THAN: Multimodal Transportation Recommendation With Heterogeneous Graph Attention Networks

Aikun Xu, Ping Zhong, *Member, IEEE*, Yilin Kang, Jiongqiang Duan, Anning Wang, Mingming Lu, *Member, IEEE*, and Chuan Shi, *Member, IEEE*

***Abstract*— Multi-modal transportation recommendation plays an important role in navigation applications. It aims to recommend a travel plan with various transport modes, such as bus, metro, taxi, bicycle, and a hybrid. Analysis of real-world large-scale navigation data shows that the correlation between the data can be represented by a graph containing different types of nodes and edges. As an emerging technology, graph neural networks (GNN) have shown powerful capabilities in representing graph data. However, existing solutions based on GNN only consider converting heterogeneous graph data into homogeneous graph data, ignoring the effects of different types of nodes and edges. In addition, those methods usually face the over-smoothing problem, which reduces the accuracy of recommendation. To this end, we propose a multi-modal *T*ransportation recommendation algorithm with *H*eterogeneous graph *A*ttention *N*etworks (THAN) based on carefully constructed heterogeneous graphs. We first design a novel graph embedding method to represent the correlation between the origin and the destination, as well as the correlation between origin-destination (OD) pairs and users. Next, a heterogeneous graph from large-scale data is built to describe the relationship between users, OD pairs, and transport modes. Then, we design a hierarchical attention mechanism with residual blocks to generate node embedding in terms of homogeneity and heterogeneity. Finally, a fusion neural layer is designed to fuse embeddings from different views and predict the proper transport mode for users. Extensive experimental results on a large-scale real-world dataset demonstrate that the performance of THAN outperforms five baselines.**

***Index Terms*— Multimodal transportation, recommendation system, graph neural network, attention, heterogeneous.**

Manuscript received 15 July 2021; revised 26 January 2022, 23 May 2022, and 8 September 2022; accepted 1 November 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62272489, Grant 62172443, and Grant 62172442; in part by the National Science Foundation of Hunan Province under Grant 2022JJ30760; in part by the Natural Science Foundation of Changsha, China, under Grant kq2202107; and in part by the Fundamental Research Funds for the Central Universities of Central South University, China, under Grant 2021zzts0735 and Grant 2021zzts0753. The Associate Editor for this article was L. Du. *(Corresponding author: Ping Zhong.)*

Aikun Xu, Ping Zhong, Jiongqiang Duan, Anning Wang, and Mingming Lu are with the School of Computer Science and Engineering, Central South University, Changsha 410073, China (e-mail: aikunxu@csu.edu.cn; ping.zhong@csu.edu.cn; 8208191025@csu.edu.cn; anningwang@csu.edu.cn; mingminglu@csu.edu.cn).

Yilin Kang is with the School of Computer Science, South Central University for Nationalities, Wuhan 430074, China (e-mail: ylkang@mail.scuec.edu.cn).

Chuan Shi is with the School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: shichuan@bupt.edu.cn).

Digital Object Identifier 10.1109/TITS.2022.3221370

I. Introduction

As THE living standards improve significantly, traveling has become a necessity on a daily basis for most people. Navigation application plays an important role in traveling, and thus it has penetrated deeply into people's daily life. Transportation recommendation is one of the core components of many navigation applications, and the research on it attracts the attention of academia [1], [2], [3], [4], [5], [6], [7] and industry [8], [9], [10], [11], [12], *e.g.,* Google Map,[1] Amap,[2] Baidu Map,[3] and Didi[4] are trying to design and develop safer, more coordinated, and more efficient intelligent transportation systems to promote social development and meet the diversified needs of citizens. People usually choose to use transport modes such as walking, biking, driving, and public transportation to complete the trip from the origin (O) to the destination (D). In fact, the navigation application will only make a route plan for the user considering one of the above transport modes, *i.e.*, uni-modal transportation recommendation. For trips with a given transport mode, it is quite easy to plan the travel path. However, there is still a lack of research to systematically solve the problem of choosing the most suitable transport mode for users from multiple transport modes, which is called multi-modal transportation recommendation.

Uni-modal transportation recommendation that only considers one transport mode, *e.g.*, bicycle, bus, taxi, or subway, which has been widely studied [10], [13], [14]. In contrast, research on multi-modal transportation recommendation that considers both uni-modal (*e.g.*, bicycle, bus, taxi, subway) and multi-modal (*e.g.*, taxi-bus, taxi-subway) in the road network is still in the preliminary exploration stage [3], [8], [9], [11], [15], [16]. The uni-modal transportation recommendation methods cannot be directly applied to multi-modal transportation recommendation. To this end, Liu et al. [8] proposed Trans2vec, which is a multi-modal transportation recommendation method based on graph embedding. Based on the Trans2vec, a multi-modal transportation recommendation system was formally established and deployed on Baidu Maps.

[1] https://www.google.com/maps
[2] https://amap.com/
[3] https://map.baidu.com/
[4] https://www.didiglobal.com/

From 2019 to 2020, the multi-modal transportation recommendation system provided hundreds of millions of users with more personalized and intelligent navigation services. While meeting the diverse needs of users, it can also reduce travel time, balance traffic flow, and reduce traffic congestion, thus greatly promoting the development of smart transportation systems [8], [9].

Although the multi-modal transportation recommendation system has been deployed for commercial use, the multi-modal transportation recommendation technology still faces many problems and challenges. The first challenge is the lack of a real multi-modal transportation recommendation open dataset as a test benchmark. As far as we know, real data is essential for deep learning (*e.g.*, graph neural networks). The data related to the transportation recommendation field is either non-open data or a single modal data (*i.e.*, it contains only one transport mode) [10], [14], which is difficult to support the research of multi-modal transportation recommendation. Therefore, it is necessary to collect and open a real multi-modal transportation dataset containing multiple transport modes. The second challenge is the heterogeneity of the transportation graph. The transportation graph is heterogeneous in nature, which contains many types of nodes and edges as well as rich semantics. One research issue is: how to represent nodes and utilize the rich semantics of transportation graphs, *i.e.*, how to mine potential relationships between different types of nodes and edges. The last challenge is the over-smoothing problem. When using a graph neural network, as the number of network layers and iterations increases, eventually all nodes converge to the same value, resulting in a decrease in the model representation ability.

We collected and analyzed the Beijing dataset[5] from Baidu Maps and found some interesting results. Specifically, as shown in Figure 1, there are eleven transport modes in total, including uni-modal and multi-modal, namely bus, metro, car, taxi, walk, bicycle, bus-metro, taxi-bus, bicycle-metro, taxi-metro, and bicycle-bus-metro. We encode these transport modes into numerical labels ranging from 1 to 11. The result shows that a mixed-mode of transportation is as important as a single mode of transportation in transportation recommendation. Hence, considering multi-modal in transportation recommendation is well worth studying. In addition, albeit 89.1% of the existing recommendation results are suitable for users [9], more than 32.5% of the recommendation results in the recommendation list are not clicked by users, as shown in Figure 2. This result shows that the improvement of the current transportation recommendation scheme is an urgent task.

In order to solve the above challenges, make full use of uni-modal modes of transportation and multi-modal as well as increase the click-through rate, we collect a large-scale real dataset and propose a novel framework called *multi-modal **T**ransportation recommendation with **H**eterogeneous graph **A**ttention **N**etworks* (**THAN**). The main contributions of our work are as follows:

- We propose THAN, a multi-modal transportation recommendation system. To the best of our knowledge, this is the first solution to implementing multi-modal transportation recommendation with heterogeneous graph attention networks.
- To obtain the features of nodes, we propose a novel latent feature mining method, named the Multi-Bigraph Embedding method (MBigraphE), which can represent the relationship between the origin and destination as well as the relationship between OD pairs and users.
- To utilize the rich semantics of transportation graphs, we design a hierarchical attention mechanism with residual blocks. It can represent features from the perspective of homogeneity, obtain potential features from a heterogeneous perspective, and ensure the network can be trained stably.
- We collect and open a large-scale real multi-modal transportation recommendation dataset. THAN conducts extensive experiments on this dataset. The results show that our scheme outperforms five state-of-the-art algorithms in four metrics.

