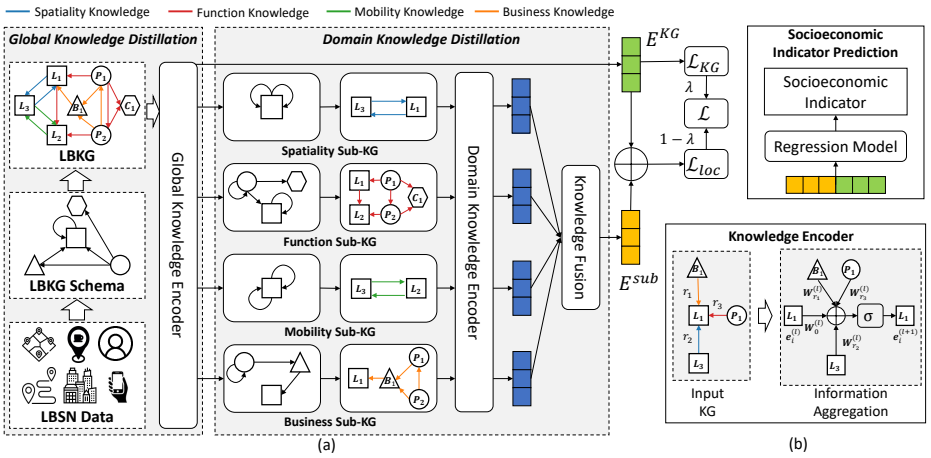


Figure 1: (a) The overall framework of our proposed model, and (b) details of knowledge encoder.

to integrate heterogeneous LBSN data, where elements in LBSN data are modeled as entities and complex relationships are modeled as relations in LBKG. As for the second challenge, we propose a hierarchical model to capture both global and domain knowledge from LBKG. At the higher level, we distil global knowledge from LBKG through KG embedding model. At the lower level, we extract several sub-KGs from LBKG which contain domain knowledge in LBSN from different aspects. Domain knowledge in sub-KGs is distilled with the assistance of global knowledge, and combined with the global knowledge to generate location embeddings. Finally, the learnt location embeddings are fed into a regression model for socioeconomic indicator prediction.

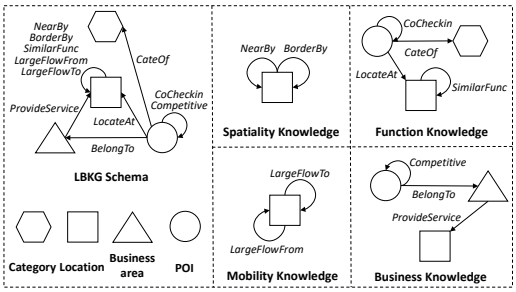


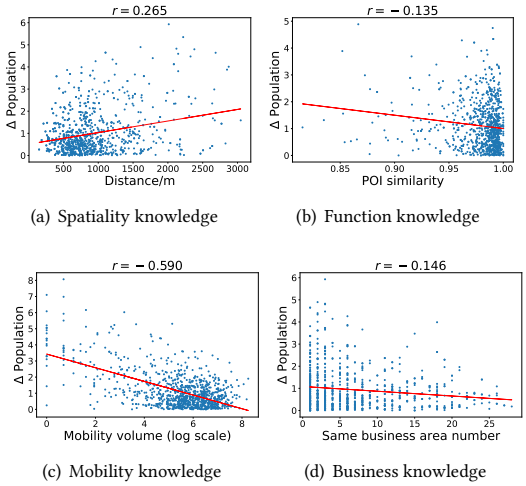
Figure 2: The schema of LBKG and four kinds of LBSN knowledge.

3.2 LBKG Construction

We first construct an LBKG including various elements and relationships in LBSN to capture semantic information therein. Inspired by existing works [17, 18, 38], we conduct correlation analysis between various kinds of LBSN knowledge and socioeconomic indicators to empirically examine their relationship. Then we incorporate LBSN knowledge that is correlated with indicators, i.e., spatiality knowledge, function knowledge, mobility knowledge and business knowledge into LBKG. Specifically, spatiality knowledge describes the spatial vicinity between locations, and function knowledge depicts the function of locations by POIs and categories. Besides, mobility knowledge characterizes the mobility transition pattern between locations and business knowledge contains information about business areas and their relationships with locations. The schema of our LBKG and each kind of LBSN knowledge is shown in Figure 2. We then explain the construction process in detail.

Spatiality knowledge. Urban regions are basic functional areas divided by the main road networks in the city where people live and work, and can be seen as basic spatial units in LBSNs. In this work, we choose urban regions as the locations to study, which are therefore identified as entities in LBKG. According to Tobler's First Law of Geography [21], near things are more related than distant things, which indicates that spatially close locations probably have more similar characteristics. We conduct analysis on our datasets to verify such correlation. Specifically, for each location, we select the nearest location and calculate the distance between them as well as the difference in population (measured on log scale). As shown in Figure 3(a), the spatial distance between regions is positively correlated with their population differences, which suggests that spatially neighboring locations are likely to have similar population. To capture such relationships, we use relation *BorderBy* to link locations that share part of the same boundary and *NearBy* to link locations whose distance is smaller than a threshold. *BorderBy* and *NearBy* describe the vicinity of locations from different scales, which further enriches the spatiality knowledge.

Function knowledge. The function of locations has a significant impact on their properties, which is largely determined by POIs and categories therein [13, 37, 38]. POIs, such as schools, parks and restaurants, are the basic functional units in LBSN, and are classified into several categories by the function and property of POIs. To evaluate the correlation between function and socioeconomic indicators, we calculate the functional similarity between locations as the cosine similarity of POI category distribution. Then for each location, we plot the functional similarity between it and its most similar location against their population difference. As shown in Figure 3(b), locations that are more functionally similar tend to have more similar population. Consequently, to capture function



(a) Spatiality knowledge (b) Function knowledge

(c) Mobility knowledge (d) Business knowledge

Figure 3: Pearson correlation between LBSN knowledge and indicators. The four kinds of LBSN knowledge are correlated with socioeconomic indicators of locations (here we present population as an example).

knowledge of locations, we incorporate each POI as an entity into LBKG and link it to the location it lies in with relation *LocateAt*. And relation *CoCheckin* is used to describe geographical influence between POIs. We also add POI categories as entities to LBKG and link each POI with the category it belongs to with *CateOf* relation. Moreover, we calculate the functional similarity for each pair of locations and link each location with the most similar k locations with relation *SimilarFunc*.

