

# Automated Feature-Topic Pairing: Aligning Semantic and Embedding Spaces in Spatial Representation Learning

Dongjie Wang  
wangdongjie@Knights.ucf.edu  
University of Central Florida  
Orlando, Florida, United States

Kunpeng Liu  
kunpengliu@knights.ucf.edu  
University of Central Florida  
Orlando, Florida, United States

David Mohaisen  
mohaisen@ucf.edu  
University of Central Florida  
Orlando, Florida, United States

Pengyang Wang  
pywang@um.edu.mo  
University of Macau  
Macau, China

Chang-Tien Lu  
ctl@vt.edu  
Virginia Tech  
Virginia, United States

Yanjie Fu\*  
yanjie.fu@ucf.edu  
University of Central Florida  
Orlando, Florida, United States

## ABSTRACT

Automated characterization of spatial data is a kind of critical geographical intelligence. As an emerging technique for characterization, Spatial Representation Learning (SRL) uses deep neural networks (DNNs) to learn non-linear embedded features of spatial data for characterization. However, SRL extracts features by internal layers of DNNs, and thus suffers from lacking semantic labels. Texts of spatial entities, on the other hand, provide semantic understanding of latent feature labels, but is insensible to deep SRL models. How can we teach a SRL model to discover appropriate topic labels in texts and pair learned features with the labels? This paper formulates a new problem: feature-topic pairing, and proposes a novel Particle Swarm Optimization (PSO) based deep learning framework. Specifically, we formulate the feature-topic pairing problem into an automated alignment task between 1) a latent embedding feature space and 2) a textual semantic topic space. We decompose the alignment of the two spaces into : 1) point-wise alignment, denoting the correlation between a topic distribution and an embedding vector; 2) pair-wise alignment, denoting the consistency between a feature-feature similarity matrix and a topic-topic similarity matrix. We design a PSO based solver to simultaneously select an optimal set of topics and learn corresponding features based on the selected topics. We develop a closed loop algorithm to iterate between 1) minimizing losses of representation reconstruction and feature-topic alignment and 2) searching the best topics. Finally, we present extensive experiments to demonstrate the enhanced performance of our method.

## CCS CONCEPTS

• **Computing methodologies** → **Learning latent representations.**

\*Concat Author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
*SIGSPATIAL '21, November 2–5, 2021, Beijing, China*  
© 2021 Association for Computing Machinery.  
ACM ISBN 978-1-4503-8664-7/21/11...\$15.00  
<https://doi.org/10.1145/3474717.3484212>

## KEYWORDS

spatial representation learning, multiple spaces alignment

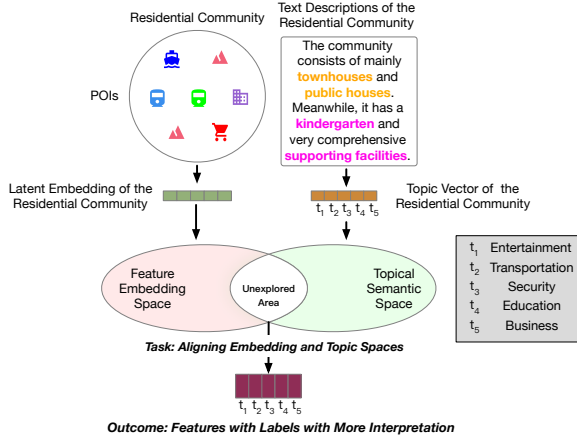
### ACM Reference Format:

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 *29th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '21)*, November 2–5, 2021, Beijing, China. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3474717.3484212>

## 1 INTRODUCTION

Spatial representation learning (SRL) refers to exploiting representation learning techniques to learn features of spatial network data, which has been successfully applied in many real-world scenarios, such as transportation networks, power networks, social networks, water supply networks [21]. In reality, many practical applications need to understand not just which features are effective, but also what these effective features stand for. This issue relates to two tasks: 1) deep representation learning; 2) label generation and matching for latent embedded features. Although there has been a rich body of work in SRL, including node embedding, autoencoder, random walk, adversarial learning, generative learning based methods with spatial data [14–17], research in unifying the two tasks is still in its early stage.