The paper is organized as follows. Section II reviews some related works. Section III presents the preliminaries. Section IV introduces THAN. Section V evaluates the performance of the proposed solutions and Section VI concludes the paper.

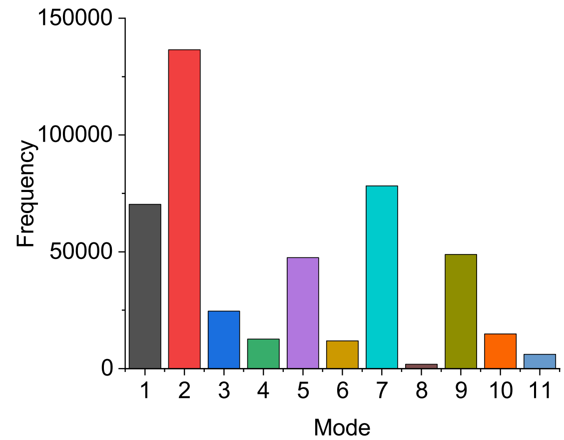


Fig. 1. Statistics on the usage of various modes of transportation.

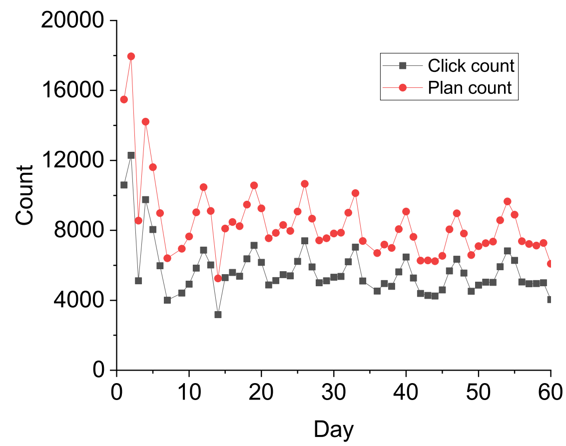


Fig. 2. Count of Beijing.

[5]https://dianshi.bce.baidu.com/competition/29/rule

II. Related Works

A. Uni-Modal Transportation Recommendation

Transportation recommendation engines usually use the shortest route query algorithm [17] with a predetermined cost function [18] to achieve path planning, which aims to recommend the most cost-effective (*e.g.*, shortest travel time, lowest price, least passing traffic lights, or shortest distance traveled) routes. It is difficult to meet the differentiated needs of users only with a fixed cost-effective function as the goal of path planning. To meet the personalized needs of different users, reference [19] takes the personalization factor into account in route planning for the first time and uses it to improve the quality of transportation recommendation. Specifically, reference [13] recommends travel routes by considering each driver's personal preferences (*e.g.*, drivers care more about time efficiency or fuel efficiency). Some schemes that consider multiple objectives based on large-scale historical trajectories are proposed to improve the quality of recommended paths [20], [21]. However, with the continuous expansion of the transportation network and the complexity of travel conditions (*i.e.*, large-scale trajectories), it is difficult for people to find the optimal route from one place to another through the transportation system, *e.g.*, the public transportation system [22]. Machine learning shows huge advantages in mining large-scale trajectories. Recently, a data-driven machine learning engine Polestar is proposed to recommend intelligent and efficient public transportation routes [10], which ignore people's mobility patterns [14], [23]. Moreover, other shared transports are considered in some works to improve the efficiency and benefits of roads have also attracted widespread attention [24], [25], *e.g.*, bike-sharing [26]. However, the above-mentioned methods only consider one transport mode, and cannot be directly used to solve the problem of multi-modal transportation recommendation.

B. Multi-Modal Transportation Recommendation

The multi-modal transportation recommendation system considers both single (*e.g.*, bus, metro, taxi) and mixed (*e.g.*, bus-metro, bus-taxi, metro-taxi) transport modes, which attract attention from academia [24], [27] and industry [8], [9], [16]. For example, a personalized route recommendation method FAVOUR is proposed to solve the problem of multi-modal transportation recommendation. However, this method requires a large amount of user privacy information to operate, it is difficult to implement [27]. Trans2vec [8] realizes multi-modal transportation recommendation by learning the embedding between users, OD pairs, and transport modes. However, this solution is plagued by the cold-start problem and requires additional models or strategies to deal with new examples. To this end, a fusion network embedding model data-driven scheme Hydra is proposed for multi-modal route recommendation [9], which ignores the significance of bike-sharing as a convenient transport mode. HMTRL [16] introduces an attention mechanism into multi-modal transportation recommendation, which aims to mine potential relationships between nodes of the same type in traffic data, thereby improving the accuracy of transportation recommendation. However, HMTRL cannot capture the relationships between different types of nodes and edges. NMTRec [3] builds a bipartite graph based on user trajectory data to describe the network structure between the user and OD pairs as well as origin and destination, which uses Word2Vec [28] to mine the relationship between nodes. The LightGBM [29] is introduced in NMTRec to predict the scores of transport modes, and complete the recommendation by ranking the scores. MTRecS-DLT [30] fuses a convolutional neural network and gradient ascent decision tree model as an integrated learning model to learn effective features in data. This model takes into account the advantages of the neural network model and machine learning model to obtain ideal data representation and realizes multi-model transportation recommendation. TTCA considering both taxi and bike-sharing takes less time than considering taxi or bike-sharing alone [24]. Some researchers considered combining traditional optimization algorithms with machine learning to improve the accuracy of multi-modal transportation recommendation. Specifically, they proposed a context-aware multi-modal transportation recommendation based on particle swarm optimization and LightGBM [31] as well as a multi-modal transportation recommendation scheme based on graph embedding and CaGBDT [32], respectively. Existing solutions are implemented based on machine learning and cannot handle the datasets used in our work (*i.e.*, graph-structured data). Instead, graph neural network-based schemes are suitable.

C. Graph Neural Network

Almost all natural data can be represented by a graph structure [33]. Graph neural network (GNN) has been proved to be able to effectively represent graph structure data [34], which aims to extend the deep neural network to process arbitrary graph structure data [35]. Recently, the emergence of graph convolutional neural networks (GCN) has promoted the development of GNN [36], [37]. Compared with GNN, GCN performs generalized convolution operation on graph structure data [38]. GCN is usually divided into two categories, namely spectral-domain [36], [39], [40] and non-spectral domain [41], [42]. For example, Kipf and Welling [36] proposes a spectral method called graph convolutional network, which designs graph convolutional networks through the local first-order approximation of spectral graph convolution. Hamilton et al. [41] designs GraphSAGE, which implements a neural network-based aggregator on fixed-size node neighbors. It can learn a function to generate embeddings by aggregating features from the local neighborhood of the node. However, neither GCN nor GraphSAGE can distinguish the importance of neighbor nodes. To solve the above problems, the attention mechanism is proposed in [43] and [42], which has been used in many studies, *e.g.*, the recommendation [44] and finance [45]. Inspired by the above schemes, HMTRL [16] is proposed to improve the accuracy of transportation recommendation. However, the above-mentioned works can only be applied to homogeneous graphs and cannot handle various types of nodes and edges (*i.e.*, heterogeneous graphs).