Mobility knowledge. Mobility data, generated by location-based devices in LBSN, depicts the human flow transition patterns across locations, which can also reflect socioeconomic properties of a location [32, 36, 37]. We aggregate the mobility data to get the volume of mobility flow between every two locations. For each location, we plot the correlation between mobility volume and population difference between this location and the location with the largest mobility flow volume. As shown in Figure 3(c), locations with a large mobility flow transition are likely to have a smaller population difference, which empirically inspires us to incorporate mobility knowledge into LBKG. Specifically, for each location L, we choose the top-k locations according to the volume of mobility flow originating from L, and draw relation *LargeFlowTo* from L to those locations. Similarly, we choose the top-k locations according to the mobility flow with destination at L, and draw relation *LargeFlowFrom* from L to those locations.

Business knowledge. As shown in previous works, business knowledge in LBSN is also correlated to socioeconomic status of locations [1, 35]. To incorporate such knowledge into LBKG, we identify core areas of commercial activities as Business Area entities in LBKG such as Sanlitun in Beijing. Business areas are linked with locations and POIs within a spatial range with relation *ProvideService* and *BelongTo*, respectively. Besides, spatially close POIs with the same brand are linked with relation *Competitive* to model their competitive relationship. To empirically evaluate the correlation, for each location, we select the location with the most common business areas they belong to, and plot the population difference against the number of common business areas. From Figure 3(d), we can observe that locations that share more common business areas tend to have a similar population.

In summary, the LBKG contains various elements in LBSN as well as complex relations between them, which incorporate rich semantic information in heterogeneous LBSN data. Furthermore, we present a visualization of LBKG in Appendix A to intuitively examine the capability of LBKG in modeling different types of entities. Besides, it should be noted that the correlations in above analysis are not very strong, which indicates that it is not straightforward to use such correlations to capture characteristics of locations. Therefore, there is a strong need for a better framework to distil and integrate the knowledge together.

3.3 Hierarchical Knowledge Distillation

After constructing the LBKG with rich semantic information, we further propose a hierarchical model to capture different kinds of knowledge comprehensively, whose architecture is shown in Figure 1. Specifically, we distil global knowledge from LBKG through a knowledge encoder at the higher level, which aims to capture the overall characteristics of LBSN. At the lower level, we extract sub-KGs to capture domain knowledge with the assistance of global knowledge distilled, which describes LBSN from different aspects such as spatiality, function, mobility, etc. Finally, domain knowledge is combined with global knowledge after a knowledge fusion module to generate the embeddings of locations.

3.3.1 Global Knowledge Distillation. Based on various kinds of knowledge therein, the LBKG models knowledge in LBSN from a global level. To distil global knowledge from LBKG, in our experiment, we adopt the widely used KG embedding model R-GCN [22] as the knowledge encoder. R-GCN is a kind of graph convolutional network designed for KG, which aggregates information from the neighborhood of entities through each type of relation separately. Therefore, it is able to capture the structural information of KG. The information aggregation process of R-GCN is shown in Figure 1 (b). Specifically, the embedding of entity e_i at the $(l+1)$-th layer can be obtained as:

$$\boldsymbol{e}_i^{(l+1)} = \sigma\Big(\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_i^r} W_r^{(l)}\boldsymbol{e}_j^{(l)} + W_0^{(l)}\boldsymbol{e}_i^{(l)}\Big), \tag{1}$$

where $\boldsymbol{e}_i^{(l)}$ is the embedding of entity e_i at the l-th R-GCN layer, $\mathcal{N}_i^r$ is the set of entities connected to entity e_i via relation r and $W_r^{(l)}$, $W_0^{(l)}$ are learnable weight matrices at the l-th R-GCN layer. σ is a nonlinear activation function.

3.3.2 Domain Knowledge Distillation. Locations play different roles with respect to different knowledge. For example, locations that are functionally similar may be far from each other geographically. Therefore, apart from global knowledge that describes LBSN at the higher level, different kinds of domain knowledge also need to be considered separately at the lower level. Specifically, we extract several sub-KGs according to different kinds of knowledge mentioned in Section 3.2 in LBKG to capture domain knowledge in LBSN.

Table 1: The basic information of three real-world datasets and basic statistics of LBKGs.

City	Basic Information		LBKG Statistics		
	#Locations	Indicators	#Entities	#Relations	#Facts
Beijing	1010	population, number of takeaway orders, number of restaurants, economic activity	36,752	11	188,985
Shanghai	2476	population, economic activity	58,175	11	368,453
NYC	2120	population, education level, average income, number of crimes, land usage	87,020	7	357,464

Spatiality sub-KG. Locations that are spatially close may also share similar characteristics. Consequently, we define the schema of sub-KG including locations and spatial relationship between them, i.e., *NearBy* and *BorderBy* to capture spatiality knowledge.

Function sub-KG. As mentioned above, POIs and their categories largely determine the function of locations. We extract locations, POIs and Categories as well as functional relations between them as Function sub-KG.

Mobility sub-KG. To capture mobility knowledge, we keep all the locations as well as relation *LargeFlowTo* and *LargeFlowFrom* between them in Mobility sub-KG.

Business sub-KG. In order to capture business knowledge in LBKG, Business sub-KG preserves Business Areas, Locations and POIs as well as *BelongTo*, *ProvideService* and *Competitive* relations.

After extracting sub-KGs from LBKG, we design a domain knowledge encoder to distil domain knowledge in each sub-KG. Note that each sub-KG we extracted is a KG as well, so we also leverage the aforementioned KG embedding model R-GCN to learn embeddings for locations. Moreover, it should be noted that the input here is embeddings after global knowledge encoder, which can also preserve global knowledge in the whole LBKG.

