Region-Wise Attentive Multi-View Representation Learning For Urban Region Embedding

Weiliang Chan
Heilongjiang University
Harbin Shi, Heilongjiang Sheng, China
chanweiliang@s.hlju.edu.cn

Qianqian Ren*
Heilongjiang University
Harbin Shi, Heilongjiang Sheng, China
renqianqian@hlju.edu.cn

ABSTRACT

Urban region embedding is an important and yet highly challenging issue due to the complexity and constantly changing nature of urban data. To address the challenges, we propose a **R**egi**o**n-Wise **M**ulti-View R**e**presentation Lea**r**ning (ROMER) to capture multi-view dependencies and learn expressive representations of urban regions without the constraints of rigid neighbourhood region conditions. Our model focuses on learn urban region representation from multi-source urban data. First, we capture the multi-view correlations from mobility flow patterns, POI semantics and check-in dynamics. Then, we adopt global graph attention networks to learn similarity of any two vertices in graphs. To comprehensively consider and share features of multiple views, a two-stage fusion module is further proposed to learn weights with external attention to fuse multi-view embeddings. Extensive experiments for two downstream tasks on real-world datasets demonstrate that our model outperforms state-of-the-art methods by up to 17% improvement.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; • **Information systems** → *Clustering*; *Data analytics.*

KEYWORDS

Graph Attention, Multi-Feature Fusion, Urban Region Embedding, Graph Neural Network

ACM Reference Format:
Weiliang Chan and Qianqian Ren. 2023. Region-Wise Attentive Multi-View Representation Learning For Urban Region Embedding. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), October 21–25, 2023, Birmingham, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3583780.3615194

1 INTRODUCTION

Urban region embedding is a classical embedding problem whose purpose is to to learn quantitative representations of regions from multi-sourced data. The problem has been found useful in many real-world applications such as socio-demographic feature prediction [11, 23], crime prediction [22, 24, 27], economic growth prediction[10] and land usage classification [15, 24, 25, 27].

Urban region embedding is challenging due to the highly complex nature of urban data, deep learning approaches have attracted much interest in recent years. In particular, GNNs have shown great potential for learning low-dimensional embeddings of graph-structured data [1, 4, 6, 12, 20, 21]. Several existing studies that have attempted to integrate POI data and human mobility to characterize regions and obtain attractive results[3, 5, 13, 14, 22, 24–27]. For example, the methods [24, 27] learns graph embeddings by combining multi-graph. The method[27] employs an attention mechanism and simple crosstalk operations to aggregate information, while MFGN[24] deeply resolves the relevance of regions in fine-grained human mobility patterns. However, existing studies have predominantly focused on capturing correlations between neighboring regions, overlook the significant influence of distant regions in complex urban systems. Attention mechanism such as GAT[19] assigns aggregation weights based solely on the influence of neighboring nodes, and self-attention[18] solely focuses on the regions themselves, both approaches fail to consider the potential correlations between regions and suffer from computational complexity[8]. As a result, these methods produce sub-optimal embeddings and limit the ability of the model to capture the underlying urban dynamics and features.

To tackle these challenges, we propose a **R**egi**o**n-Wise **M**ulti-Graph R**e**presentation Lea**r**ning (ROMER) for effective urban region representation with multi-view data. Our ROMER model adopts heterogeneous graph neural framework with respect to human movement patterns, POI semantics and check-in dynamics. In addition, a multi-graph aggregation module is design to capture region-wise dependencies and non-linear correlations among regions. Finally, we design an efficient and cost-effective attentive fusion module that learning adaptive weights for information sharing across diverse views with external attention and gating mechanism to fuse multi-view in an efficient and deeply collaborative manner.

We have extensively evaluated our approach through a series of experiments using real-world data. The results demonstrate the superiority of our method compared to state-of-the-art baselines, achieving improvements of up to 17%. In addition, our method exhibits significant computational efficiency.