TABLE I
EXAMPLE OF INTERACTION BETWEEN USERS AND BAIDU MAP

sid	pid	req_time	origin (O)	destination (D)	plan_time	plans: distance (m), price (RMB $cent$), eta (s), transport_mode	click_time	click_mode
387056	234590	2018-11-01 15:15:36	116.30,40.05	116.35,39.99	2018-11-01 15:15:40	plan1: 14541, 600, 4207, 7 plan2: 9697, -, 1822, 3 plan3: 9697, 3900, 1492, 4	2018-11-01 15:15:47	7
902489	849336	2019-01-16 19:57:41	117.33,39.08	117.32, 39.09	2019-01-16 19:57:44	plan1: 609, -, 560, 5, plan2: 857, -, 255, 6 plan3: 1839, -, 421, 3	2019-01-16 19:57:46	5
156976	221455	2018-12-17 09:05:12	121.48,31.21	121.44,31.11	2018-12-17 09:05:15	plan1: 6577, 300, 680, 2 plan2: 9506, 200, 3549, 1	2018-12-17 09:05:30	2

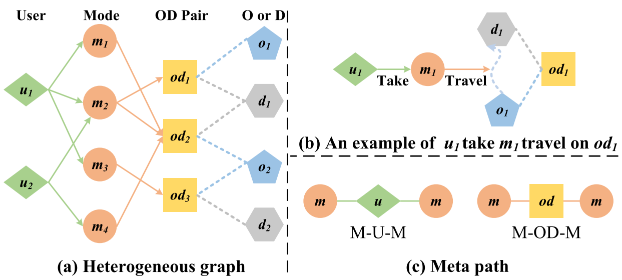


Fig. 3. Heterogeneous transportation graph in navigation data.

At present, some researchers focus on heterogeneous graphs [38], [46], [47], [48], [49]. For example, HAN [38] aims to convert heterogeneous graphs into homogeneous graphs and uses hierarchical attention to describe node-level and semantic-level structures. Although HAN has made great efforts in dealing with heterogeneous graph data, it still remains in the processing of homogeneous graphs, and it is difficult to obtain features from heterogeneous views (neighbors of different node types). To this end, a novel heterogeneous graph neural network algorithm HeCo is proposed to try to capture the influence of different types of nodes [49]. However, as the network deepens, there will be more and more training errors because of the problems of over-smooth in HAN and HeCo.

III. PRELIMINARIES

In this section, we give an explanation of the dataset representation, some definitions, and problem formulation.

A. Dataset Representation

In this paper, the navigation data we collect comes from Baidu Maps,[6] which records the process from the user first inputting the OD pair, then the route planning given by Baidu Maps, and finally, the user clicking and selecting the plan, as shown in Table I. The dataset consists of user attribute data, origin data, destination data, and transport mode data and is presented in a time-series format. Existing solutions cannot capture the relationship between the user, transport mode, origin, and destination when dealing with those data. To this end, we integrate the datasets and convert them into graph-structured data,[7] as shown in Figure 3. The graph-structured data contains nodes such as users, transport modes, OD pairs, origin and destination, and edges with relationships such as "Take" and "Travel". Moreover, it also implies that user "u" uses transport mode "m" to travel on OD pair "od". To represent the data more clearly, we use equations (1) and (2) to represent the node data in Figure 3,

$$\mathcal{X}_m = \{X_u, X_o, X_d, X_{con}\} \in \mathbb{R}^{N\times(m+p+p+k)}, \tag{1}$$
$$X_{od} = \{X_o, X_d\} \in \mathbb{R}^{N\times n}, \tag{2}$$

where $X_u \in \mathbb{R}^{N\times m}$, $X_{con} \in \mathbb{R}^{N\times k}$, $X_o \in \mathbb{R}^{N\times p}$, and $X_d \in \mathbb{R}^{N\times p}$ represent the user profile data, the context data, the origin data, and the destination data, respectively. m, k represent the dimension of the user profile data, and the context data, respectively. p is the dimension of the origin and destination data. n is the dimension of OD pairs data, $n = p + p$. N is the number of transport modes. Specifically, the total data volume (*i.e.*, N), the number of users, and the number of OD pairs are 308,507, 42,342, and 156,958, respectively.

B. Definition

Definition 1 (Origin-Destination (OD) Pair): An origin o is usually the starting point of path planning. In the navigation APP, it is usually the user's current location or other locations set by the user. A destination d is the end of path planning, and it is usually changed by the user. An OD pair $od = (o, d)$ is composed of origin and destination, which is a pair of regions.

Definition 2 (Heterogeneous Transportation Graph): A heterogeneous transportation graph is denoted as $G = (V, E)$, which consists of an object set V and an edge set E. A heterogeneous transportation graph is also associated with an object type mapping function $\phi : V \rightarrow A$ and an edge type mapping $\psi : E \rightarrow R$, where A and R denote the sets of object types and edge types, and $|A| + |R| > 2$.

As shown in Figure 3 (a), we build a heterogeneous graph to model the transportation graph, which consists of five types of objects (*i.e.*, User (u), Mode (m), OD Pair (od), Origin (o), and Destination (d)) and two types of relations (*i.e.*, "take" and "travel", *e.g.*, user takes mode, and mode travels OD pair).

Definition 3 (Network Schema): The network is defined as $T_G = (A, R)$, which is a meta template of a heterogeneous transportation graph G with the object type mapping function $\phi : V \rightarrow A$ and the edge type mapping $\psi : E \rightarrow R$. Meanwhile, T_G is a directed graph defined on the object type set A, with the relationship on R as the edge.

[6]https://map.baidu.com/
[7]https://github.com/xuaikun/THAN

For example, Figure 3 (b) describes the network schema of heterogeneous transportation graph. We usually use network schema to describe the direct connections between different nodes in G. The network schema is the local structure of a heterogeneous transportation network. In Figure 3 (b), we can know that mode m_1 is took by user u_1 travel on OD pair od_1.

Definition 4 (Meta Path): A meta path is defined as a path in the form of $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \cdots \xrightarrow{R_l} A_{l+1}$ (simplified to $A_1 A_2 \cdots A_{l+1}$), which describes a composite relation $R = R_1 \circ R_2 \circ \cdots \circ R_l$ between objects A_1 and A_{l+1}, where $\circ$ represents the composition operator on relations.

For example, Figure 3 (c) shows that modes can be connected via multiple meta paths, *e.g.*, Mode-User-Mode (MUM) and Mode-OD-Mode (MODM). Different meta paths describe semantic relations in different views. Specifically, MUM describes that two modes are taken by the same user, while MODM means two modes travel on the same OD pair. The meta path is regarded as a high-level structure because it is composed of multiple relationships and contains complex semantics.

Problem (Multi-Modal Transportation Recommendation Problem): Given a heterogeneous transportation graph G, a user u, an OD pair od, and some context information $cont$, we aim to recommend the most suitable transport mode m for users to travel on OD pair. Specifically, the optimal mapping function $f(\cdot)$ of transport mode is learned from the input data.

$$m = f(G, u, od, cont). \tag{3}$$

IV. THAN

This section presents the framework of THAN in detail.

A. Overview

Figure 4 shows an overview of THAN. It consists of four major parts: *Bigraph*, *Heterogeneous graph*, *Hierarchical attention*, and *Fusion neural layer*. The *Bigraph* module builds an origin-destination bipartite graph and a user-OD bipartite graph based on the existing dataset, and uses GCN and GAT to represent the feature of origin, destination, OD pair, and user, respectively. Thereafter, the *Heterogeneous graph* module extracts heterogeneous transportation graphs from large-scale data and assigns initial features to each node. Meanwhile, the *Hierarchical attention* module represents the embedding of transport modes from the view of homogeneity and heterogeneity. Finally, the *Fusion neural layer* module integrates homogeneous features and heterogeneous features to make recommendation.