Since different knowledge may contribute differently to socioeconomic indicator prediction, we further adopt a knowledge fusion module [31] to adaptively fuse the knowledge. Specifically, let $\{G_1, G_2, ...G_M\}$ denotes the set of sub-KGs, we calculate the importance of each sub-KG as:

$$w_{G_k} = \frac{1}{|\mathcal{S}_L|} \sum_{j \in \mathcal{S}_L} q^\top \tanh(W \boldsymbol{e}_j^{G_k} + b), \tag{2}$$

where $\mathcal{S}_L$ is the set of locations, $e_j^{G_k}$ is the embedding of location L_j in sub-KG G_k, q is the attention vector, W is the weight matrix and b is the bias vector. Then the weight of each sub-KG can be calculated by normalizing w_{G_k} with softmax function:

$$\beta^{G_k} = \frac{\exp(w_{G_k})}{\sum_{j=1}^{M} \exp(w_{G_j})}. \tag{3}$$

Finally, we fuse the embedding from each sub-KG to obtain the sub-KG embeddings of locations as $\boldsymbol{e}_j^{sub} = \sum_{j=1}^{M} \beta^{G_k} \boldsymbol{e}_j^{G_k}$.

3.4 Framework Optimization

So far we have obtained two embeddings for each location $\boldsymbol{e}^{KG}$ and $\boldsymbol{e}^{sub}$ with encoded global knowledge in LBKG and domain knowledge in sub-KGs. To better preserve the semantic knowledge in KG as well as similarity between locations, we design the KG completion loss to capture the plausibility of facts in KG at the higher level, and the location loss to preserve similarity between locations at the lower level.

3.4.1 KG Completion Loss. To better preserve global knowledge in LBKG, after global knowledge encoder, scoring function DistMult [34] is used to calculate the plausibility of each triplet (h, r, t) in LBKG:

$$\phi(h, r, t) = (\boldsymbol{e}_h \odot \boldsymbol{e}_r)^\top \boldsymbol{e}_t, \tag{4}$$

where $\boldsymbol{e}_h$, $\boldsymbol{e}_r$ and $\boldsymbol{e}_t$ are the embeddings of head entity h, relation r and tail entity t. We aim to correctly calculate the plausibility for triplets in KG, i.e., triplets existing in KG should have a higher score. As a result, we optimize a cross-entropy loss for KG completion [12]:

$$\mathcal{L}_{KG} = \sum_{(h,r,t) \in \mathcal{F}} -\log \frac{\exp(\phi(h, r, t))}{\sum_{t' \in \mathcal{E}} \exp(\phi(h, r, t'))}. \tag{5}$$

3.4.2 Location Loss. While KG completion loss aims to better preserve global knowledge in LBKG, we also design location loss to capture location similarity. In order to incorporate global knowledge and domain knowledge, we first fuse location embeddings learnt from LBKG and sub-KGs as $\boldsymbol{e}^{fuse} = \boldsymbol{e}^{KG} + \boldsymbol{e}^{sub}$. Then we adopt a widely used loss function to measure location similarity in previous works [13, 32, 38]. Specifically, We estimate the distribution of mobility flow originating from location L_i as:

$$\hat{p}(L_j|L_i) = \frac{\exp(\boldsymbol{e}_i^{fuse\top} \boldsymbol{e}_j^{fuse})}{\sum_{k=1}^{n} \exp(\boldsymbol{e}_i^{fuse\top} \boldsymbol{e}_k^{fuse})}. \tag{6}$$

We then obtain source and destination location pairs from mobility data, which is denoted as a set $\mathcal{M}$, and each element $(L_i, L_j) \in \mathcal{M}$ corresponds to a mobility record from location L_i to location L_j. The location loss is calculated as:

$$\mathcal{L}_{loc} = \sum_{(L_i, L_j) \in \mathcal{M}} -log\hat{p}(L_j|L_i). \tag{7}$$

The final loss is the combination of $\mathcal{L}_{KG}$ and $\mathcal{L}_{loc}$:

$$\mathcal{L} = \lambda \mathcal{L}_{KG} + (1 - \lambda)\mathcal{L}_{loc}, \tag{8}$$

where λ is a hyperparameter. Finally, we get the embeddings of locations as the concatenation of e^{KG} and e^{sub}, which are fed into regression model for socioeconomic indicator prediction.

4 EXPERIMENTS

4.1 Datasets

To examine the effectiveness of our model, we conduct experiments on three LBSN datasets from Beijing, Shanghai and New York City (NYC). The datasets contain LBSN data including locations and their socioeconomic indicators, and we construct a LBKG for each dataset according to methods in 3.2. The basic information of datasets and statistics of LBKGs are presented in Table 1. Specifically, in Beijing dataset, we remove all food category POIs to avoid data leakage in restaurant number prediction. Besides, in NYC dataset, due to

Table 2: Performance comparison with baselines on Beijing and Shanghai dataset. Best results are presented in bold, and the second best results are underlined.