2 PRELIMINARIES

In this section, we first give some notations and define the urban region embedding problem. We partition a city into N regions $V = \{v_1, v_2, \cdots, v_N\}$, where v_i denotes the i-th region. For a **trip** p =

*Corresponding author

CIKM '23, October 21–25, 2023, Birmingham, United Kingdom

ACM ISBN 979-8-4007-0124-5/23/10…$15.00
https://doi.org/10.1145/3583780.3615194

(v_o, v_d), v_o denotes the origin region and v_d denotes the destination region where $1 \leq o, d, \leq N$. Given the set of regions V, we further define the following.

Definition 1. (Human mobility feature). The human mobility feature is defined as a trip sets $\mathcal{P} = \{p_i | p_i.v_o, p_i.v_d \in V\}$ that occur in urban areas. $i = \{1, 2, \cdots, M\}$, where M is the number of trips.

Definition 2. (Semantic feature). The semantic feature describes the functional similarity among regions. Similar regions may necessarily be close in space. In this work, the semantic feature of the region is characterized with POIs in the located region. Given a region v_i, its semantic feature is represented as:

$$\mathcal{S} = \{s_i \mid s_i \in \mathbb{R}\}, i = 1, 2, \cdots, n \tag{1}$$

Definition 3. (Dynamic feature). The dynamic feature describes the activities of POIs in regions, which integrates human activities and POIs information. Given a region v_i, its dynamic feature is represented as:

$$\mathcal{G} = \{g_i \mid g_i \in \mathbb{R}\}, i = 1, 2, \cdots, m \tag{2}$$

Urban region embedding problem. We denote the three families of features for a region v_i as a vector $e^i \in \mathbb{R}^r$, where r is the number of features. Then, the our final goal is to learn a mapping function F,

$$\mathcal{E} = F(\mathcal{E}^{\mathcal{V}}) \tag{3}$$

where $\mathcal{E}^{\mathcal{V}} \in \mathbb{R}^{M \times r}$ are three features of all regions in V.

3 METHODOLOGY

The overall architecture of ROMER proposed in this paper is shown in Figure. 1, which consists of three components: the region-wise graph learning module, the multi-graph aggregation module and attentive fusion module.

3.1 Region Wise Graph Learning Module

In this section, we elucidate the utilization of various types of region dependencies for encoding multi-graph.

3.1.1 Mobility based Region Graph. The movement of people within urban spaces across regions can be understood by examining their interactions. When people travel between different origins (O) and destinations (D), we can observe similarities in their patterns if they have the same O/D region. In simpler terms, by analysing the similarity of O/D patterns, we can identify important potential features related to human mobility. Given a set of human mobility M, we can use

$$s_{v_j}^{v_i} = \left|(v_i, v_j) \in M\right| \tag{4}$$

to calculate the similarity value between region v_o and region v_d, where $|.|$ counts the size of the trip. Then we employ distributions $p_o(v \mid v_i)$ and $p_d(v \mid v_i)$ to describe the origin and destination contexts of a region v_i as follows:

$$p_o(v \mid v_i) = \frac{s_{v_i}^v}{\sum_v s_{v_i}^v}, \quad p_d(v \mid v_i) = \frac{s_v^{v_i}}{\sum_v s_v^{v_i}}. \tag{5}$$

The two types of dependencies were defined by us based on the source and destination context of each region, as follows,

$$\mathcal{D}_O^{ij} = \text{sim}\left(p_o(v \mid v_i), p_o(v \mid v_j)\right), \tag{6}$$

$$\mathcal{D}_D^{ij} = \text{sim}\left(p_d(v \mid v_i), p_d(v \mid v_j)\right), \tag{7}$$