B. Bigraph

In the navigation system, the destination will always appear with the origin, and the OD pair is formed by associating them. Obtaining the OD pair is one of the most critical tasks in the navigation task. The OD pair used by each user is almost different. Associating the OD pair with a specific user is very important for mining the user's personalized characteristics. To establish the link between origin data and destination data and the link between OD pairs data and user data, we utilize the origin data (O), the destination data (D), the OD pairs data, and the user profile which belongs to the multi-modal data to construct a bipartite graph. For each origin (or destination), an edge is built between the origin and the destination, only if the corresponding OD pairs exist. Then, we construct an origin-destination bipartite graph via $X_o \in \mathbb{R}^{N\times p}$ and $X_d \in \mathbb{R}^{N\times p}$, and then the node of bipartite graph aggregates information of neighbors by GCN model. Combining the origin and destination data can update the features of OD pairs data $X_{od} \in \mathbb{R}^{N\times n}$. Obviously, we can intuitively understand that the origin (or destination) affects the destination (or origin) from the origin-destination bipartite graph. For each OD pair or user, an edge will be built between the OD pair and the user, only if the user has traveled on the OD pair. Finally, a user-OD bipartite graph is constructed via $X_u \in \mathbb{R}^{N\times m}$ and $X_{od} \in \mathbb{R}^{N\times n}$, and the node information is updated by the GAT model. To achieve the above functions, we propose a method named Multi-Bigraph Embedding (MBigraphE) to extract the latent features of o, d, od, and u, which includes the GCN layer and GAT layer.

1) GCN Layer: After getting the origin-destination bipartite graph, we update the characteristics of origin and destination separately by GCN,

$$e_{o_i}^{(l+1)} = \sum_{d_j \in N(o_i)} \frac{1}{\sqrt{|N(o_i)||N(d_j)|}} e_{d_j}^{(l)}, \tag{4}$$

$$e_{d_i}^{(l+1)} = \sum_{o_j \in N(d_i)} \frac{1}{\sqrt{|N(d_i)||N(o_j)|}} e_{o_j}^{(l)}, \tag{5}$$

where $e_{d_j}^{(l)}$ and $e_{o_j}^{(l)}$ are the embeddings of the l^{th} layer of d_j and o_j, respectively. $N(o_i)$ and $N(d_i)$ represent the set of neighbors of origin o_i and destination d_i, respectively. $|N(o_i)|$, $|N(d_i)|$ are the degrees of origin o_j and destination d_i respectively. $e_{od} = e_o||e_d$, when the corresponding OD pairs is exist.

2) GAT Layer: After getting the user-OD bipartite graph, the initial representation matrix of user-OD bipartite E is

$$E = [\underbrace{e_{od_1}^{(0)}, e_{od_2}^{(0)}, \cdots, e_{od_M}^{(0)}}_{OD\ pair\ embedding}, \underbrace{e_{u_1}^{(0)}, e_{u_2}^{(0)}, \cdots, e_{u_Q}^{(0)}}_{user\ embedding}], \tag{6}$$

where M and Q represent the number of OD pair and user, respectively. $e_{od}^{(0)}$, $e_u^{(0)}$ are the initialization of OD pair embedding and user embedding respectively.

In order to obtain the feature information of the user more efficiently, we aggregate the messages propagated from the neighborhood of u to refine the embedding of u. Specifically, we define the aggregation function as

$$e_{u_i}^{(l+1)} = \sum_{od_j \in N_{(u_i)}} a_{u_i od_j} W_1^{(l)} e_{od_j}^{(l)}, \tag{7}$$

where $e_{u_i}^{(l+1)}$ denotes the embedding of u_i obtained at $(l+1)^{th}$ GAT layer, $N_{(u_i)}$ is neighbor set of u_i, $W_1^{(l)}$ is the trainable weight matrices. $a_{u_i od_j}$ is attention value between u_i and od_j

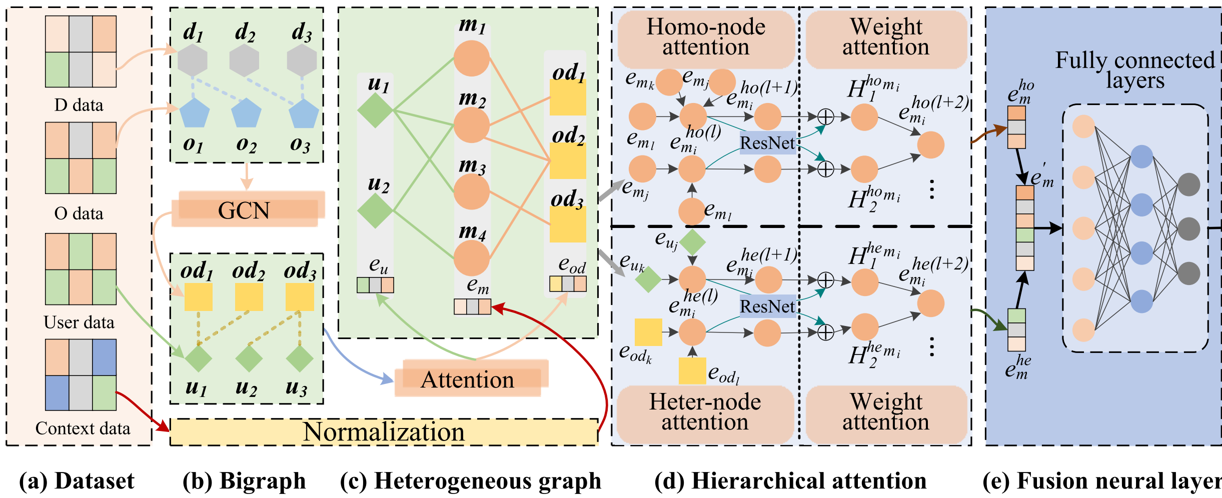


Fig. 4. THAN overview.

that can be shown as

$$a_{u_i od_j}^{(l)} = softmax(edge_{u_i od_j}^{(l)}), \tag{8}$$

$$edge_{u_i od_j}^{(l)} = \rho(\mathbf{a}^T[W_2^{(l)} e_{u_i}^{(l)} || W_3^{(l)} e_{od_j}^{(l)}]), \tag{9}$$

where $W_2^{(l)}$, $W_3^{(l)}$ are the trainable weight parameters, ρ denotes the activation function *LeakyReLU*, $\mathbf{a}$ is the attention weight vector, $||$ denotes the concatenate operation, $e_{u_i}^{(l)}$ and $e_{od_j}^{(l)}$ are the representation of u_i and od_j in l layer of GAT network. Therefore, the embedding of user (or OD pair) based on the user-OD bipartite graph, through the GAT layer, is e_u (or e_{od}). The context data after standard normalization is denoted as e_m.

C. Heterogeneous Graph

After analyzing a large number of navigation data, we found that the data not only includes entities such as users, origins, destinations, OD pairs, and traffic modes, but also includes use (*i.e.*, user use the mode) and travel (*i.e.*, the mode travel on OD pair) relationships. This fits perfectly with the definition of a heterogeneous graph, as shown in Section III-B, so we can project navigation data onto a heterogeneous transportation graph for processing. The heterogeneous graph contains the structure of (u, m, od), which represents user u uses transport mode m on od. We aim to recommend the most suitable transport mode m for user u to travel od. Based on previous results, the embedding vector for each node of the heterogeneous graph can be shown as

$$e_{init} = e_u || e_{od} || e_m, \tag{10}$$

where $||$ denotes the concatenate operation. Specifically, the final embedding of the transport mode can be shown as

$$\mathcal{E}_m = \{E_u, E_{od}, E_m\} \in \mathbb{R}^{N \times (f_1 + f_2 + f_3)}, \tag{11}$$

where $E_u \in \mathbb{R}^{N \times f_1}$, $E_{od} \in \mathbb{R}^{N \times f_2}$, and $E_m \in \mathbb{R}^{N \times f_3}$ respectively represent the heterogeneous nodes embedding from multi-modal data, f_1, f_2, and f_3 represent the size of the embedding dimension. $e_m' \in \mathcal{E}_m$ denotes the multi-modal embedding vector for each transport mode.

D. Hierarchical Attention

In a heterogeneous transportation graph, most nodes have no less than one neighbor node, but each neighbor node usually has a different effect on the current node, so the importance of identifying neighbors is particularly important. In our work, neighbor nodes are obtained according to different meta-paths, and the importance of neighbors obtained through different meta-paths to the current node also needs to be distinguished. To distinguish the importance of nodes and semantics as well as solve gradient instability, we design hierarchical attention with residual block to mine the latent features of the temporary embedding of the transport mode, which includes the ResNet layer, Node attention layer, and Weight attention layer.