	Beijing dataset												Shanghai dataset					
	Population			#Orders			#Restaurants			Economic activity			Population			Economic activity		
Model	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
Node2vec	0.680	0.864	0.593	2.009	2.541	0.504	0.741	0.927	0.739	1.030	1.286	0.627	0.702	0.954	0.534	1.190	1.581	0.385
GCN	0.627	0.803	0.648	2.070	2.702	0.439	0.823	1.044	0.668	1.074	1.338	0.596	0.716	0.972	0.517	1.265	1.640	0.338
GAT	0.725	0.913	0.545	1.969	2.599	0.481	0.773	0.987	0.703	1.130	1.409	0.551	0.771	1.026	0.461	1.349	1.721	0.271
ZE-Mob	0.798	1.003	0.452	2.160	2.846	0.378	0.944	1.199	0.563	1.274	1.614	0.412	0.839	1.145	0.329	1.384	1.750	0.247
MGFN	0.705	0.900	0.558	1.979	2.655	0.458	0.858	1.119	0.619	1.174	1.504	0.489	0.812	1.097	0.384	1.348	1.727	0.266
MV-PN	0.860	1.082	0.362	1.889	2.631	0.468	0.862	1.093	0.637	1.097	1.456	0.521	0.964	1.281	0.160	1.514	1.885	0.126
HDGE	0.686	0.871	0.586	1.961	2.692	0.442	0.918	1.166	0.585	1.220	1.564	0.447	0.765	1.038	0.448	1.365	1.728	0.265
HUGAT	0.703	0.878	0.580	1.903	2.578	0.489	0.834	1.080	0.645	1.159	1.464	0.516	0.831	1.100	0.380	1.364	1.745	0.251
MVURE	0.626	0.806	0.646	1.829	2.475	0.530	0.756	0.980	0.708	1.057	1.377	0.572	0.749	1.020	0.467	1.341	1.720	0.272
ours	**0.559**	**0.710**	**0.725**	**1.633**	**2.253**	**0.610**	**0.612**	**0.801**	**0.805**	**0.935**	**1.196**	**0.677**	**0.637**	**0.870**	**0.612**	**1.149**	**1.516**	**0.434**
Improv.	10.7%	11.6%	11.9%	10.7%	9.0%	15.1%	17.4%	13.6%	8.9%	9.2%	7.0%	8.0%	9.3%	8.8%	14.6%	3.4%	4.1%	12.7%

Table 3: Performance comparison with baselines on NYC dataset. Best results are presented in bold, and the second best results are underlined.

	Population			Education level			Income level			Crime			Land use	
Model	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	NMI	ARI
Node2vec	0.449	0.661	0.077	0.090	0.118	0.675	0.228	0.298	0.505	0.497	0.637	0.429	0.809	0.520
GCN	0.446	0.664	0.069	0.089	0.118	0.677	0.213	0.281	0.559	0.443	0.577	0.531	0.814	0.522
GAT	0.439	0.644	0.124	0.092	0.120	0.664	0.227	0.293	0.522	0.479	0.614	0.468	**0.817**	**0.527**
ZE-Mob	0.473	0.682	0.018	0.128	0.158	0.419	0.291	0.368	0.243	0.601	0.752	0.203	0.535	0.194
MGFN	0.450	0.664	0.070	0.109	0.137	0.559	0.254	0.326	0.408	0.499	0.642	0.420	0.724	0.365
MV-PN	0.456	0.649	0.110	0.140	0.178	0.257	0.315	0.398	0.117	0.631	0.773	0.157	0.256	0.013
HDGE	0.445	0.664	0.068	0.104	0.135	0.575	0.237	0.315	0.446	0.525	0.668	0.370	0.761	0.341
HUGAT	0.466	0.673	0.043	0.131	0.162	0.383	0.296	0.375	0.215	0.525	0.669	0.368	0.784	0.466
MVURE	0.447	0.647	0.116	0.095	0.121	0.658	0.228	0.293	0.523	0.472	0.608	0.478	0.728	0.375
ours	**0.423**	**0.617**	**0.196**	**0.084**	**0.108**	**0.728**	**0.203**	**0.266**	**0.604**	**0.397**	**0.505**	**0.640**	0.812	0.512
Improv.	3.6%	4.2%	58.1%	5.6%	8.5%	7.5%	4.7%	5.3%	8.1%	10.4%	12.5%	20.5%	-0.6%	-2.8%

the lack of business area data and some POI related data, we omit business knowledge in our model. The details of datasets can be found in Appendix B.

4.2 Experiment Settings

4.2.1 Baselines. We compare the performance of our model with several state-of-the-art baselines. Specifically, for single view methods, we choose **Node2vec** [5], **GCN** [14], **GAT** [26], **ZE-Mob** [36] and **MGFN** [32]. For multi-view methods, we choose **MV-PN** [3], **HDGE** [28], **HUGAT** [13] and **MVURE** [38]. All the baselines use the same settings as ours, i.e., they first learn location embeddings, and use regression or clustering model for socioeconomic indicator prediction. The details of baselines and implementation can be found in Appendix C.

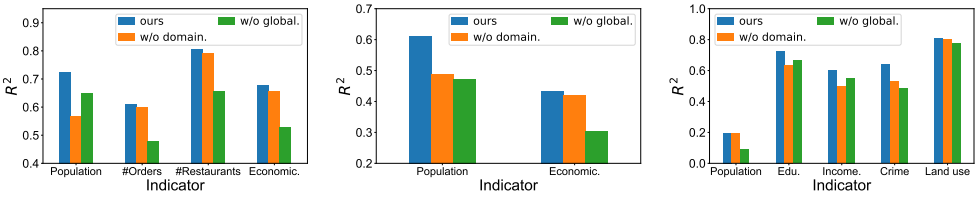
4.2.2 Evaluation Metrics. We evaluate the performance of models on two kinds of socioeconomic prediction tasks, prediction and clustering. For all indicators except for land usage in NYC dataset, we apply Ridge regression model [7] to predict the value of socioeconomic indicator by embeddings of locations. Specifically, we split all locations into train, validation and test set by 6:2:2, and adopt widely used metrics Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and coefficient of determination (R^2) to measure the performance. For clustering task, we apply K-means algorithm on embeddings of all locations and use community districts in NYC as the ground truth. The performance is measured by Normalized Mutual Information (NMI) [23] and Adjusted Rand Index (ARI) [10], the details of which are presented in Appendix D.

4.3 Overall Performance

The overall performance of our model and baselines on three datasets is shown in Table 2-3, from which we have the following findings.