In response, we formulate the problem as a task of feature-topic pairing (**Figure 1**), which is to align a latent embedding feature space, consisting of multiple latent features, and a textual semantic topic space, consisting of multiple topic labels during SRL. The basic idea is to teach a machine to extract topic labels from texts, and then pair the labels with learned features. To that end, we propose to develop a novel deep learning framework to unify feature learning, topic selection, feature-topic matching. There are three unique challenges in addressing the problem: (1) Label Generation Challenge, in which a textual semantic topic space is difficult to construct due to the unstructured spatial texts; (2) Measurement Challenge, in which a promising measurement is highly desired to evaluate the alignment or quantify the matching score between the topic label space and the embedding feature space; (3) Optimization Challenge, in which a deep optimization framework is needed for to jointly and simultaneously unify the three tasks of feature learning, topic label selection, and feature-topic pairing.



**Figure 1: The motivation of the feature-topic pairing problem: bridging the gap between feature embedding space and topic semantic space in representation learning.**

To solve the three challenges, we develop a new PSO-based framework (named AutoFTP) that enclose the optimizations of feature learning, topic selection, and feature-topic pairing in a loop. Specifically, our contributions are: (1) formulating the feature-topic pairing problem for relieving the scarcity of semantic labels; (2) proposing a three-step method for generating candidate topic labels; (3) deriving a feature-topic alignment measurement by point-wise alignment between an embedding feature vector and a categorical topic distribution, and pair-wise alignment for the consistency of feature-feature similarity matrix and topic-topic similarity matrix; (4) developing a Particle Swarm Optimization (PSO)-based algorithm for unified optimization.

## 2 PROPOSED METHOD

### 2.1 The Feature-Topic Pairing Problem

The feature-topic pairing problem aims to pair the latent features extracted by representation learning, with the explicit topics of texts of a spatial entity. Formally, given a set of  $N$  spatial entities, the  $n$ -th entity is described by multiple graphs (e.g., a POI-POI distance graph  $G_n^d$  and a POI mobility connectivity  $G_n^m$ , defined in Section 3.3) and a topic distribution  $\mathbf{t}_n$  extracted from textual descriptions  $\mathcal{E}_n$ . Let  $\tilde{\mathbf{r}}_n$  be the embedding vector of the  $n$ -th entity. The objective is to optimize a function that measures representation loss and feature-topic alignment:

$$\operatorname{argmax}_{\tilde{\mathbf{R}}} \sum_{n=1}^N f(\tilde{\mathbf{r}}_n | \mathbf{t}_n, G_n^d, G_n^m, K), \quad (1)$$

where  $\tilde{\mathbf{R}} = \{\tilde{\mathbf{r}}_n\}_{n=1}^N \in \mathbb{R}^{N \times K}$  are the embeddings of all spatial entities,  $K$  is the number features of an embedding vector.

### 2.2 Preprocessing

**Textual Topic Extraction.** We employ a pre-trained deep word embedding model [4] to generate topics. Specifically, we first collect the text descriptions of all entities. Besides, we extract keywords from texts using the TextRank algorithm [10] and leverage a pre-trained language model [4] to learn the corresponding word embedding of each keyword. Moreover, we exploit a Gaussian Mixture Model (GMM) to cluster the keyword embeddings into  $T$  topics. The clustering model provides a topic label for each keyword.

**Embedding of Spatial Entities.** We construct a graph to capture the spatial autocorrelation between spatial entities. Specifically, we describe a spatial entity in terms of its POIs, by building two graphs. (i) **POI-POI distance graph:** denoted by  $G^d$ , where POI categories are nodes and the average distances between POI categories are edge weights. (ii) **POI-POI mobility graph:** denoted by  $G^m$ , where nodes are POI categories, and edge weights are human mobility connectivity, which is extracted by the method in [16]. We then apply Graph Auto Encoder (GAE) [7] as the spatial representation learner to learn spatial embeddings over these two constructed graphs respectively. Finally, we aggregate the embeddings of these two graphs by averaging, so as to construct the unified spatial embedding of the entity, denoted by  $\mathbf{r}_n \in \mathbb{R}^K$ .

