

The potential of explainable deep learning for public transport planning

Wenzhe Sun, Jan-Dirk Schmöcker, Youxi Lai

Department of Urban Management,
Kyoto University,
Kyoto, Japan

{wz.sun, schmoecker, ylai}@trans.kuciv.kyoto-u.ac.jp

Koji Fukuda

Global Center for Social Innovation,
Hitachi, Ltd.,
Tokyo, Japan

koji.fukuda.jf@hitachi.com

Abstract— Deep learning has been widely used in transportation studies, especially for prediction tasks. It has been proven powerful in predicting travel demand for a transportation node or a general spatial unit. Recently, Graph Convolutional Networks (GCNs) have drawn much attention with their ability to model complex interactions between spatial units of an area by describing them with graphs. Multi-relational GCNs can be used for cases that consider multiple node features and relations between nodes. These have a variety of applications for public transport planning. To introduce GCNs and their variants to a wider range of audiences, this chapter aims to provide a general and easy-to-follow framework based on several examples. Furthermore, this chapter discusses the potential to explain such black boxes by using SHapley Additive exPlanations (SHAP). SHAP may address the concern that arises among transport planners and operators regarding model interpretability.

Keywords— *Explainable deep learning; Graph Convolutional Networks; Shapley Additive Explanations; Demand prediction; Public transport planning*

I. INTRODUCTION

New challenges and opportunities confront public transport planning in the current era of intelligent transport systems blessed with a big data environment. The enhanced data availability with higher spatial and temporal resolutions enable transport planners and operators to obtain more accurate and informative estimates or predictions. These contribute to data-driven transport planning and operation. Public transport operators have welcomed this possibility of better service provision because the capacity of traditional models to deal with big data and real-time dynamics has prevented them from developing advanced tools to further improve their services.

The emerging deep learning approaches can process comprehensive data in a significantly efficient way. Among them, Graph Convolutional Networks (GCNs) and their variants are better at capturing spatial relationships for a natural network or derived network structure, including molecular structures (Ryu et al. 2018), social networks (Wu et al. 2018), and a number of network-structured real-world phenomena. Therefore, they are receiving increasing attention from transportation researchers. They have been widely used to predict a variety of transportation service attributes including traffic flow, traffic speed, and travel demand (Yu et al. 2017; Guo et al. 2019; Yu et al. 2020; Ke et al. 2021; Lee et al. 2022). Despite some academic contributions, practical solutions are still limited. In order to bridge the gap between traditional and new models, this chapter aims to generalize the existing advanced GCN-based approaches and conceptualize a deep learning framework for a wide range of readers. Long-term predictions of travel demand at the city-level as well as short-term predictions of public transport travel demand at more detailed spatial levels are covered.

Arguments also arise about the lack of interpretability in deep learning approaches. Traditional regression techniques still enjoy popularity if the sensitivity of output changes to input changes is desired, e.g. the impact magnitude of a factor on travel demand. The recent efforts made by the computer science community to make deep learning explainable are noteworthy. Among them, a series of methods based on SHapley Additive exPlanations (SHAP) are well recognized. SHAP methods are in analogy to Shapley Values (Shapley, 1953; Hart, 1989), breaking down any prediction result associated with the number of input variables as a payoff to each player. A variant tailored for GCNs can distribute payoffs considering the incorporated network structure and make GCN-based deep learning explainable. These efforts are expected to pave the way for explainable deep learning and can address the current concerns about the incapability of black-box approaches to interpret how the results are derived.

The remainder of this chapter is organized as follows. The following section discusses the advances in data collection environments for public transport planning and new needs of public transport modelling. It points out the mismatch between the traditional methods and new big data environments. In response to the challenges facing traditional modelling approaches, this chapter then introduces a comprehensive graph-based deep learning framework. It goes through several examples to ease the understating of this deep learning concept. To address the concerns of practitioners that deep learning approaches are unexplainable and therefore not suitable for planning and policy-making purpose, a new method based on SHAP modified for GCNs is introduced. It significantly strengthens model interpretability of graph-based deep learning. This method can decompose a prediction result into several parts

and link each part to each model input. An example using Swiss data following and extending the work presented in Lai et al. (2022) is presented. Finally, this chapter provides conclusions and proposes some future research directions that the authors believe are noteworthy to pursue in future studies.