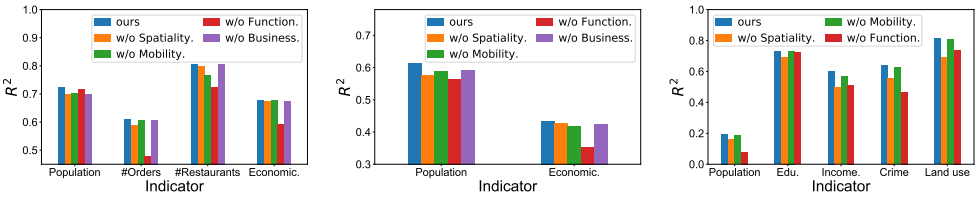
First, our model outperforms all baselines on almost all indicators owing to its capability of integrating various kinds of LBSN knowledge comprehensively. For example, on Beijing dataset, our model outperforms existing methods with an improvement in R^2 by 8.0% to 15.1%. Such great improvement shows the capability of our model to integrate heterogeneous LBSN data and capture knowledge related to socioeconomic indicators. Besides, some state-of-the-art baseline models such as MGFN, MVURE and HUGAT only conduct experiments on a single dataset with less than 300 locations. In comparison, our model achieves the best performance on various socioeconomic indicators across three much larger datasets from different cities with totally different LBSN structures and environments, which further demonstrates the robustness and generalizability of our model.

Second, graph neural networks achieve rather good performance, suggesting rich semantics in LBKG. Take NYC dataset as an example, the GAT model achieves similar or even higher performance with the best baseline MVURE. This is probably because spatiality, mobility and part of function knowledge is already included in the



(a) Beijing dataset (b) Shanghai dataset (c) NYC dataset

Figure 4: Performance comparison of models without global knowledge or domain knowledge. "domain." and "global." represent domain knowledge and global knowledge, respectively.



(a) Beijing dataset (b) Shanghai dataset (c) NYC dataset

Figure 5: Performance comparison of models without specific domain knowledge.

edges between locations in LBKG, which further shows that LBKG can integrate various knowledge effectively.

Third, among different baselines, MVURE and HUGAT generally perform better because they incorporate information in LBSN from various aspects, while models considering only mobility flow data, i.e., ZE-Mob and MGFN, perform rather worse. This finding shows the importance of considering various kinds of LBSN knowledge. However, even though MVURE and HUGAT utilize various information, they still perform worse than our model, which further shows that KG can better integrate knowledge from LBSN data.

In summary, our proposed model achieves considerable improvement over state-of-the-art baselines across three datasets, demonstrating the capability of LBKG in integrating heterogeneous LBSN data and effectiveness of our hierarchical design. Further analysis on the impact of hyperparameters can be found in Appendix E. We also evaluate the efficiency of our model in Appendix F.

4.4 Ablation Study

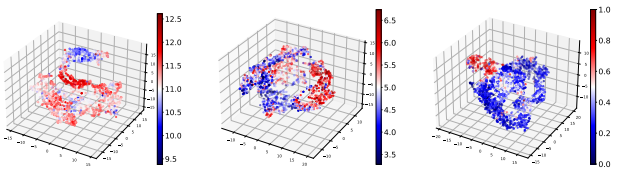
To evaluate the influence of global and domain knowledge, we remove the LBKG or sub-KGs and use the rest embeddings for socioeconomic indicator prediction. As shown in Figure 4, performance on all indicators drops when omitting LBKG or sub-KGs, which demonstrates the effectiveness of global knowledge in LBKG as well as domain knowledge in sub-KGs. Besides, we find that two kinds of knowledge may contribute differently to different indicators. Take Beijing dataset as an example, embeddings with only global knowledge (w/o sub-KGs) perform better than that with only domain knowledge (w/o LBKG) on number of orders, number of restaurants and economic activity prediction. But embeddings with domain knowledge perform better on population prediction. It further indicates that knowledge from global and local level is both important and needed to be modeled in a hierarchical way.

Besides, we analyze the effectiveness of different LBSN knowledge by removing different sub-KGs and corresponding entities and relations in LBKG. From Figure 5, we can observe that performance becomes worse on almost all indicators when removing each sub-KG, which shows the necessity of incorporating various kinds of LBSN knowledge in our model. Besides, on Beijing and Shanghai datasets, the impact of function knowledge is rather large on all indicators, which suggests that it is important to consider POI and POI categories in LBKG.

4.5 Result Visualization

To intuitively examine correlations between location embeddings and socioeconomic indicators, we map the learnt location embeddings into 3-dimensional vectors by t-SNE [25], and visualize them in Figure 6, where the color represents the value of socioeconomic indicators. It can be observed that locations can be distinguished based on socioeconomic indicators. For example, as shown in Figure 6(a), red dots mostly lie in the middle part of the figure, which correspond to locations with high income level. Meanwhile, blue dots, which correspond to low income level locations, mostly locates at the upper part. Similarly, locations with fewer crimes are likely to lie in the right part of Figure 6(b). In addition, in Figure 6(c), locations with high education level and those with lower education level can be easily separated spatially. The findings demonstrate that location embeddings learnt by our model can distinguish locations by socioeconomic indicators, and thus help socioeconomic indicator prediction.

Moreover, to get an intuitive understanding of the clustering results, we visualize location clusters of different models in Figure 7.



(a) Income level (b) Crime (c) Education level

Figure 6: Visualization of location embeddings, where color represents the value of indicators. (Income level and crime are measured on log scale.)

The results of some other models are shown in Appendix G. For better visual effect, here we only visualize locations in the borough of Manhattan, which has only 12 clusters according to community districts. It can be observed that our result fits the ground truth better than baselines. For example, HDGE fails to distinguish the four districts in the northern part, MVURE generates fewer clusters than ground truth in the southern part of Manhattan, while the clusters of HUGAT are more likely to mix together. The visualization result further shows the effectiveness of our hierarchical KG learning model in capturing various LBSN knowledge.

(a) Ground truth (b) ours (c) HUGAT (d) MVURE (e) HDGE

Figure 7: Visualization of clustering results in borough of Manhattan, NYC. Red squares mark where baselines perform worse than our model.