1) ResNet Layer: For deep neural networks, as the network deepens, it should be trained better and better. However, as the network deepens, there will be more and more training errors because of the problems of over-smooth. Preliminary work has confirmed that ResNet does help to solve the over-smooth problem. It makes it possible to ensure good performance when training deeper networks. Therefore, we added residual block in the training process of heterogeneous graph network, *i.e.*, the value of $H_{\Psi_s}^{hx_m}$ is directly affected by both $e_m^{hx(l+1)}$ and $e_m^{hx(l)}$, where $hx \in (ho, he)$.

$$H_{\Psi_p}^{hx_m} = \sigma(W_4^{hx(l+2)} e_m^{hx(l+1)} + b^{hx(l+2)} + e_m^{hx(l)}), \tag{12}$$

where hx is the type of node (*i.e.*, homogeneous nodes (ho) and heterogeneous nodes (he)). Give the meta path set $\{\Psi_1, \Psi_2, \cdots, \Psi_S\}$, which included Ψ_s. σ represents activation function *eLU*, $W_4^{(l+2)}$ is the trainable weight matrix, $b^{(l+2)}$ is the bias vector. $\mathbf{e}_m^{(l)}$ denotes the temporary representation for transport mode m at the layer l.

2) Node Attention Layer: Based on the graph attention mechanism, $e_{m_i}^{(l+1)}$ is computed as

$$e_{m_i}^{hx(l+1)} = \sum_{m_j \in N_{(m_i)}} a_{m_i m_j} W_5^{(l)} e_{m_j}^{hx(l)}, \tag{13}$$

where $e_{m_i}^{(l+1)}$ is the temporary embedding of transport mode m at the layer $l+1$, which contains information about neighbor nodes of m_i.

3) Weight Attention Layer: It is worth noting that the importance of edges between neighbors will also affect the representation of transport mode m. The final embedding of each transport mode m is aggregated by all edges embedding in equation (14). Then we can apply the final embedding to specific tasks, *e.g.*, node classification, node prediction, and recommended system,

$$e_m^{hx(l+2)} = \sum_{s=1}^{S} softmax(H_{\Psi_s}^{hx_m}) \cdot e_m^{hx(l+1)}. \tag{14}$$

E. Fusion Neural Layer

As shown in Figure 4, the final transport mode feature is determined by both the heterogeneous view feature and the homogeneous view feature. Therefore, it is necessary to merge the above two features for prediction. In this paper, hx is represents the type of homogeneous node and heterogeneous nodes (*i.e.*, $hx \in (ho, he)$), homogeneous nodes (ho) and heterogeneous nodes (he)). The multi-modal embedding e'_m with multiple fully connected layers is used to predict the most suitable transport mode m for user u to travel od,

$$e'_m = e_m^{ho} || e_m^{he}, \tag{15}$$

$$y'_m = softmax(\varrho(W_6^{(l)} e'^{(l)}_m + b^{(l)})), \tag{16}$$

where $W_6^{(l)}$ is the trainable weight matrix, ϱ is the activation function *ReLU*, $e_m^{'(l)}$ denotes the final multi-modal representation for transport mode m at the layer l, and $b^{(l)}$ is the bias vector. Finally, we use softmax function and obtain the final prediction score y'_m.

F. Optimization

For model optimization, our method could be trained with a supervised setting. Based on the cross-entropy loss, the objective function could be defined as

$$Loss = -\sum_{i=1}^{C} y_{m_i} log(y'_{m_i}), \tag{17}$$

where C is the number of classification, y_{m_i} is the actual label, y'_{m_i} is the predictive scores. Under the guidance of labeled data, we can optimize the proposed model through backpropagation and learn the embedding of transport mode m.

V. Experiments

In this section, we describe the experimental setups and results. The results obtained with THAN are compared with five of the state-of-the-art methods for multi-modal transportation recommendation.

A. Experimental Setup

1) Dataset: We conduct experiments on our own dataset to verify the performance of the proposed framework THAN. In this dataset, users, transport modes, and OD pairs respectively represent nodes of heterogeneous graphs. Moreover, users clicking transport mode and transport modes travel on OD pairs represent relationships (*i.e.*, edges) of heterogeneous graphs. In this paper, we consider eleven transport modes {*bus, metro, car, taxi, walk, bicycle, bus-metro, taxi-bus, bicycle-metro, taxi-metro, and bicycle-bus-metro*}. We encode these transport modes from mode 1 to mode 11. The speed comparison between different modes of transportation is shown in Figure 5. Mode 3 (car) has the fastest speed, and mode 5 (walk) is the slowest. The utilization rate of the transport mode is $u\_rate = \frac{part\_times}{total\_times}$, where $part\_times$ is the number of times the transport mode is used within a certain path length and $total\_times$ is the total number of times the transport mode is used. u_rate has a very strong correlation with the travel distance. As shown in Figure 6, modes 5 and 6 are usually used for trips less than 8 kilometers.

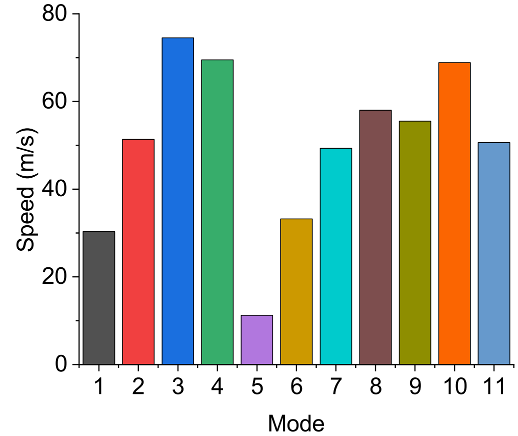


Fig. 5. Speed of transport mode.

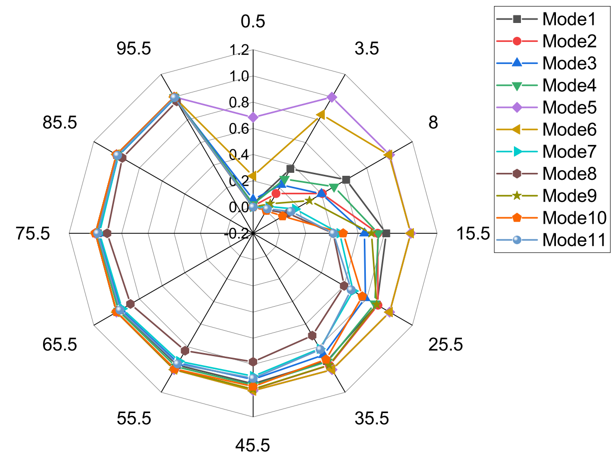


Fig. 6. Relationship between transport mode utilization rate u_rate and travel distance.

2) Parameter Settings: The proposed method is implemented with Pytorch and optimized by Adam with a learning rate of 0.0001 in DGL. The number of heads in the attention mechanism is 8. The hidden layer dimensions in homogeneous and heterogeneous structures are 128 and 64, respectively. The parameters of NMTRec and Hydra remain the same as the work [3], [9], *i.e.*, the maximum depth is 5 and the columns sample rate is 0.8. The related hyperparameters settings of

TABLE II
COMPARISON OF PRE, REC, NDCG AND F1 BETWEEN NMTREC, HYDRA, HMTRL, HAN, HECO AND THAN METHODS (TRAIN AND TEST ON BEIJING DATASET)

Methods	F1	Pre	Rec	NDCG
NMTRec	0.6908	0.6932	0.7245	0.8878
Hydra	0.7506	0.7508	0.7786	0.9082
HMTRL	0.7411	0.7634	0.7731	0.9344
HAN	0.6706	0.7337	0.7115	0.9098
HeCo	0.6895	0.7661	0.7135	0.9462
THAN	**0.7566**	**0.797**	**0.7848**	**0.9523**

heads in the multi-head attention of HAN, HeCo, and HMTRL are the same THAN, *i.e.*, the number of heads is 8, where the dropout of attention to 0.6. And we use early stopping with the patience of 200, *i.e.*, we stop training if the validation loss does not decrease for 200 consecutive epochs. The source code and datasets are publicly available on Github[8] after the paper is accepted. The number of transport mode, users, and OD pairs are respectively $N = 308507$, $Q = 42342$, and $M = 156958$. The dimension of user data, OD pair, and context data are respectively $m = 66$, $n = 8$, and $k = 57$. The dimension of o and d data are $p = 4$. The dimension of user embedding, context embedding, and od embedding are respectively $f_1 = 66$, $f_2 = 2071$, and $f_3 = 611$. We split all data randomly, the training set 70%, the validation set 10%, and the test set 20%.