II. PUBLIC TRANSPORT PLANNING

Compared to road traffic planning, public transport planning is arguably far more complex. The planning and design of a public transportation service includes a number of steps as outlined comprehensively in the books by Ceder (2016) and by Vuchic (2017): The road infrastructure often serves as a constraint on top of which then stops are placed, routes planned and timetables designed. Vehicle and staff requirements and constraints are also considered. Capacity, waiting time and transfer issues also arise as a consequence of the discrete nature of the public transport services. Furthermore, fare settings, service quality requirements and catering for passengers with special needs are additional issues that need to be considered in the planning process.

A key for successful planning is appropriate demand estimates. Since the demand will respond to the service design, it is important to define “demand functions” and “demand sensitivity”. There is a significant body of literature discussing demand elasticity to fares, access time and travel time in general, but clearly the values are location and time of day specific as the willingness to take public transport will also depend on other available modes.

Furthermore, establishing demand sensitivity to “demand generators” and “demand attractors” is also critical. Sun and Schmöcker (2021) discuss the difference in planning for a new service and planning for service modifications. Whereas in the latter case a base demand is available and elasticities might be more easily established, that is not the case in the former case. Here one has to rely on the family of “gravity type models” to establish a base demand for a new network. Though there is a wide range of specific models, the basic idea that demand is a function of the population and its attributes at one place and attractiveness of an area weighted by the travel discomfort is common to most models. Functional forms may vary as well as to whether mode choice should be completed first and whether, for example, models consider trips, home-to-home tours or other forms of activity chains.

The complexity of a gravity model is limited by the observations available to calibrate it as well as the accuracy of the measure of the attractiveness of one area. Traditional models presented among others in Ortúzar and Willumsen (2011), discuss that number of workplaces and shops are useful measures. A comprehensive discussion can also be found in Chapter 4 of “Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems” (Nökel and Gentile, 2017). The book highlights that in the current era the basic structure still remains. Advances are mainly made in the details of the choice models that reflect the sensitivity of travellers to attributes including the role of psychological attributes. Advanced models will furthermore split places that are likely to generate and attract travel demand into further subgroups, and consider interactions with external factors such weather, time-of-day and seasonal effects.

Clearly though the consideration of multiple interactions between city attraction points, also known as Points of Interests (POIs) and such time factors, becomes quickly cumbersome. Furthermore, the effects of POIs are likely to have interdependencies. A specific area of a city is often attractive because of its accumulation of different POIs.

As part of this discussion, the need to estimate “point demands” in public transport also has to be highlighted. Whereas for traffic planning the required information is mostly the number of cars travelling from and to an area of the city, public transport planners instead need to understand the bus stop or train station demand. This difference highlights the need to discuss the “network of attraction” from a specific point. If this can be established, it will help the planner in the public transport network design stage, as well as the detailed planning of a public transport line.

Finally, the recent data developments motivate a discussion on the feasibility of machine learning approaches for demand estimation. In particular recent advances in digital maps, crowd-sourced data collection technologies, and availability of a range of tracking data are providing such detailed information about POIs, their static attributes as well as who visits them during different periods of the day. These data cannot anymore be appropriately analysed with “analytical” models. Commonly used by many researchers, OpenStreetMap data include hundreds of POI types. Furthermore, Google provides ratings and opening times for POIs. There is further a recent stream of research using “Google Popular times data” to analyse temporal demand patterns. Vongvanich et al. (2022) used regression models to understand which POIs are visited before people visit train stations. Similarly, other publicly available GIS data provide information about other specific attributes of an area such as geographical, sociodemographic or economic features. On the demand side, Application Programming Interface (API) data from transit planners and other applications provide a rich source of information on which (combinations of) POIs people visit before or after using transit.