5 RELATED WORK

5.1 Socioeconomic Indicator Prediction in LBSN

Socioeconomic indicator prediction in LBSN has attracted extensive attention, and the main concern of such task is how to utilize LBSN data properly. Early studies usually use feature engineering methods, manually extracting features from LBSN data to infer socioeconomic indicators. Yang et al. [35] use social context of LBSN such as number of visitors and number of POIs to predict the commercial activeness of a location through a linear model. Dong et al. [1] extract features from restaurant data, such as restaurant number and review number, to predict socioeconomic indicators like population and consumption level through LASSO regression. Besides, Wang et al. [27] consider the features of locations as well as interplay between locations. Feature engineering can reveal the importance of different features, but it is labor-intensive and the performance is largely affected by feature design.

In recent years, many socioeconomic indicator prediction models based on representation learning have been proposed. For example, ZE-Mob [36] borrow the idea of word embedding methods and model location as word and mobility event as context to learn location embeddings for socioeconomic indicator prediction. GMEL [19] considers geographic contextual information of locations for flow prediction. Some studies focus on mobility data in LBSN and model the mobility patterns between locations at different time [9, 32] or with different features [33]. Some studies further try to incorporate LBSN data from various sources through multiple graphs. For example, HDGE [28] constructs a flow graph and a spatial graph to predict prime rate, income and house price. MV-PN [3] constructs POI networks considering both geographical and human mobility views. Some studies [20, 38] model various relationships in LBSN such as mobility, POI and spatial vicinity in different graphs and adaptively fuse them together. HUGAT [13] designs several meta-paths to capture different relationships in LBSN from a heterogeneous graph, and aggregate information from meta-path based neighbors of locations. However, existing works do not leverage KG to model LBSN data, and they fail to integrate heterogeneous LBSN data in a single graph comprehensively.

5.2 Knowledge Graph Application in LBSN

KG has been widely applied in LBSNs thanks to its capability of modeling complex relationships between various elements in LBSN. For example, STKG [29] utilizes spatio-temporal KG to combine information in LBSN, and convert mobility prediction problem to a KG completion problem. KnowSite [17] applies KG to site selection problem and shows great performance and interpretability. UKGC [15] models geographical and functional knowledge as well as interactions between users and POIs for location recommendation. RFP-KMN [16] models flow transitions as relations in KG to predict flow patterns. Besides, Tan et al. [24] constructs a KG of traffic system to discover implicit traffic knowledge in LBSN. These works show the ability of KG to model complex entities and relations in LBSN, while they focus on different problems from ours. In this work, we introduce KG to a new problem of socioeconomic indicator prediction in LBSN.

6 CONCLUSION

In this paper, we propose a hierarchical KG learning model for socioeconomic indicator prediction in LBSN. To capture semantic information from heterogeneous LBSN data, we construct a LBKG consisting of various types of entities and complex relations between them. Moreover, we design a hierarchical model to learn global knowledge and various kinds of domain knowledge in a comprehensive way. We conduct extensive experiments on three real-world datasets, and the results demonstrate the effectiveness and robustness of our model.

This work also has some limitations. For example, we use LBSN data from various sources, some of which may be not easy to collect, e.g., business area data. Besides, we do not consider the dynamic change of socioeconomic indicators and other characteristics of locations over time, which can be a future work. In the future, another promising direction is to integrate inductive KG techniques like NodePiece [4] for inductive learning. Moreover, we plan to change our model to an end-to-end framework for socioeconomic indicator prediction.

ACKNOWLEDGMENTS

This work was supported in part by the BNRist, the National Key Research and Development Program of China under 2020YFB2104005, the National Natural Science Foundation of China under U20B2060, U21B2036, and the Guoqiang Institute, Tsinghua University under 2021GQG1005. We sincerely thank Yunke Zhang for collecting and preprocessing NYC data from Safegraph.

REFERENCES

[1] Lei Dong, Carlo Ratti, and Siqi Zheng. 2019. Predicting neighborhoods' socioeconomic attributes using restaurant data. *Proceedings of the National Academy of Sciences* 116, 31 (2019), 15447–15452.

[2] Lei Dong, Xiaohui Yuan, Meng Li, Carlo Ratti, and Yu Liu. 2021. A gridded establishment dataset as a proxy for economic activity in China. *Scientific Data* 8, 1 (2021), 1–9.

[3] Yanjie Fu, Pengyang Wang, Jiadi Du, Le Wu, and Xiaolin Li. 2019. Efficient region embedding with multi-view spatial networks: A perspective of locality-constrained spatial autocorrelations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 906–913.

[4] Mikhail Galkin, Etienne Denis, Jiapeng Wu, and William L Hamilton. 2021. Node-piece: Compositional and parameter-efficient representations of large knowledge graphs. *arXiv preprint arXiv:2106.12144* (2021).

[5] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 855–864.

[6] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. 2020. A survey on knowledge graph-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[7] Arthur E Hoerl and Robert W Kennard. 1970. Ridge regression: applications to nonorthogonal problems. *Technometrics* 12, 1 (1970), 69–82.

[8] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia D'amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. 2021. Knowledge Graphs. *Comput. Surveys* 54, 4 (2021), 1–37.

[9] Mingliang Hou, Feng Xia, Haoran Gao, Xin Chen, and Honglong Chen. 2022. Urban Region Profiling With Spatio-Temporal Graph Neural Networks. *IEEE Transactions on Computational Social Systems* (2022).

[10] Lawrence Hubert and Phipps Arabie. 1985. Comparing partitions. *Journal of Classification* 2, 1 (1985), 193–218.

[11] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and S Yu Philip. 2021. A Survey on Knowledge Graphs: Representation, Acquisition and Applications. *IEEE Transactions on Neural Networks and Learning Systems* (2021).

[12] Rudolf Kadlec, Ondrej Bajgar, and Jan Kleindienst. 2017. Knowledge base completion: Baselines strike back. *arXiv preprint arXiv:1705.10744* (2017).

[13] Namwoo Kim and Yoonjin Yoon. 2022. Effective Urban Region Representation Learning Using Heterogeneous Urban Graph Attention Network (HUGAT). *arXiv preprint arXiv:2202.09021* (2022).

[14] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[15] Chang Liu, Chen Gao, Depeng Jin, and Yong Li. 2021. Improving Location Recommendation with Urban Knowledge Graph. *arXiv preprint arXiv:2111.01013* (2021).