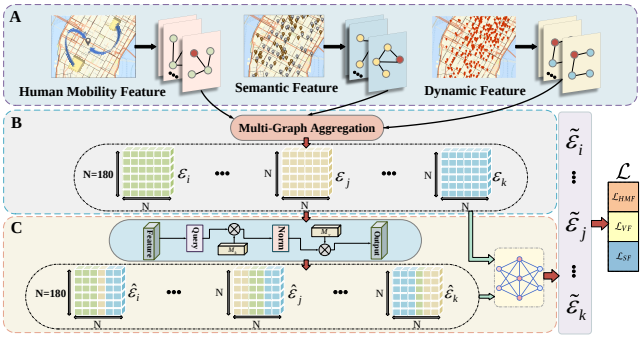

Figure 1: The architecture of ROMER consists of three key components: A) the region-wise graph learning module, B) the multi-graph aggregation module, and C) the attentive fusion module.

where $\mathcal{D}_O^{ij}$ denotes the dependencies between two origins, $\mathcal{D}_D^{ij}$ represents the dependencies between two destinations, $\text{sim}(\cdot)$ denotes the cosine similarity. Based on Equation (6) and (7), we construct region wise graphs $\mathcal{G}_O$ and $\mathcal{G}_D$.

3.1.2 Semantic Region Graph. In urban environments, the description of regional semantics relies on the utilization of Point of Interest (PoI) information. The PoI attributes encapsulate the semantic features associated with specific regions. To incorporate the PoI context into region embeddings, we leverage semantic dependencies to effectively capture and integrate region functionality information into the representation space. The calculation for this process can be described as follows:

$$\mathcal{D}_S^{ij} = \text{sim}\left(\vec{s}_i, \vec{s}_j\right). \tag{8}$$

where $\mathcal{D}_S^{ij}$ is the semantic dependency between region v_i and v_j. We thus obtain the semantic region graph $\mathcal{G}_S$.

3.1.3 Dynamic Region Graph. In contrast to PoI attributes, which solely provide information regarding the quantity of PoIs, check-in data takes into account human activity and reflects the popularity of each PoI category. When characterizing regions with check-in attributes, we employ a dynamic dependency measure to determine the significance of each check-in type within a given region. The calculation of this measure can be described as follows:

$$\mathcal{D}_G^{ij} = \text{sim}\left(\vec{a}_i, \vec{a}_j\right) \tag{9}$$

where $\mathcal{D}_G^{ij}$ is the dynamic dependency between region v_i and v_j. Now we construct dynamic region wise graph $\mathcal{G}_G$.

3.2 Multi-Graph Aggregation Module

It is observed that not only adjacent regions are relevant, but also many distant regions are correlated. However, existing GAT based methods [14, 27] only considers the influence of neighbouring nodes. Inspired by [9], we utilize an improved GAT mechanism to extract any two relevant regions in a city to assign learning weights. Given the vertex feature $\mathbf{h} = \left\{\vec{h}_1, \vec{h}_2, \ldots, \vec{h}_n\right\}, \vec{h}_i \in \mathbb{R}^F$, where F is

Table 1: Performance comparison of different approaches for check-in prediction and land usage classification.

Models	Check-in Prediction			Land Usage Classification	
	MAE	RMSE	R^2	NMI	ARI
LINE	564.59	853.82	0.08	0.17	0.01
node2vec	372.83	609.47	0.44	0.58	0.35
HDGE	399.28	536.27	0.57	0.59	0.29
ZE-Mob	360.71	592.92	0.47	0.61	0.39
MV-PN	476.14	784.25	0.08	0.38	0.16
MVGRE	297.72	495.27	0.63	0.78	0.59
MGFN	280.91	436.58	0.72	0.76	0.58
ROMER(ours)	**252.14**	**413.96**	**0.74**	**0.81**	**0.68**

the input dimension, the MGA layer works as follows:

$$A_{ij} = cosine(h_i, h_j) \cdot w_{ij} = \frac{h_i (h_j)^T \cdot w_{ij}}{\|h_i\| \|h_j\|}, \quad \hat{A_{ij}} = \sigma(A_{ij}), \tag{10}$$

where σ denotes the softmax function, w_{ij} is the weight matrix. A_{ij} is the similarity between h_i and h_j. Softmax function is used to normalize the coefficients. $\|h_i\|$ denotes the norm of vector h_i^l and $\cdot$ is the dot product of vectors.