3) Evaluation Metrics: We evaluate the THAN with four metrics including the Precision (Pre), Recall (Rec), Normalized Discounted Cumulative Gain (NDCG), and F1. The NDCG metric considers all transport modes, while the remaining metrics only focus on the top-1 recommendation. There are many classes of transport mode, we define F1 as the weight F1. For each class, the F1 score is

$$F_{1,class_i} = \frac{2Pre \times Rec}{Pre + Rec}, \tag{18}$$

The weighted F1 considers the weight of each class, and can be formulated as

$$F1 = w_1 F_{1,class_1} + w_2 F_{1,class_2} + \cdots + w_i F_{1,class_i} \tag{19}$$

where the weight w_i is calculated by the ratio of true instances for class i.

B. Results and Analysis

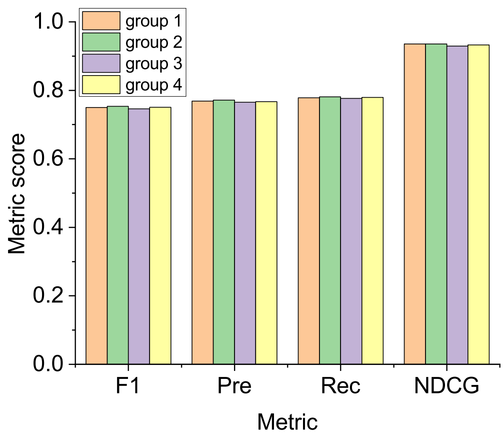
We analyze the experimental results on a large-scale navigation dataset to verify the effectiveness of THAN, including performance comparison, robustness verification, ablation tests, MBigraphE rationality, and parameter sensitivity.

1) Performance Comparison: We compare the results of our method for transport mode prediction with that of five state-of-the-art models including NMTRec [3], Hydra [9], HMTRL [16], HAN [38], and HeCo [49], as shown in Table.II. Overall, THAN outperforms all the baselines on the Beijing dataset using all metrics, which demonstrates the advance of our model. Specifically, THAN achieves (0.8%, 6.2%, 0.8%, 4.9%) F1, Pre, Rec, and NDCG improvement compared with the state-of-the-art scheme (Hydra) on the Beijing dataset. Moreover, we can make the following observations, (1) the performance of NMTRec is much worse than Hydra. This observation indicates that using too many features to train makes the training process over-fitting. (2) Hydra is a baseline algorithm by incorporates fine-grained handcrafted features and high-order embedding features. However, compared with graph learning-based methods, *e.g.*, HAN, HeCo, and HMTRL, the manually extracted features limit the recommendation capability of the model. (3) HAN, HeCo, and HMTRL are graph learning-based methods, which can make up for the shortcomings of the former. However, HAN, HeCo, and HMTRL ignore the interaction of constituent nodes and the importance of residual structure in the training process, therefore performing worse than our method. In conclusion, NMTRec and Hydra are worse than THAN. This is because, there are rich spatial relationships between nodes in traffic datasets, which are usually not negligible. THAN can learn spatial features, while NMTRec and Hydra cannot. Compared with existing graph neural network schemes (*i.e.*, HAN, HeCo, and HMTRL), THAN performs better. This is because THAN can not only learn the relationship between homogeneous nodes, but also enhance the feature representation from the perspective of heterogeneity.

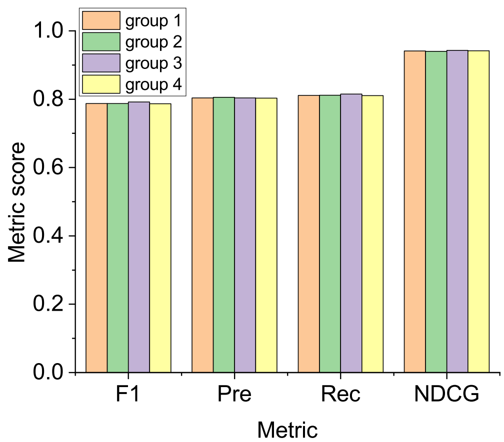
2) Robustness Verification: A robust algorithm should perform evenly on different query subgroups and have similar performance. To verify the robustness of THAN, we divide the data evenly into four groups and apply our model to test the performance of each group's data. Figure 7 shows the performance of THAN on different subgroups on the Beijing dataset. For different groups, the results are strongly stable on four metrics. Specifically, the difference in values does not exceed 2% in all our tests using the four metrics. The result validates that our THAN is robust for different transport mode recommendation. Moreover, the validation results on the Guangzhou dataset show that the values of F1, Pre, Rec, and NDCG are similar to the results obtained on the Beijing dataset and also maintain a difference within 2%, which further explains THAN has good robustness.

3) Ablation Test: To further explore the effect of multimodal fusion in our proposed model, we compare THAN with THAN(ResNet), THAN(MBigraphE), THAN(HierAtt), and THAN(ResNet&MBigraphE), where THAN(ResNet), THAN(MBigraphE), THAN(HierAtt), and THAN(ResNet&MBigraphE) are parts of our model and described as no consider ResNet layer, MBigraphE, HierAtt, and both ResNet and MBigraphE to make transport modes prediction, respectively. As shown in Figure 8, the performance of THAN outperforms THAN(ResNet), THAN(MBigraphE), THAN(HierAtt), and THAN(ResNet&MBigraphE). Moreover, the THAN(MBigraphE) and THAN(HierAtt) perform better than THAN(ResNet), which demonstrates the residual block plays a more important role in multi-modal transportation recommendation. This is because ResNet does help to solve the problem of over-smooth, making training better, or at least ensuring that the current result is no worse than the last. What's more, MBigraphE can balance the importance of features. This is because MBigraphE not only learns

[8]https://github.com/xuaikun/THAN



(a) Beijing.



(b) Guangzhou.

Fig. 7. Robustness by group on different datasets.

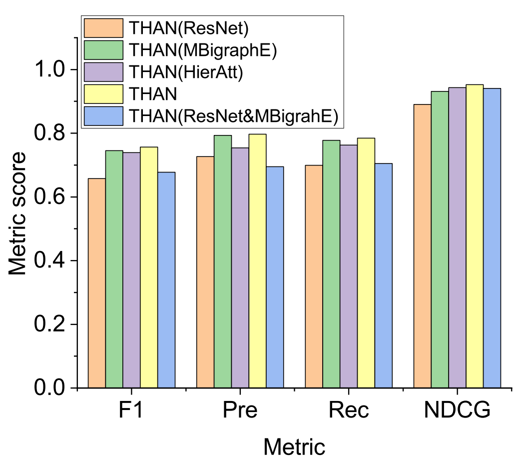


Fig. 8. Ablation test of THAN.

neighbor embeddings but also distinguishes the importance of neighbors.