In the next section it is discussed that machine learning - and in particular the introduced graph-based deep learning approach - is a tool that offers rich possibilities for doing so. We do not claim that this chapter provides a comprehensive framework for specific demand estimation for public transport with all these data but the broad framework should provide ideas to adapt it to specific problems. The mainstream research based on the introduced deep learning framework uses lagged demand or an attribute as one of the model inputs and obtains future predictions. It has to be noted that this family of prediction tasks differs from the aforementioned estimation problems as the demand scale is not unknown if the lagged status is fed to a model. The framework remains applicable if lagged information is removed from the model input. This converts a prediction problem to a joint estimation and prediction problem for which other time-dependent inputs are required, e.g. lagged population distribution of a city to predict travel demand.

III. THE EMERGING GRAPH CONVOLUTIONAL NETWORKS

A. A Generalized Deep Learning Framework and Methodology

Network structure is an inherent property of any transportation infrastructure and its related travel demand. A network or “graph” consists of nodes and edges and both can have multiple attributes or “features”. A node can be an aerial unit such as a census tract, a city zone, and a municipality if the prediction task is for national-level travel demand and infrastructure. It can also refer to transportation infrastructure such as a railway station and a bus stop. Among deep learning approaches, Graph Convolutional Networks (GCNs) can directly reflect network structures in the learning process. The key idea of GCNs is to convert data to computational graphs and to learn the spatial and topological dependencies between nodes according to their relationships. This is in contrast to “standard” neural networks where learning can be considered to be carried out on a complete graph. With increasing data availability and computational capacity, GCN-based deep learning approaches make it possible to take the so-far overlooked interactions between nodes into consideration. Therefore, they are expected to produce more precise predictions on node-specific characteristics.

In Fig. 3.1, a general framework is given based on the growing body of existing literature and some ongoing works conducted by the authors’ research group on the applications of GCNs to transportation studies, mainly for prediction problems. The framework provides an abstraction of the methodologies used by comprehensive literature, covering a variety of prediction tasks that can help city planners and transport operators to foresee the demand and infrastructure density.

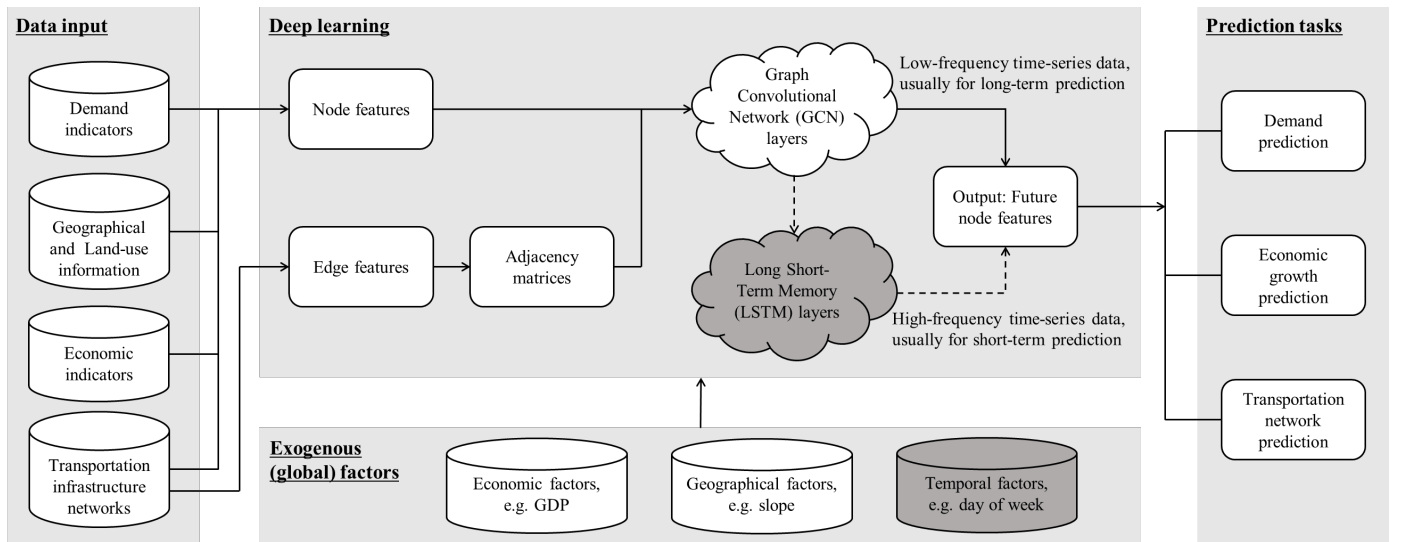


Figure 3.1 A general deep learning framework for demand and transportation infrastructure prediction
Source: Author provided