### 2.3 PSO Based Feature-Topic Pairing

#### 2.4.1 Measuring the Alignment of Embedding and Semantic Spaces.

To pair features with topics, we conduct space alignment from the point-wise and pair-wise perspectives, with considering the alignment of the coordinate system and information contents respectively. To be convenient, we take the  $n$ -th entity as an example to explain the calculation process.

1) **Point-wise Alignment Loss:**  $\mathcal{L}_P$ . Intuitively, the embedding feature of the spatial entity and corresponding topic should reach a consensus on describing an spatial entity, thus correlations are expected to be maximized between them. Therefore, we first select  $K$  values from the topic vector  $\mathbf{t}_n$  as the vector  $\check{\mathbf{t}}_n \in \mathbb{R}^K$ , which contains the most representative semantics in the semantic space. Then, we maximize the correlation between  $\check{\mathbf{t}}_n$  and the spatial embedding  $\mathbf{r}_n$ , which is equal to minimize the negative correlation between the two vectors. The formula of the minimizing process as follows:

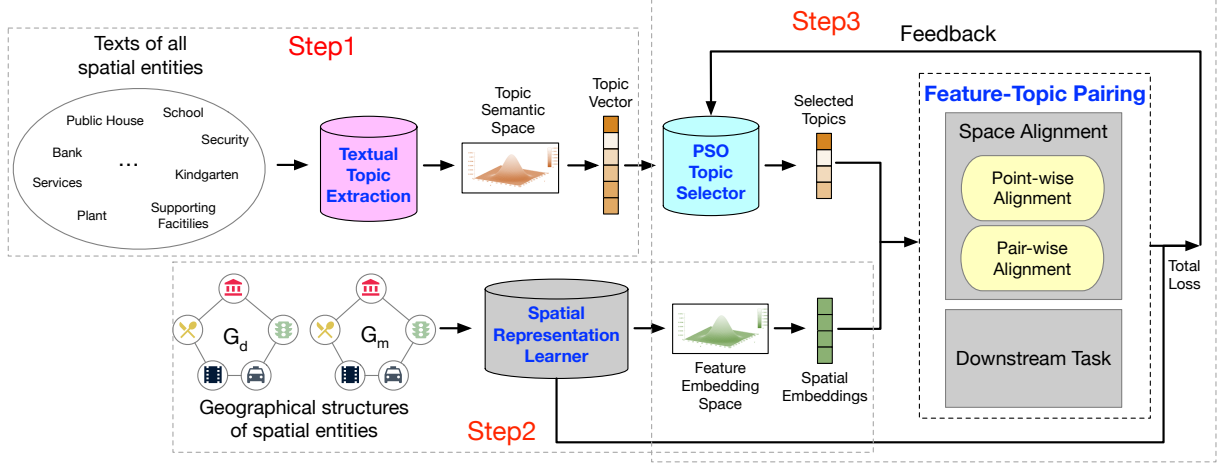
$$\mathcal{L}_P = - \sum_{n=1}^N \frac{\operatorname{cov}(\check{\mathbf{t}}_n, \mathbf{r}_n)}{\delta(\check{\mathbf{t}}_n)\delta(\mathbf{r}_n)}, \quad (2)$$

where  $\operatorname{cov}(\cdot)$  denotes the covariance calculation;  $\delta(\cdot)$  denotes the standard deviation.

2) **Pair-wise Alignment Loss:**  $\mathcal{L}_C$ . On the other hand, the embedding feature and the corresponding topic should show consistency on the pair-wise similarity in each space to reflect the pair-wise alignment. Therefore, we minimize the difference between the pair-wise similarity between these two spaces. Specifically, we first construct the topic-topic similarity matrix  $\mathbf{S}$  and the feature-feature similarity matrix  $\mathbf{S}'$ . Specifically, for  $\mathbf{S} \in \mathbb{R}^{K \times K}$ , we calculate the similarity between any two topics. For  $\mathbf{S}' \in \mathbb{R}^{K \times K}$ , we calculate the similarity between two features of spatial embeddings. We keep the pair-wise consistency between  $\mathbf{S}$  and  $\mathbf{S}'$  by minimizing the Frobenius norm, as follows:

$$\mathcal{L}_C = \|\mathbf{S} - \mathbf{S}'\|_F. \quad (3)$$

**2.4.2 Supervised PSO For Automatic Topic Selection.** To select best  $K$  topics for feature-topic alignment, we propose to formulate the joint task of feature learning, topic selection, topic and feature pairing into a PSO problem. Specifically, we first randomly initialize a number of particles in PSO, where a particle is a binary topic mask (i.e., the mask value of 1 indicates “select” and the mask value of 0 indicates “deselect”). In other words, a set of particles select a subset of topics. A multi-objective deep learning model, whose