Next, $\hat{A_{ij}}$ is utilized to aggregate information from all other features in the network to each feature, by

$$\hat{h_i} = \sum_{j \in \widetilde{N}_i} \hat{A_{ij}} h_j w_{ij}, \tag{11}$$

where $\widetilde{N}_i$ is all nodes in the graph except node h_i. $\hat{h_i}$ is the information aggregation from the global features to the feature h_i.

In our model, $\mathcal{G}_O, \mathcal{G}_D, \mathcal{G}_S$ and $\mathcal{G}_G$ are fed in to the MAG block, then we obtain the corresponding representation results as $\mathcal{E}_O, \mathcal{E}_D$, $\mathcal{E}_S$, and $\mathcal{E}_G$.

3.3 Attentive Fusion Module

In this section, we design attentive fusion mechanism to efficient interchange information among multiple views.Self-attention is widely used in existing fusion methods[18], which is at the price of high computation. Inspired by [8], We adopt external attention to allow information to propagate across multiple views. Given the representations of M views $\{\mathcal{E}_1, \mathcal{E}_2, \cdots, \mathcal{E}_M\}$, for each $\mathcal{E}_i$, we then propagate information among all views as follows:

$$[A_i]_{i=1}^M = \text{Norm}\left(\left[\mathcal{E}_i M_k^T\right]_{i=1}^M\right), \hat{\mathcal{E}}_i = \sum_{i=1}^M A_i M_v \tag{12}$$

where $M_k \in \mathbb{R}^{S \times d}$ and $M_v \in \mathbb{R}^{S \times d}$ are learnable parameters independent of the representations, which act as the key and the value memory of the whole training dataset. And d and S are hyper-parameters. $\hat{\mathcal{E}}_i$ is considered as the relevant global information for i-th view. Meanwhile the normalization method we follow [8]. The embedding results $\hat{\mathcal{E}}_O$, $\hat{\mathcal{E}}_D$, $\hat{\mathcal{E}}_S$, and $\hat{\mathcal{E}}_G$ are generated from the above modules, linking the global information to subsequent fusions in the model.

In order to integrate global and local region representation. We follow the fusion mechanism in [27]. The fusion layer operates as

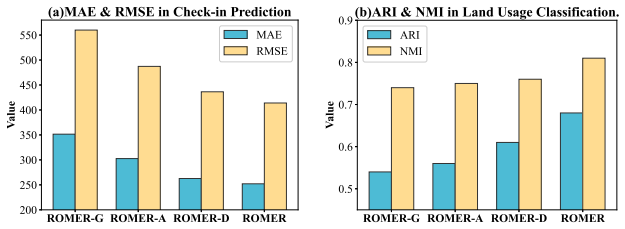

Figure 2: Ablation studies for two tasks on NYC dataset. (a) MAE and RMSE in Check-in Prediction. (b) ARI and NMI in Land Usage Classification.

follows:

$$\mathcal{E}_i' = \alpha \hat{\mathcal{E}}_i + (1 - \alpha)\mathcal{E}_i, \quad 0 \le \alpha \le 1 \tag{13}$$

$$\mathcal{E}_{\mathcal{F}} = \sum_i^M w_i \mathcal{E}_i, w_i = \sigma\left(\mathcal{E}_i W_f + b_f\right) \tag{14}$$

where $\mathcal{E}_i'$ is the representation for i-th view with global information, and α is the weight of global information, wi is the weight of i-th view, which is learned by a single layer MLP network with the i-th embeddings as input.