In “Data input”, the data used to obtain quantitative node and edge features for GCNs are listed. A node can be defined as an aerial unit or a transportation node according to the studied problem and the available data. Several types of node features have been considered useful. Demand indicators measure the travel demand related to an area, and it can be the population representing the potential travel demand or the observed travel demand. It can be based on a transportation node or aggregated for a spatial unit. Geographical and land-use information describes the land-use attributes of the area. It can be the number of a specific important facility or the distribution of various Points of Interest (POIs). Economic indicators quantify the capability of an area to develop transportation infrastructure and the busyness of social and economic activities, providing crucial added value for long-term predictions. Lastly, data of transportation infrastructure networks can be processed into both node and edge features. For node features, a straightforward way is to obtain the density of transportation infrastructure of an area, such as the density of a railway network in terms of station number and rail length. More specifically for public transport, the demand and infrastructure networks can be classified into several modes such as railway, tram, and bus. GCNs can also deal with multiple edge types and accordingly multiple adjacency matrices. This allows in-depth investigation into the interactions between different public transport modes. Moreover, if a road network is included, the cooperation and competition between public and non-public transport modes will become measurable.

In “Exogenous (global) factors”, both data for which node heterogeneity can and cannot be estimated are found. As an example for the former, when economic indicators are not available at node-level, Gross Domestic Product (GDP) can be used to describe the general economic situation of the study area. As an example of the latter, if a part of the studied country has unique geographical characteristics, it is essential to distinguish this with regional variables. For example, in a mountainous country, clearly travel demand

and transportation development will be influenced by the topography. Temporal and seasonal dummies are also important global factors (or the region of the study area is large). The demand on workdays and weekends can be significantly different, and so is the demand in summer and winter. Furthermore, in long-term prediction, it is necessary to distinguish special periods such as pandemics from “normal” periods.

In “Deep learning” in Fig. 3.1, the flowcharts for long-term and short-term predictions are distinguished. Long-term predictions, e.g. population predictions, usually require data frequency compromises as data is mostly not available for shorter intervals and/or the dynamics over shorter time periods are minor. Hence, also defining sub-models that can distinguish “recent” and longer-term dynamics are not significant and/or can not be established due to few temporal data points. A long-term prediction model that fits these characteristics will be introduced later.

In short-term prediction models with many temporal data points, introducing weights according to the recency of observations has become a common technique. A general term for this is the “long-short-term-memory” (LSTM) approach. A new trend is to include Temporal Convolutional Network (TCN) layers, which will be introduced with more details in the forthcoming examples. The detailed layer setting for GCN layers and LSTM layers is based on trial-and-error tests. The best layer settings are still an open-ended question. Therefore, the focus of the following section is mainly to explain the steps to prepare node features and adjacency matrices that form the links of the GCN framework, and to generalize the steps for a variety of problems for public transport planning.

B. Applications of the Deep Learning Framework: Long-term and Short-term Predictions

In this section, two examples of the aforementioned deep learning framework are provided. The first one is long-term railway network prediction using multiple time-series datasets over a century in Switzerland, including population data, road and railway network data, and topographical data. The second one is short-term bus travel demand prediction using one-month smartcard data, static road network and built environment data from China. They present two applications of the generalized framework to long-term and short-term prediction, respectively. An in-depth understanding of the framework and the GCN-based approaches is expected to be gained by going through these two examples in detail. The abstraction of the methodologies can be found in Fig. 3.2.