4) MBigraphE Rationality: Then, We verified the rationality of MBigraphE by evaluating three variants of THAN, (1) THAN_o_d_od_pid is not considering MBigraphE in THAN, (2) THAN_o_d means that the interaction between o and d is not considered, (3) THAN_od_pid represents than the interaction between od and u is not considered. As shown in Figure 9, considering the interaction between o and d or od and u is better than not considering the effect. Moreover,

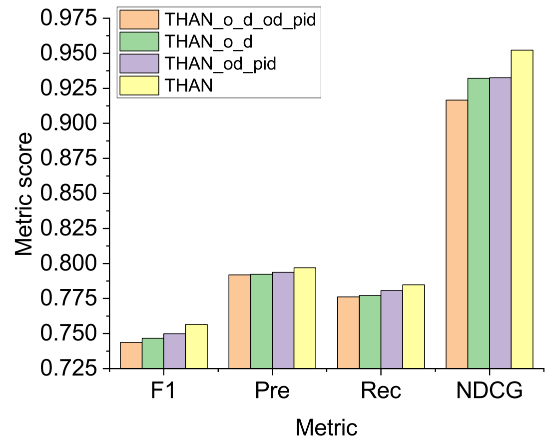


Fig. 9. MBigraphE validity verification.

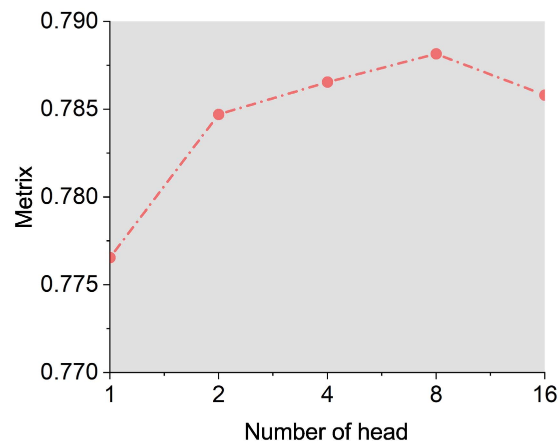


Fig. 10. Evaluation of the number of heads.

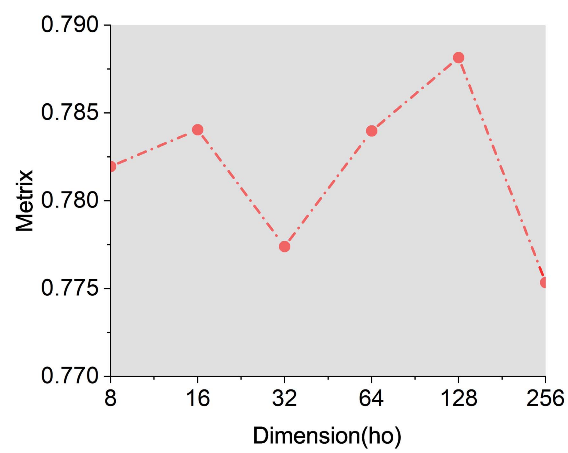


Fig. 11. Dimensional evaluation of the hidden layer of homogeneous structure.

considering the interaction between o and d and od and u at the same time is better than considering the effect separately, demonstrating the effectiveness of MBigraphE.

5) Parameter Sensitivity: We further study the parameter sensitivity of THAN. Each time we vary a parameter, we set others to their default values.

Firstly, we vary the number of heads from 1 to 16. The results are reported in Figure 10. As the head increases, the performance first increases and then decreases.

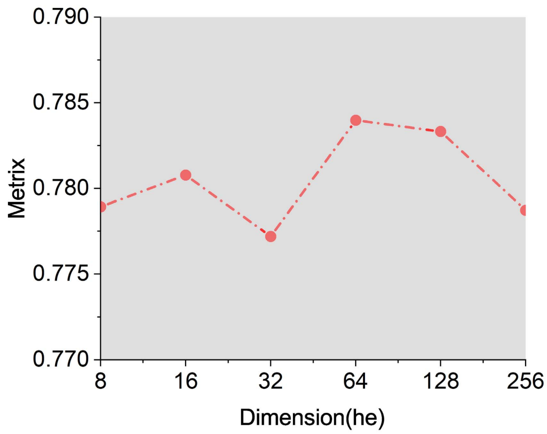


Fig. 12. Dimensional evaluation of hidden layers of heterogeneous structure.

Therefore, setting the number of heads to 8 can achieve the best performance.

Then, we vary the dimension of homogeneous structure from 8 to 256. The results are reported in Figure 11. Using 128 dimensions is good enough to capture representation information in the homogeneous structure.

Finally, we vary the dimension of heterogeneous structure from 8 to 256. The results are reported in Figure 12. The best performance can be achieved when the dimension is 64 in the heterogeneous structure.

VI. Conclusion

Although transportation recommendation is increasingly popular and frequently used, the existing solutions still cannot meet the diverse needs of users. After analyzing the data, we found that the navigation data can be represented as a graph. The graph neural network has shown a powerful ability in processing graph data. In this paper, we propose a novel and base-graph learning framework, named THAN, which can recommend the most suitable transport mode for users. Firstly, based on the powerful expression capabilities of GCN and GAT in bipartite graphs, we design a mining method called MBigraphE to represent the potential relationship between heterogeneous nodes. Then, we construct a heterogeneous transportation graph based on the relationship between nodes in a large-scale dataset and initialize node features. In order to obtain the representation of nodes, we further study the features of nodes from both the perspectives of homogeneity and heterogeneity, and use the residual block to stabilize network representation. Finally, we design a fusion neural layer to fuse embeddings from different views, it can predict and recommend the most suitable transport mode for users. We conduct extensive experiments on a large-scale real dataset. The results show that the system has achieved state-of-the-art performance.

References

[1] H. Bast et al., "Route planning in transportation networks," in *Algorithm Engineering*, vol. 9220. Cham, Switzerland: Springer, 2016, pp. 19–80.

[2] S. Wang, W. Lin, Y. Yang, X. Xiao, and S. Zhou, "Efficient route planning on public transportation networks: A labelling approach," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, May 2015, pp. 967–982.

[3] Y. Liu, C. Lyu, Z. Liu, and J. Cao, "Exploring a large-scale multi-modal transportation recommendation system," *Transp. Res. C, Emerg. Technol.*, vol. 126, May 2021, Art. no. 103070.

[4] J. Wang, J. Jiang, W. Jiang, C. Li, and W. X. Zhao, "Libcity: An open library for traffic prediction," in *Proc. 29th Int. Conf. Adv. Geographic Inf. Syst.*, 2021, pp. 145–148.

[5] J. Wang, N. Wu, and W. X. Zhao, "Personalized route recommendation with neural network enhanced A* search algorithm," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5910–5924, Dec. 2021.

[6] C. Li, L. Bai, W. Liu, L. Yao, and S. T. Waller, "A multi-task memory network with knowledge adaptation for multimodal demand forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 131, Oct. 2021, Art. no. 103352.

[7] T. Litman, "Introduction to multi-modal transportation planning," Victoria Transp. Policy Inst., Victoria, BC, Canada, Technol. Rep., 2021.

[8] H. Liu, T. Li, R. Hu, Y. Fu, J. Gu, and H. Xiong, "Joint representation learning for multi-modal transportation recommendation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, Jul. 2019, pp. 1036–1043.

[9] H. Liu, Y. Tong, P. Zhang, X. Lu, J. Duan, and H. Xiong, "Hydra: A personalized and context-aware multi-modal transportation recommendation system," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 2314–2324.

[10] H. Liu et al., "Polestar: An intelligent, efficient and national-wide public transportation routing engine," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2020, pp. 2321–2329.

[11] H. Liu, Y. Tong, J. Han, P. Zhang, X. Lu, and H. Xiong, "Incorporating multi-source urban data for personalized and context-aware multi-modal transportation recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 2, pp. 723–735, Feb. 2020.

[12] J. Ke, S. Feng, Z. Zhu, H. Yang, and J. Ye, "Joint predictions of multi-modal ride-hailing demands: A deep multi-task multi-graph learning-based approach," *Transp. Res. C, Emerg. Technol.*, vol. 127, Jun. 2021, Art. no. 103063.