In a bid to enable the learning of the multi-view fusion layer, we engage $\mathcal{E}$ in the learning objective of each view. Formally, we update the representation of each view as:

$$\tilde{\mathcal{E}}_i = \left(\mathcal{E}_i' + \mathcal{E}_{\mathcal{F}}\right)/2. \tag{15}$$

By incorporating the outputs of the base model into proposed joint learning module, we derive region embeddings $\tilde{\mathcal{E}}_O, \tilde{\mathcal{E}}_D, \tilde{\mathcal{E}}_S$, and $\tilde{\mathcal{E}}_G$, on which we work out of the various learning goals.

3.4 Prediction Objectives

In order to efficiently train our model, according to [27], given the source region v_i, we model the distribution of the target region v_j as follows:

$$\hat{p}_O\left(v_j \mid v_i\right) = \frac{\exp\left(e_O^{iT} e_D^j\right)}{\sum_j \exp\left(e_O^i e_D^j\right)}. \tag{16}$$

Similarly, we model $\hat{p}_D\left(v_j \mid v_i\right)$ in the same way for the distribution of source region r_i for a given destination region r_j. Then, $\mathcal{L}_{HMF}$ is constructed by maximizing the probability of O/D occurrence. Hence, the $\mathcal{L}_{HMF}$ between region r_i and region r_j can be computed as:

$$\mathcal{L}_{HMF} = \sum_{(v_i, v_j) \in \mathcal{M}} -\log \hat{p}_o\left(v_j \mid v_i\right) - \log \hat{p}_d\left(v_i \mid v_j\right). \tag{17}$$

In order for the learned region embeddings to preserve the region similarity across region attributes, we designed tasks to reconstruct region correlations based on the corresponding embeddings. Taking the Check-in property as an example, the learning objective is based on C_G and $\tilde{\mathcal{E}}_G = \left\{e_G^i\right\}_{i=1}^n$ defined as follows.

$$\mathcal{L}_{VF} = \sum_{i,j}\left(C_G^{ij} - e_G^{iT} e_G^j\right)^2. \tag{18}$$

Similarly, we define the learning objective $\mathcal{L}_{SF}$ of check-in attributes. In this way, the final learning objective is:

$$\mathcal{L} = \mathcal{L}_{HMF} + \mathcal{L}_{VF} + \mathcal{L}_{SF}. \tag{19}$$

(a) Districts (b) HDGE (c) ZE-Mob

(d) MV-FN (e) MVGRE (f) Ours

Figure 3: Districts in Manhattan and region clusters.

4 EXPERIMENTS

In this section, extensive experiments are conducted to verify the superiority of the proposed model.

4.1 Datasets and Baselines

Experiments are conducted on several real-world datasets of New York City from NYC open data website[1]. We follow [27], taxi trip data was analysed as the source of human movement data, and the borough of Manhattan was divided into 180 zones to serve as the study area. As shown in Figure. 3(a), the borough of Manhattan was divided into 12 regions based on land use according to community board [2].

This paper compares the ROMER model with the following region representation methods, including LINE[16], node2vec[7], HDGE[22], ZE-Mob[25], MV-PN[5], MVGRE[27], MGFN[24] etc.

4.2 Experimental Settings

For land usage classification, we clustered the region embeddings using K-means with Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI) with settings [25]. In the case of check-in prediction, we utilize the Lasso regression [17] with metrics of Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and coefficient of determination (R^2).

4.3 Experimental Results

4.3.1 Main Results. Table. 1 shows the results of the check-in prediction task and the land usage classification task. We draw the following conclusions: (1) Our method (ROMER) outperforms all baseline tasks, in particular achieving over 10% improvement in MAE in the check-in prediction task and over 17% improvement in ARI in the land usage classification task. (2) Traditional graph embedding methods (LINE, node2vec) perform poorly because of the local sampling approach, which may not fully express the relationships between nodes. (3) While HDGE, ZE-Mob, and MV-PN utilize multi-scale graph structures and embedding methods to capture the multi-level features and complex relationships within the urban system, they lack an attention mechanism that accounts for the varying importance of nodes. (4) Both MVGRE and MGFN employ multi-view fusion methods and attention mechanisms. However, the long-range dependence of the regions mined by these models is poor.