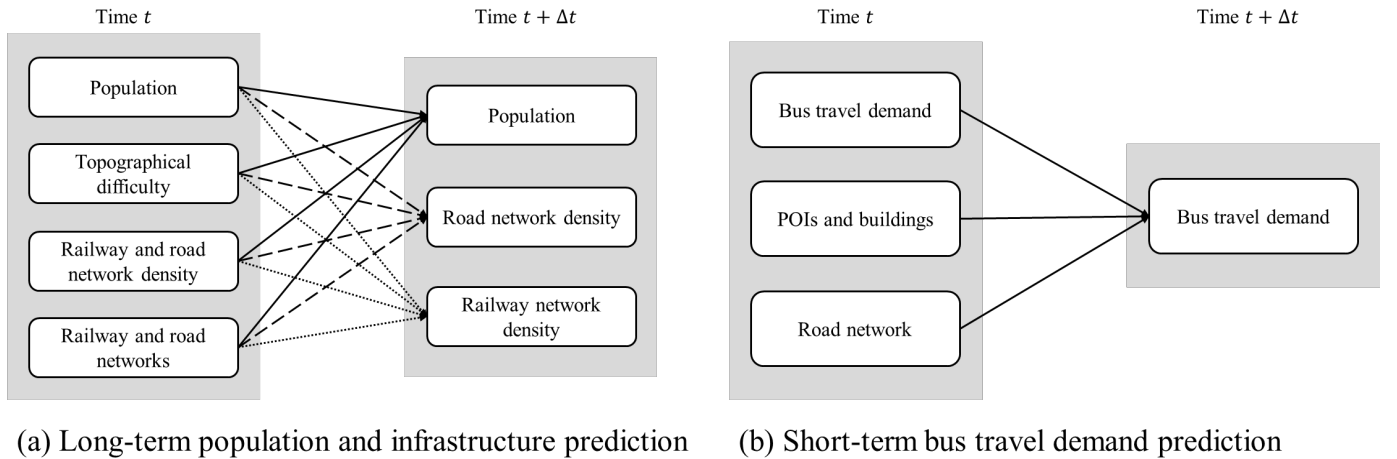


Figure 3.2 Two applications of the deep learning framework to forecast public transport infrastructure: (a) abstraction of the methodology of Lai et al. (2022), (b) abstraction of the methodology of Zhao et al. (2022)

Source: Author provided

As is illustrated in the framework, the initial steps are to prepare edge and node features from data and to create time-series graphs. For each time step t , an undirected graph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$, with N nodes $v_{i \in \{1, 2, \dots, N\}} \in \mathcal{V}$ and R types of edges $(v_i, v_j)_t^{(r)} |_{r \in \{1, 2, \dots, R\}} \in \mathcal{E}_t$ is created from data. Each node has F^{in} input features and a node feature matrix for the graph at time t is denoted as $\mathbf{X}_t \in \mathbb{R}^{N \times F^{in}}$. Edges can be written as R weighted adjacency matrices $\mathbf{A}_t^r \in \mathbb{R}^{N \times N}$, $r \in \{1, 2, \dots, R\}$. These adjacency matrices are model inputs and describe the existence of transportation networks or their qualities, e.g. travel time. To distinguish input and output, $\mathbf{Y}_{t+\Delta t} \in \mathbb{R}^{N \times F^{out}}$ is used to denote the output node feature matrix at time $t + \Delta t$. F^{out} is different from F^{in} as is shown in Figure 3.2 because the number of output node feature types is not necessarily identical to that of the input node feature. More specifically, $x_{i,k,t}$ represents

the k^{th} input node feature of node i at time t , where $k \in \{1, 2, \dots, F^{\text{in}}\}$. Instead $y_{i,k,t+\Delta t}$ represents the k^{th} node feature of node i at time $t + \Delta t$, where $k \in \{1, 2, \dots, F^{\text{out}}\}$. This notation will be used throughout this chapter.