[16] Jia Liu, Tianrui Li, Shenggong Ji, Peng Xie, Shengdong Du, Fei Teng, and Junbo Zhang. 2021. Urban Flow Pattern Mining based on Multi-source Heterogeneous Data Fusion and Knowledge Graph Embedding. *IEEE Transactions on Knowledge and Data Engineering* (2021).

[17] Yu Liu, Jingtao Ding, and Yong Li. 2021. Knowledge-driven site selection via urban knowledge graph. *arXiv preprint arXiv:2111.00787* (2021).

[18] Yu Liu, Jingtao Ding, and Yong Li. 2022. Developing knowledge graph based system for urban computing. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Geospatial Knowledge Graphs*. 3–7.

[19] Zhicheng Liu, Fabio Miranda, Weiting Xiong, Junyan Yang, Qiao Wang, and Claudio Silva. 2020. Learning geo-contextual embeddings for commuting flow prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 808–816.

[20] Yan Luo, Fu-lai Chung, and Kai Chen. 2022. Urban Region Profiling via A Multi-Graph Representation Learning Framework. *arXiv preprint arXiv:2202.02074* (2022).

[21] Harvey J Miller. 2004. Tobler's first law and spatial analysis. *Annals of the Association of American Geographers* 94, 2 (2004), 284–289.

[22] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*. Springer, 593–607.

[23] Alexander Strehl and Joydeep Ghosh. 2002. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research* 3, Dec (2002), 583–617.

[24] Jiyuan Tan, Qianqian Qiu, Weiwei Guo, and Tingshuai Li. 2021. Research on the Construction of a Knowledge Graph and Knowledge Reasoning Model in the Field of Urban Traffic. *Sustainability* 13, 6 (2021), 3191.

[25] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing Data Using t-SNE. *Journal of Machine Learning Research* 9, 11 (2008).

[26] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[27] Hongjian Wang, Daniel Kifer, Corina Graif, and Zhenhui Li. 2016. Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 635–644.

[28] Hongjian Wang and Zhenhui Li. 2017. Region representation learning via mobility flow. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 237–246.

[29] Huandong Wang, Qiaohong Yu, Yu Liu, Depeng Jin, and Yong Li. 2021. Spatio-Temporal Urban Knowledge Graph Enabled Mobility Prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–24.

[30] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Transactions on Knowledge and Data Engineering* 29, 12 (2017), 2724–2743.

[31] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous graph attention network. In *The World Wide Web Conference*. 2022–2032.

[32] Shangbin Wu, Xu Yan, Xiaoliang Fan, Shirui Pan, Shichao Zhu, Chuanpan Zheng, Ming Cheng, and Cheng Wang. 2022. Multi-Graph Fusion Networks for Urban Region Embedding. *arXiv preprint arXiv:2201.09760* (2022).

[33] Fengli Xu, Yong Li, and Shusheng Xu. 2020. Attentional multi-graph convolutional network for regional economy prediction with open migration data. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2225–2233.

[34] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575* (2014).

[35] Su Yang, Minjie Wang, Wenshan Wang, Yi Sun, Jun Gao, Weishan Zhang, and Jiulong Zhang. 2017. Predicting commercial activeness over urban big data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–20.

[36] Zijun Yao, Yanjie Fu, Bin Liu, Wangsu Hu, and Hui Xiong. 2018. Representing urban functions through zone embedding with human mobility patterns. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*.

[37] Jing Yuan, Yu Zheng, and Xing Xie. 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.

[38] Mingyang Zhang, Tong Li, Yong Li, and Pan Hui. 2021. Multi-view joint graph representation learning for urban region embedding. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*. 4431–4437.

A LBKG VISUALIZATION

We leverage KG embedding model TuckER to learn a embedding for each entity in LBKG (on Beijing dataset), an visualize the embeddings using t-SNE in Figure 8, where different colors represent different types of entities. Specifically, we sample 1000 POIs and preserve all entities of other types. It can be observed that different types of entities can be spatially distinguished, which demonstrate the ability of KG to model various types of entities in LBSN.

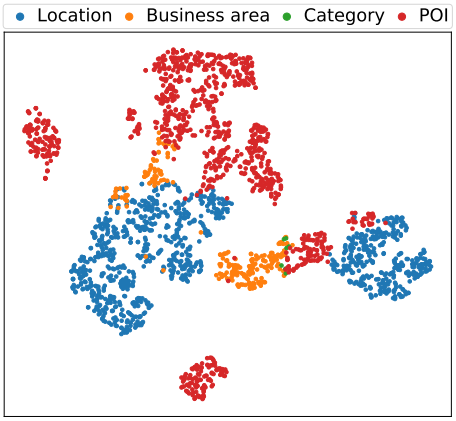


Figure 8: Visualization of different types of entities in LBSN on Beijing dataset.

B DETAILS OF DATASETS

Here we present the details of three datasets.

- **Beijing Dataset**. This dataset contains 1010 regions in Beijing partitioned by main road networks. Besides, each region has four socioeconomic indicators. Population data is collected from WorldPop[1], which contains the estimated population in 2018. The number of takeaway orders is obtained from Meituan, a life service platform. Restaurant data [1] in 2017 is collected from Dianping, the largest rating platform of restaurants in China. We also use firm data from [2] to reflect the economic activity of regions.
- **Shanghai Dataset**. This dataset includes 2476 regions in Shanghai, which are also separated by main road networks. In addition, the data of two socioeconomic indicators, i.e., population and economic activity (reflected by number of firms), is collected from the same sources as Beijing dataset.
- **NYC Dataset**. In this dataset, we use census tracts in NYC as the regions to study. The population, education level and average income data is collected from Safegraph [2], where the education level is measured by the ratio of population with bachelor degree. Besides, we collect crime data from NYC Open Data[3]. As for land usage, we follow [38] and use community districts as the ground truth of clustering.

[1]https://hub.worldpop.org/geodata/summary?id=24924
[2]https://www.safegraph.com/
[3]https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i

C DETAILS OF BASELINES AND IMPLEMENTATION

The details of baselines are as follows.