**Figure 2: An overview of AutoFTP.** In the framework, we first construct a topic semantic space based on the texts of spatial entities. Then, we initialize an embedding feature space based on the geographical structures of spatial entities. Later, we employ a PSO-based framework to conduct feature-topic pairing through jointly optimizing representation learning, point-wise alignment, pair-wise alignment, and downstream task over learning iterations.

objective function includes the losses of graph reconstruction, semantic alignment, and the regression estimator in the downstream task, is trained to learn spatial representations, using each selected topic subset. As an application, we use the embedding of spatial entities (residential communities) to predict their real estate prices, and the loss of the regression model  $\mathcal{L}_{Reg}$  is:

$$\mathcal{L}_{Reg} = \frac{1}{N} \sum_{n=1}^N (c_n - c_n^*)^2, \quad (4)$$

where  $c_n$  is the golden standard real estate price and  $c_n^*$  is the predicted price. Next, we calculate the fitness of each particle according to the total loss of the deep model. The fitness can be calculated by:

$$Fitness = \mathcal{L}_C + \mathcal{L}_P + \mathcal{L}_R + \mathcal{L}_{Reg}. \quad (5)$$

Then, we utilize the fitness to inform all particles how far they are from the best solution. Next, each particle moves forward to the solution based on not only its current status but also all particles' movement. After the fitness value of PSO converges, PSO identifies the best topic subset. Finally, the semantically-rich embeddings of spatial entities, given by:  $\tilde{\mathbf{R}} = \{\tilde{\mathbf{r}}_n\}_{n=1}^N$ .

## 3 EXPERIMENTAL RESULTS

### 3.1 Evaluation Task

In this work, we apply the proposed AutoFTP to the price prediction of real-estate as the evaluation task. Specifically, we first apply AutoFTP to learn a series of representations of spatial entities based on their geographical structural information and related text descriptions. Then, we build up a deep neural network (DNN) model for predicting average real estate price of each spatial entity according to its corresponding representation. We use RMSE, MAE, MAPE and MLSE as the evaluation metric.

### 3.2 Data Description

Table 2 shows the statistics of five data sources used in the experiments. Specifically, the taxi traces data describes the GPS trajectory of taxis in Beijing in three months; the residential regions, texts,

and real estate price data sources are crawled from [www.fang.com](http://www.fang.com); and the POIs information are extracted from [www.dianping.com](http://www.dianping.com).

**3.2.1 Baseline Algorithms.** We compared our proposed method with seven baseline algorithms: **AttentionWalk** [1], **ProNE** [22], **GatNE** [2], **GAE** [7], **DeepWalk** [11], **Node2Vec** [3], and **Struc2Vec** [12]. Besides, regarding the are four losses in AutoFTP: reconstruction loss  $\mathcal{L}_R$ , point-wise alignment loss  $\mathcal{L}_P$ , pair-wise alignment loss  $\mathcal{L}_C$ , and regression loss  $\mathcal{L}_{Reg}$ , we also derive four variants: (ii) **AutoFTP<sup>(R+P)</sup>**, which keeps  $\mathcal{L}_R$  and  $\mathcal{L}_P$  of AutoFTP; (iii) **AutoFTP<sup>(R+C)</sup>**, which keeps  $\mathcal{L}_R$  and  $\mathcal{L}_C$  of AutoFTP; (iv) **AutoFTP<sup>(R+P+C)</sup>**, which keeps  $\mathcal{L}_R$ ,  $\mathcal{L}_P$ , and  $\mathcal{L}_C$  of AutoFTP.