1) Long-term Prediction of Population and Transportation Infrastructure Density

This example, based on Lai et al (2022), attempts to predict future population, railway network density, and road network density for 2,833 municipalities in Switzerland. Time-series population and network data in the 20th century are collected for each municipality. Complete population, road, and railway network data are available for the years 1910, 1930, 1950, 1960, 1970, 1980, 1990, and 2000. Although data is not available for every year, it is fairly sufficient for model training and testing with this sample size. Considering the frequency of the data and that significant network changes are not expected to occur in a short time interval, the time step for this prediction is 20 years, e.g. the node features in 1930 are supposed to be successfully predicted by using the data in 1910. To make full use of the data, six pairs are created: [1910,1930], [1930,1950], [1950,1970], [1960,1980], [1970,1990], [1980,2000]. For each pair, the earlier year denotes the year of input data and the later year is the year of prediction results. A planner can decide how to split the data into training and testing sets. It is intuitive to use the pairs in the early 20th century for training and those in the late 20th century for testing.

In this case, there are four types of input node feature, three types of output node feature, four types of edge, and 2,833 nodes so that $F^{\text{in}} = 4$, $F^{\text{out}} = 3$, $R = 4$, and $N = 2833$. Among the four types of node features, population can be directly obtained from the data for each municipality and for each year. Topographical difficulty is a special input for this case. The evolvement of transportation infrastructure and population of a country may show strong cross-region inequality as a result of the differences in regional attractiveness and ability to develop transportation infrastructure. This inequality may be exacerbated by the mountainous characteristics of Switzerland. A municipality in a mountainous area with high elevation may have difficulty in developing infrastructure and have low attractiveness to population immigration. In order to take this into account, the topographical obstacle and economic capability of a municipality are integrated as one factor constraining (the speed of) regional development and Lai et al (2022) refer to it as “topographical difficulty”, in the sense of “difficulty of developing transportation infrastructure”. The slopes for all the meshes within the polygon of a municipality are extracted by the digital elevation model. The 200-meter grid digital elevation model from the Federal Office of Topography “swisstopo” was used. By summing up the slopes and then dividing the summation by the area of that municipality, the average slope of a municipality is obtained to describe its topographical obstacle. Regarding economic capability, the spatial resolution of economic indicators such as Gross Domestic Product (GDP) may be not in the municipality unit, and it is simplified by a national indicator Gross National Income (GNI) also due to its fine availability in the full 20th century. Note that topographical difficulty is both node and time-dependent as it is a product of node-dependent slope which does not change over time and exogenous global time-dependent data. Let D_i^t denote the topographical difficulty of municipality i in time t , δ_i be the average slope in degree unit of municipality i , and G_t be the GNI in time t , D_i^t can be computed as in (1). Since difficulty is the second input node feature, D_i^t can also be written as $x_{i,2,t}$. It can be observed in Fig. 3.3 that the municipalities in southern Switzerland had much larger topographical difficulty than those in northern Switzerland, and the node difficulties in 1990 were much reduced compared to those in 1910 due to the increase in GNI.

$$D_i^t = \frac{\delta_i}{\log_{10} G_t}$$

(1)

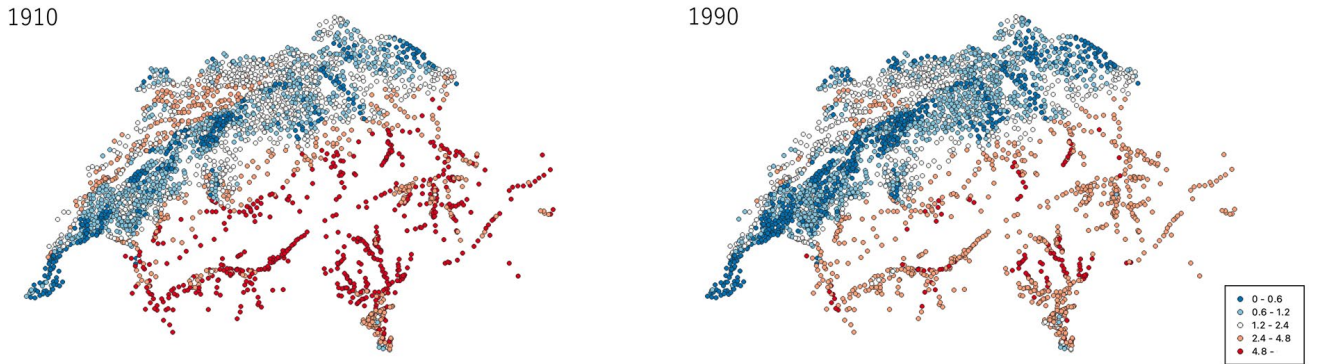


Figure 3.3 Topographical difficulty of 2,833 municipalities in Switzerland in 1910 and 1990