4.3.2 Ablation Study. To verify the effect of key components on the proposed model, this paper conducts an ablation study in land usage classification and check-in prediction tasks respectively. The variants of ROMER are named as follows:

- ROMER-G: It is ROMER without multi-graph aggregation Module(MGA). MGA module is replaced with the GAT[19].
- ROMER-A: It is ROMER without Attentive Fusion which is replaced with the self-attention[18].
- ROMER-D: It is ROMER without Attentive graph fusion. The extracted spatial features are concatenated directly.

Figure. 2 displays the experimental results of ROMER and its variants in check-in prediction and land use classification tasks. The MGA module (ROMER-G) exhibits the most substantial impact on performance. Without the MGA component, the MAE for the prediction task increases significantly from 252.14 to 351.52, and the RMSE increases from 413.96 to 559.98. The attentive fusion module (ROMER-A) demonstrates the second-largest impact, affirming the effectiveness of our constructed module in enhancing single-view performance.

4.3.3 Visualized Analysis. To visually evaluate the clustering results, we plotted the clustering results of five methods in Figure. 3, where the same color marks the regions in the same cluster. We observe that the clustering results based on our method are optimal in terms of consistency with the real boundaries of ground conditions. These results suggest that the regional embeddings learned by our model are able to represent regional functions effectively.

5 CONCLUSION

In this paper, we try to solve the urban region embedding problem with attentive multi-view neural networks. Specially, we designed a graph aggregation module to captures region-wise dependencies within the urban network. To comprehensively share information across multiple views, we designed a attentive fusion module and fuse view embeddings with external attention and gating mechanism. Extensive experimental results on real-world data demonstrated the effectiveness of our proposed ROMER. In the future, we will apply the proposed frewmork to addition gragh based applications.

ACKNOWLEDGMENTS

This work was supported by the China Postdoctoral Science Foundation under Grant No.2022M711088.

[1] https://opendata.cityofnewyork.us/

REFERENCES

[1] Sami Abu-El-Haija, Bryan Perozzi, Rami Al-Rfou, and Alex Alemi. 2018. Watch Your Step: Learning Node Embeddings via Graph Attention. arXiv:1710.09599 [cs, stat]

[2] Bruce F Berg. 2007. New York City Politics : Governing Gotham. *Rutgers University Press* (2007), 352. https://doi.org/10.36019/9780813586694

[3] Ximing Chang, Jianjun Wu, Zhengbing He, Daqing Li, Huijun Sun, and Weiping Wang. 2020. Understanding User's Travel Behavior and City Region Functions from Station-Free Shared Bike Usage Data. *Transportation Research Part F: Traffic Psychology and Behaviour* 72 (July 2020), 81–95. https://doi.org/10.1016/j.trf.2020.03.018

[4] Peng Cui, Xiao Wang, Jian Pei, and Wenwu Zhu. 2017. A Survey on Network Embedding. arXiv:1711.08752 [cs]

[5] 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. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (July 2019), 906–913. https://doi.org/10.1609/aaai.v33i01.3301906

[6] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. 2017. Neural Message Passing for Quantum Chemistry. arXiv:1704.01212 [cs]

[7] Aditya Grover and Jure Leskovec. 2016. Node2vec: Scalable Feature Learning for Networks. arXiv:1607.00653 [cs, stat]

[8] Meng-Hao Guo, Zheng-Ning Liu, Tai-Jiang Mu, and Shi-Min Hu. 2021. Beyond Self-attention: External Attention Using Two Linear Layers for Visual Tasks. arXiv:2105.02358 [cs]

[9] Rongzhou Huang, Chuyin Huang, Yubao Liu, Genan Dai, and Weiyang Kong. 2020. LSGCN: Long Short-Term Traffic Prediction with Graph Convolutional Networks. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, Yokohama, Japan, 2355–2361. https://doi.org/10.24963/ijcai.2020/326