### 3.3 Overall Performance

Table 1 shows the comparison of all the 11 models. As can be seen, AutoFTP, in overall, outperforms the baseline algorithms in terms of RMSE, MAE, MAPE and MSLE. A possible reason for this observation is that compared with other baseline algorithms, AutoFTP not just captures geographical structural information but also preserves rich semantics of spatial entity. Besides, the regression estimator (the downstream task) of AutoFTP provides a clear learning direction (accuracy) for spatial representation learning. Thus, in the downstream predictive task, the spatial embedding features learned by AutoFTP beats all baselines.

## 4 RELATED WORK

**Graph Representation Learning with Latent Semantics.** Graph representation learning refers to techniques that preserve the structural information of a graph into a low-dimensional vector [13, 18]. However, owing to traditional graph representation learning models are implemented by deep neural networks, the learned embeddings lack interpretability. Recently, to overcome this limitation, researchers leveraged the texts related to graphs to learn semantically rich representations [9, 19].

**Topic Models in Spatio-temporal Domain.** Topic models aim to automatically cluster words and expressions patterns for characterizing documents [8, 20]. Recently, to understand the hidden semantics of spatial entities, many researchers applied topic models

**Table 1: Overall Performance with respect to RMSE, MAE, MAPE and MSLE. (The smaller value is, the better performance is)**

	RMSE	Outperform	MAE	Outperform	MAPE	Outperform	MSLE	Outperform
<b>AutoFTP</b>	<b>18.646</b>	-	<b>16.192</b>	-	<b>58.851</b>	-	<b>0.2267</b>	-
AttentionWalk	21.418	+14.9%	19.712	+21.7%	68.590	+16.6%	0.2907	+28.2%
ProNE	21.830	+17.1%	19.929	+23.1%	69.188	+17.6%	0.2949	+30.1%
GatNE	21.229	+13.9%	19.288	+19.1%	67.043	+13.9%	0.2854	+25.9%
GAE	21.338	+14.4%	19.676	+21.5%	68.579	+16.5%	0.2894	+27.6%
DeepWalk	23.561	+26.4%	21.987	+35.8%	76.038	+29.2%	0.3321	+46.5%
Node2Vec	22.688	+21.7%	21.084	+30.2%	73.135	+24.3%	0.3152	+39.0%
Struc2Vec	21.589	+15.8%	19.937	+23.1%	69.423	+17.9%	0.2942	+29.7%
AutoFTP <sup>R</sup>	21.965	+17.8%	20.283	+25.3%	70.991	+20.6%	0.2928	+29.1%
AutoFTP <sup>(R+P)</sup>	20.509	+9.99%	18.921	+16.8%	66.477	+12.9%	0.2681	+18.3%
AutoFTP <sup>(R+C)</sup>	21.014	+12.7%	19.413	+19.8%	67.920	+15.4%	0.2773	+22.3%
AutoFTP <sup>(R+P+C)</sup>	20.211	+8.39%	18.676	+15.3%	65.685	+11.6%	0.2636	+16.3%

**Table 2: Statistics of the Experimental Data**

Data Sources	Properties	Statistics
Taxi Traces	Number of taxis	13,597
	Time period	Apr. - Aug. 2012
Residential Regions	Number of residential regions	2,990
	Time period of transactions	04/2011 - 09/2012
POIs	Number of POIs	328668
	Number of POI categories	20
Texts	Number of textual descriptions	2,990
	Time Period	04/2011 - 09/2012
Real Estate Prices	Number of real estate prices	41,753
	Time Period	12/2011 - 06/2012

in the spatio-temporal data mining domain [5, 6]. Thus, in this paper, we employ a pre-trained language model to get the embeddings of keywords and utilize Gaussian Mixture Model to extract topic distribution based on the embeddings.

## 5 CONCLUSION

We presented a novel spatial representation learning (SRL) framework, namely AutoFTP. The spatial embeddings produced by traditional SRL models lack semantic meaning. To overcome this limitation, we formulated the feature-topic pairing problem. We proposed a novel deep learning framework to unify representation learning, topic label selection, and feature-topic pairing through a PSO-based optimization algorithm. Extensive experiments demonstrated the effectiveness of AutoFTP by comparing it with other baseline models. For future work, we plan to extend our approach from geospatial networks to other applications that consist of graphs and texts, such as social media and software code safety.

## ACKNOWLEDGMENTS

This research was partially supported by the National Science Foundation (NSF) via the grant numbers: 1755946, 2040950, 2006889, 2045567, 2141095.