Source: Author provided, adapted from the Supplemental Material of Lai et al. (2022), permission obtained by authors

There are in general two ways to define transportation infrastructure density: station density and network density. In this case, network density is used. Road or railway network density of a municipality is defined as the distance of road or railway passing through it divided by its area. One can also multiply the distance-based density with speed or number of tracks to describe the actual capacity of transportation infrastructure.

After preparing node features, one will move on to edge features. In this case, four types of year-dependent adjacency matrices are input to the GCN framework. Fig. 3.4 shows the four types of edges: (a) neighbouring, (b) railway network and road network, (c) accessibility based on travel cost. Among them, the neighbouring edges are the simplest edge types. Each edge is a binary entry to the correspondent adjacency matrix and indicates whether two nodes are located in such a way that they can be reached without traversing another node. Further, road and railway networks are processed into integer entries to two other adjacency matrices. If a transportation link crosses the boundary between two municipalities, it is regarded as an edge connecting them. If there are multiple transportation links crossing the boundary between the same pair of nodes, the number of links will be accumulated as the weight in the adjacency matrix to show a stronger transportation connection, as is illustrated in Fig. 3.5. For the last edge feature, accessibility, a larger entry is fed to the adjacency matrix if the travel cost between two municipalities is lower. A strength of GCN is that it can handle a variety of types of node and edge features, and one can adjust model inputs according to data availability. With the development of transportation networks, the adjacency matrices of railway network, road network, and accessibility should be time-dependent. The neighbouring matrix may vary from time to time as well due to the changes in administrative boundary.

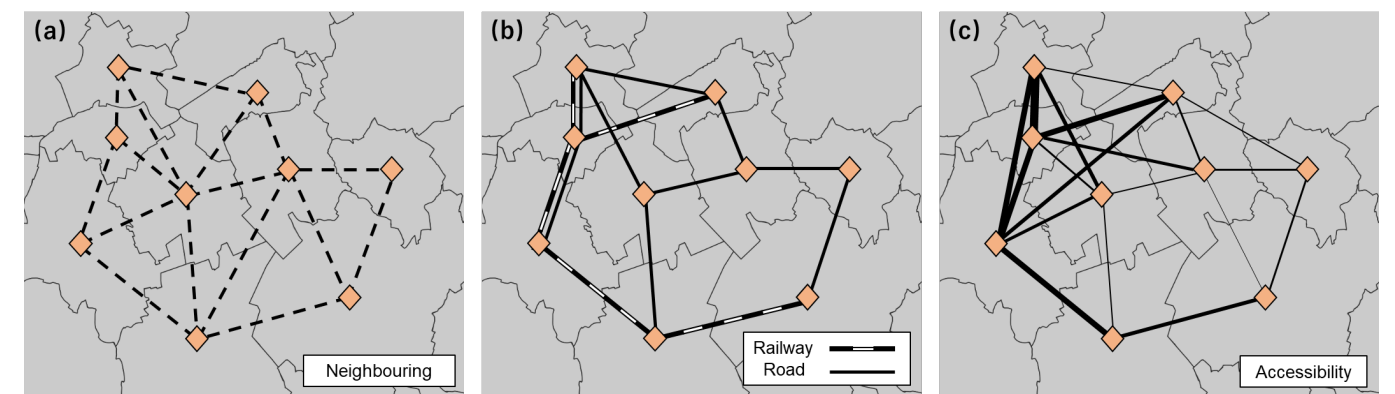


Figure 3.4 Four types of edge feature: (a) neighbouring, (b) railway network and road network, (c) accessibility based on travel cost

Source: Author provided, adapted from Lai et al. (2022), permission obtained by authors