Single view methods.

- **Node2vec** [5]: It uses random walks to learn node embeddings with skip-gram models.
- **GCN** [14]: GCN aggregates information from neighborhood for each node.
- **GAT** [26]: GAT applies attention modules to aggregate information from neighborhood with different weights.
- **ZE-Mob** [36]: This model uses the co-occurrence of origin-destination locations to learn embeddings of locations from mobility flow data.
- **MGFN** [32]: It first fuses mobility graphs with similar patterns, and then learns embeddings of locations via multi-level attention mechanism.

Multi-view methods.

- **MV-PN** [3]: It constructs multi-view POI-POI networks for each location and learns location embeddings through an encoder-decoder framework.
- **HDGE** [28]: It constructs a spatial graph and a flow graph and jointly learns location embeddings from them.
- **HUGAT** [13]: It defines several meta-paths to capture semantics in LBSN from a heterogeneous graph, and adopts heterogeneous graph attention network to learn location embeddings.
- **MVURE** [38]: This work models different types of correlations between locations with different graphs, and proposes a joint learning module to learn location embeddings.

Since graph neural networks models take homogeneous graphs as input, we use our LBKG with only the locations and all edges between locations preserved as the input graph for Node2vec, GCN and GAT, where the edge types are ignored. For all models, the embedding dimension is set as 64 for fair comparison. We tune the hyperparameters for each model and choose the best one. Besides, to guarantee the robustness of models, we run all experiments for 5 times and report the average results.

D DETAILS OF CLUSTERING EVALUATION METRICS

We adopt widely used Normalized Mutual Information (NMI) [23] and Adjusted Rand Index (ARI) [10] to evaluate the performance of clustering.

- **NMI**. It measures the purity of clustering results and is defined as:

$$NMI = \frac{2 \times I(Y;C)}{H(Y) + H(C)}, \tag{9}$$

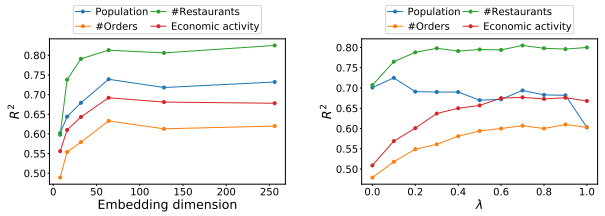
  where $I(Y;C)$ is the mutual information between true labels set Y and cluster labels set C. $H(Y)$ and $H(C)$ represent the entropy of true labels and cluster labels. NMI ranges from 0 to 1, and a larger NMI indicates a better clustering result.
- **ARI**. It measures the similarity between true clusters and clustering results, which is defined as:

$$ARI = \frac{RI - Expected(RI)}{max(RI) - Expected(RI)}, \tag{10}$$

where Rand Index (RI) calculate the correctness of each pair of locations in clustering results by $RI = \frac{TP+TN}{TP+FP+TN+FN}$. For example, if two locations have the same true label and lie in the same cluster, they are considered as a true positive (TP) pair. True negative (TN), false positive (FP) and false negative (FN) pairs are defined similarly. A larger ARI indicates a better clustering result.

E IMPACT OF HYPERPARAMETERS

In this section, we investigate the impact of two important hyperparameters in our model, i.e., embedding dimension and the weight λ in the loss function Equation 8.



(a) Embedding dimension (b) Lambda

Figure 9: Performance comparison under different hyperparameters on Beijing dataset.

Figure 9(a) shows the performance on Beijing dataset with embedding dimension in {8,16,32,64,128,256}. It can be observed that the R^2 goes up on all indicators when embedding dimension increases from 8 to 64 and then remains stable. This is because embedding with a larger dimension can preserve more information, while a too large dimension may make the embedding sparse. Besides, the result shows that it is appropriate to choose 64 as the embedding dimension in our experiment.

We also evaluate our model under different λ ranging from 0.0 to 1.0 and present the result on Beijing dataset in Figure 9(b). It can be observed that for different indicators, the trend of performance curves are different. For example, as λ increases, the performance on economic activity, number of orders and number of restaurants goes larger, while the performance on population drops. This finding indicates that importance of global knowledge and domain knowledge varies across indicators, which further shows the necessity to consider such knowledge in our model.

F EFFICIENCY EVALUATION

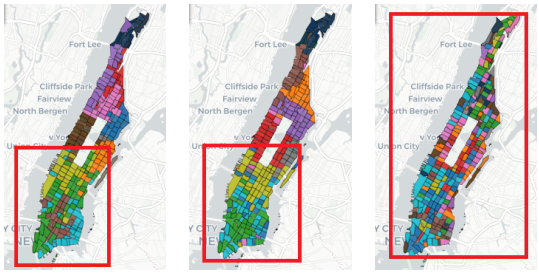
To evaluate the efficiency of our model, we compare the training time of our model and baselines in Table 4. It can be observed that our model is more efficient than some state-of-the-art baselines like ZE-Mob, MGFN and MVURE. Moreover, the training time of our model on all datasets is within an hour, which is acceptable in practice.

G CLUSTERING RESULTS

Here we present the clustering results of MGFN, ZE-Mob and MV-PN in Figure 10. It can be observed that they generally fit the ground truth worse than models in Figure 7, because these three models do not incorporate spatial information.

Table 4: Training time comparison with baselines on three datasets.

Model	Dataset		
	Beijing	Shanghai	NYC
Node2vec	3 min	4 min	3 min
GCN	1 min	2 min	2 min
GAT	1 min	2 min	2 min
ZE-Mob	1.3 h	7h	3.6 h
MGFN	1.4 h	12 h	8 h
MV-PN	1 min	1 min	2 min
HDGE	17 min	47 min	35 min
HUGAT	1 min	3 min	3 min
MVURE	40 min	3 h	2.3 h
ours	25 min	47 min	53 min



(a) MGFN (b) ZE-Mob (c) MV-PN

Figure 10: Visualization of clustering results in borough of Manhattan, NYC. Red squares mark where baselines perform worse than our model.