## REFERENCES

- [1] Sami Abu-El-Hajja, Bryan Perozzi, Rami Al-Rfou, and Alexander A Alemi. 2018. Watch Your Step: Learning Node Embeddings via Graph Attention. *Advances in Neural Information Processing Systems* 31 (2018), 9180–9190.
- [2] Yukuo Cen, Xu Zou, Jianwei Zhang, Hongxia Yang, Jingren Zhou, and Jie Tang. 2019. Representation learning for attributed multiplex heterogeneous network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1358–1368.
- [3] 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.
- [4] Han He. 2014. *HanLP: Han Language Processing*. <https://github.com/hankcs/HanLP>
- [5] Liang Huang, Yuanqiao Wen, Wei Guo, Xinyan Zhu, Chunhui Zhou, Fan Zhang, and Man Zhu. 2020. Mobility pattern analysis of ship trajectories based on semantic transformation and topic model. *Ocean Engineering* 201 (2020), 107092.
- [6] Qionghao Huang, Changqin Huang, Jin Huang, and Hamido Fujita. 2019. Adaptive resource prefetching with spatial-temporal and topic information for educational cloud storage systems. *Knowledge-Based Systems* 181 (2019), 104791.
- [7] Thomas N Kipf and Max Welling. 2016. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308* (2016).
- [8] Hakyoon Lee and Pilsung Kang. 2018. Identifying core topics in technology and innovation management studies: A topic model approach. *The Journal of Technology Transfer* 43, 5 (2018), 1291–1317.
- [9] Gengchen Mai, Krzysztof Janowicz, and Bo Yan. 2018. Combining Text Embedding and Knowledge Graph Embedding Techniques for Academic Search Engines. In *Semdeep/NLIWoD@ ISWC*. 77–88.
- [10] Rada Mihalcea and Paul Tarau. 2004. Texttrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*.
- [11] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 701–710.
- [12] Leonardo FR Ribeiro, Pedro HP Saverese, and Daniel R Figueiredo. 2017. struc2vec: Learning node representations from structural identity. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 385–394.
- [13] Daixin Wang, Peng Cui, and Wenwu Zhu. 2016. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1225–1234.
- [14] Dongjie Wang, Pengyang Wang, Jingbo Zhou, Leilei Sun, Bowen Du, and Yanjie Fu. 2020. Defending Water Treatment Networks: Exploiting Spatio-temporal Effects for Cyber Attack Detection. In *2020 IEEE International Conference on Data Mining (ICDM)*. IEEE, 32–41.
- [15] 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.
- [16] Pengyang Wang, Yanjie Fu, Jiawei Zhang, Xiaolin Li, and Dan Lin. 2018. Learning urban community structures: A collective embedding perspective with periodic spatial-temporal mobility graphs. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9, 6 (2018), 1–28.
- [17] Pengyang Wang, Yanjie Fu, Jiawei Zhang, Pengfei Wang, Yu Zheng, and Charu Aggarwal. 2018. You are how you drive: Peer and temporal-aware representation learning for driving behavior analysis. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2457–2466.
- [18] Pengyang Wang, Yanjie Fu, Yuanchun Zhou, Kunpeng Liu, Xiaolin Li, and Kien A Hua. 2020. Exploiting Mutual Information for Substructure-aware Graph Representation Learning. In *IJCAI*. 3415–3421.
- [19] Han Xiao, Minlie Huang, Lian Meng, and Xiaoyan Zhu. 2017. SSP: semantic space projection for knowledge graph embedding with text descriptions. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- [20] Guangxu Xun, Yaliang Li, Wayne Xin Zhao, Jing Gao, and Aidong Zhang. 2017. A Correlated Topic Model Using Word Embeddings. In *IJCAI*. 4207–4213.
- [21] Daokun Zhang, Jie Yin, Xingquan Zhu, and Chengqi Zhang. 2018. Network representation learning: A survey. *IEEE transactions on Big Data* 6, 1 (2018), 3–28.
- [22] Jie Zhang, Yuxiao Dong, Yan Wang, Jie Tang, and Ming Ding. 2019. ProNE: Fast and Scalable Network Representation Learning. In *IJCAI*, Vol. 19. 4278–4284.