[13] J. Dai, B. Yang, C. Guo, and Z. Ding, "Personalized route recommendation using big trajectory data," in *Proc. IEEE 31st Int. Conf. Data Eng.*, Apr. 2015, pp. 543–554.

[14] J. Wang, N. Wu, W. X. Zhao, F. Peng, and X. Lin, "Empowering A* search algorithms with neural networks for personalized route recommendation," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 539–547.

[15] H. Liu, J. Han, Y. Fu, J. Zhou, X. Lu, and H. Xiong, "Multi-modal transportation recommendation with unified route representation learning," *Proc. VLDB Endowment*, vol. 14, no. 3, pp. 342–350, Nov. 2020.

[16] H. Liu, J. Han, Y. Fu, Y. Li, K. Chen, and H. Xiong, "Unified route representation learning for multi-modal transportation recommendation with spatiotemporal pre-training," *VLDB J.*, pp. 1–18, May 2022.

[17] L. Fu, D. Sun, and L. R. Rilett, "Heuristic shortest path algorithms for transportation applications: State of the art," *Comput. Oper. Res.*, vol. 33, no. 11, pp. 3324–3343, Nov. 2006.

[18] R. J. Szczerba, P. Galkowski, I. S. Glicktein, and N. Ternullo, "Robust algorithm for real-time route planning," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 3, pp. 869–878, Jul. 2000.

[19] S. Rogers and P. Langley, "Personalized driving route recommendations," in *Proc. Amer. Assoc. Artif. Intell. Workshop Recommender Syst.*, 1998, pp. 96–100.

[20] Z. Yu, "Trajectory data mining: An overview," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, pp. 1–41, 2015.

[21] J. Yuan et al., "T-drive: Driving directions based on taxi trajectories," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 99–108.

[22] T. Litman, *Evaluating Public Transit Benefits Costs*. Victoria, BC, Canada: Victoria Transport Policy Institute Victoria, 2015.

[23] Y. Liu et al., "Exploiting heterogeneous human mobility patterns for intelligent bus routing," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2014, pp. 360–369.

[24] H. Zhou, Y. Zhao, J. Fang, X. Chen, and K. Zeng, "Hybrid route recommendation with taxi and shared bicycles," *Distrib. Parallel Databases*, vol. 38, no. 3, pp. 563–583, 2020.

[25] M. Hu, "Bike sharing," in *Sharing Economy: Making Supply Meet Demand*, vol. 6. New York, NY, USA: Springer, 2019, pp. 435–459.

[26] A. Faghih-Imani, S. Anowar, E. J. Miller, and N. Eluru, "Hail a cab or ride a bike? A travel time comparison of taxi and bicycle-sharing systems in New York city," *Transp. Res. A, Policy Pract.*, vol. 101, pp. 11–21, Jul. 2017.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[27] P. Campigotto, C. Rudloff, M. Leodolter, and D. Bauer, "Personalized and situation-aware multimodal route recommendations: The FAVOUR algorithm," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 92–102, Jan. 2016.

[28] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 1–9.

[29] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–9.

[30] A. Abedalla et al., "MTRecS-DLT: Multi-modal transport recommender system using deep learning and tree models," in *Proc. 6th Int. Conf. Social Netw. Anal., Manage. Secur. (SNAMS)*, Oct. 2019, pp. 274–278.

[31] Q. Sun, Z. Qu, and C. Ren, "Context-aware multi-modal transportation recommendation based on particle swarm optimization and LightGBM," *Acta Electonica Sinica*, vol. 49, no. 5, p. 894, 2021.

[32] Q. Sun, L. Chang, C. Ma, and Z. Qu, "Multi-modal transportation recommendation based on graph embedding and cagbdt," *J. Beijing Univ. Posts Telecommun.*, vol. 44, no. 5, p. 81, 2021.

[33] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2008.

[34] J. Gao, T. Lyu, F. Xiong, J. Wang, W. Ke, and Z. Li, "MGNN: A multimodal graph neural network for predicting the survival of cancer patients," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2020, pp. 1697–1700.

[35] M. Gori, G. Monfardini, and F. Scarselli, "A new model for learning in graph domains," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 2, Jul. 2005, pp. 729–734.

[36] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[37] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2020.

[38] X. Wang et al., "Heterogeneous graph attention network," in *Proc. World Wide Web Conf.*, 2019, pp. 2022–2032.

[39] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," 2013, *arXiv:1312.6203*.

[40] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 3844–3852.

[41] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 1025–1035.

[42] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.

[43] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[44] X. Han, C. Shi, S. Wang, P. S. Yu, and L. Song, "Aspect-level deep collaborative filtering via heterogeneous information networks," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3393–3399.

[45] X. Ying, C. Xu, J. Gao, J. Wang, and Z. Li, "Time-aware graph relational attention network for stock recommendation," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 2281–2284.

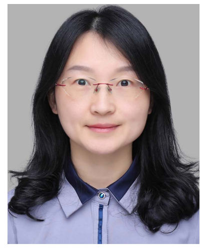
[46] X. Fu, J. Zhang, Z. Meng, and I. King, "MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding," in *Proc. Web Conf.*, 2020, pp. 2331–2341.

[47] S. Yun, M. Jeong, R. Kim, J. Kang, and H. J. Kim, "Graph transformer networks," 2019, *arXiv:1911.06455*.

[48] C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla, "Heterogeneous graph neural network," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 793–803.

[49] X. Wang, N. Liu, H. Han, and C. Shi, "Self-supervised heterogeneous graph neural network with co-contrastive learning," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2021, pp. 1726–1736.

Aikun Xu received the M.S. degree from Central South University, China, in 2022, where he is currently pursuing the Ph.D. degree. His research interests include machine learning, scheduling, electric vehicles, and edge computing.



Ping Zhong (Member, IEEE) received the Ph.D. degree in communication engineering from Xiamen University, China, in 2011. She is currently an Associate Professor with the School of Computer Science and Engineering, Central South University. Her research interests include machine learning, data mining, and network protocol design. She is a member of ACM, CCF, and IEICE.

Yilin Kang received the Ph.D. degree in computer science from Nanyang Technological University, Singapore. She is currently an Assistant Professor with the School of Computer Science, South Central University for Nationalities, China. Prior to joining SCUN, she was a Researcher Fellow with the NTU-UBC Joint Research Centre of Excellence in Active Living for the Elderly (LILY). Her current research interests include human-centric computing, brain-inspired intelligent agent, cognitive and neural systems, and human–computer interaction. She serves as a PC Member for AAAI 2020, AAAI 2019, and IJCAI 2016, and an OC Member for IEEE WI/IAT 2015. She serves as a Reviewer for several major journals, such as IEEE TRANSACTIONS ON NEURAL NETWORKS LEARNING SYSTEMS, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, and *JAAMAS*.

Jiongqiang Duan is currently pursuing the bachelor's degree with the School of Computer Science and Engineering, Central South University, Changsha, China. His research interests include machine learning, scheduling, electric vehicles, and wireless networks.

Anning Wang received the B.S. degree from the School of Computer Science and Engineering, Central South University, Changsha, China, in 2021, where she is currently pursuing the master's degree. Her research interests include machine learning, data mining, and urban intelligence.

Mingming Lu (Member, IEEE) received the Ph.D. degree in computer science from Florida Atlantic University, USA, in 2008. He is currently an Associate Professor with the School of Computer Science and Engineering, Central South University, China. He has published over 60 journals and conference papers. His current research interests include deep learning, source code mining, and edge computing.

Chuan Shi (Member, IEEE) received the B.S. degree from Jilin University in 2001, the M.S. degree from Wuhan University in 2004, and the Ph.D. degree in ICT from the Chinese Academic of Sciences in 2007. He joined the Beijing University of Posts and Telecommunications as a Lecturer in 2007. He is currently a Professor and the Deputy Director of the Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia. He has published more than 100 papers in refereed journals and conferences. His research interests include data mining and machine learning.