[10] Bo Hui, Da Yan, Wei-Shinn Ku, and Wenlu Wang. 2020. Predicting Economic Growth by Region Embedding: A Multigraph Convolutional Network Approach. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. ACM, Virtual Event Ireland, 555–564. https://doi.org/10.1145/3340531.3411882

[11] Neal Jean, Sherrie Wang, Anshul Samar, George Azzari, David Lobell, and Stefano Ermon. 2018. Tile2Vec: Unsupervised Representation Learning for Spatially Distributed Data. arXiv:1805.02855 [cs, stat]

[12] Thomas N. Kipf and Max Welling. 2016. Variational Graph Auto-Encoders. arXiv:1611.07308 [cs, stat]

[13] Hao Liu, Ting Li, Renjun Hu, Yanjie Fu, Jingjing Gu, and Hui Xiong. 2019. Joint Representation Learning for Multi-Modal Transportation Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (July 2019), 1036–1043. https://doi.org/10.1609/aaai.v33i01.33011036

[14] Y. Luo, F. Chung, and K. Chen. 2022. Urban Region Profiling via A Multi-Graph Representation Learning Framework. arXiv:2202.02074 [cs]

[15] Yan Luo, Fu-lai Chung, and Kai Chen. 2022. Urban Region Profiling via Multi-Graph Representation Learning. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management* (Atlanta, GA, USA) *(CIKM '22)*. Association for Computing Machinery, New York, NY, USA, 4294–4298. https://doi.org/10.1145/3511808.3557720

[16] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. LINE: Large-Scale Information Network Embedding. In *Proceedings of the 24th International Conference on World Wide Web* (Florence, Italy) *(WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1067–1077. https://doi.org/10.1145/2736277.2741093

[17] Robert Tibshirani. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 1 (1996), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x jstor:2346178

[18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. arXiv:1706.03762 [cs]

[19] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. arXiv:1710.10903 [cs, stat]

[20] Dongjie Wang, Kunpeng Liu, David Mohaisen, Pengyang Wang, Chang-Tien Lu, and Yanjie Fu. 2021. Automated Feature-Topic Pairing: Aligning Semantic and Embedding Spaces in Spatial Representation Learning. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems* (Beijing, China) *(SIGSPATIAL '21)*. Association for Computing Machinery, New York, NY, USA, 450–453. https://doi.org/10.1145/3474717.3484212

[21] Dongjie Wang, Kunpeng Liu, David Mohaisen, Pengyang Wang, Chang-Tien Lu, and Yanjie Fu. 2021. Towards Semantically-Rich Spatial Network Representation Learning via Automated Feature Topic Pairing. *Frontiers in Big Data* 4 (Oct. 2021), 762899. https://doi.org/10.3389/fdata.2021.762899

[22] 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*. ACM, Singapore Singapore, 237–246. https://doi.org/10.1145/3132847.3133006

[23] Zhecheng Wang, Haoyuan Li, and Ram Rajagopal. 2020. Urban2Vec: Incorporating Street View Imagery and POIs for Multi-Modal Urban Neighborhood Embedding. arXiv:2001.11101 [cs, stat]

[24] 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:2201.09760 [cs]

[25] 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*. International Joint Conferences on Artificial Intelligence Organization, Stockholm, Sweden, 3919–3925. https://doi.org/10.24963/ijcai.2018/545

[26] Chao Zhang, Keyang Zhang, Quan Yuan, Haoruo Peng, Yu Zheng, Tim Hanratty, Shaowen Wang, and Jiawei Han. 2017. Regions, Periods, Activities: Uncovering Urban Dynamics via Cross-Modal Representation Learning. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Perth Australia, 361–370. https://doi.org/10.1145/3038912.3052601

[27] Mingyang Zhang, Tong Li, Yong Li, and Pan Hui. 2020. Multi-View Joint Graph Representation Learning for Urban Region Embedding. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, Yokohama, Japan, 4431–4437. https://doi.org/10.24963/ijcai.2020/611