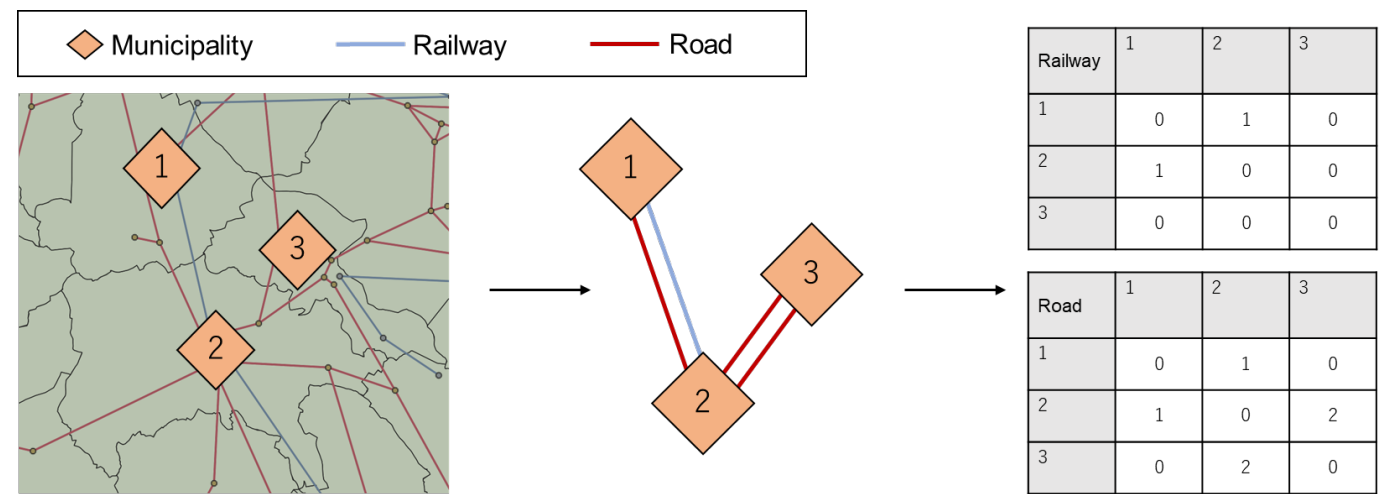


Figure 3.5 Data processing to convert road and railway network into adjacency matrices

Source: Author provided, adapted from the Supplemental Material of Lai et al. (2022), permission obtained by authors

2) Short-term Public Transport Travel Demand Prediction

This section is mainly based on two cases of using GCNs to predict short-term public transport travel demand, one for bus transit (Zhao et al., 2022) and the other for metro (He et al., 2022). Although the steps to tackle these short-term prediction tasks share common tasks, a number of differences should be noticed. In long-term prediction, it is always essential to determine nodes and edges for a graph. In short-term prediction, a graph structure is not time-varying since the changes in transportation networks are usually negligible and therefore not considered. As a result, for each time step t , the graph has the same nodes and edges so that the time step subscript is omitted. The graph $G = (\mathcal{V}, \mathcal{E})$, with N nodes $v_{i \in \{1, 2, \dots, N\}} \in \mathcal{V}$ and R types of edges $(v_i, v_j)^{(r)} \mid r \in \{1, 2, \dots, R\} \in \mathcal{E}$ is created from data. In bus travel demand prediction, which is similar to the previous case, a node is defined as a spatial unit instead of a bus stop. The node feature “demand” is aggregated from the demand of bus stops within each spatial unit. Prediction may perform better under this setting as the distribution of bus stop demand is fairly skewed. Some stops attract and generate few passengers, merely playing the role of route connector. This skewness of the distribution is amplified and exacerbated if the stop demand is fed to a stop-based origin-destination (OD) matrix. Bus-stop OD demand estimation in large networks where some stops have few demand is therefore more challenging, and zonal aggregation can ease the difficulty. A recent study by Zhuang et al. (2022) recognizes this “sparse stop-based travel demand prediction problem” and develops spatial-temporal GCNs considering zero entries. To note is that although describing bus transit demand generation and attraction on a zone level is more suitable for initial planning purposes (Hussain et al., 2021), explicit boundaries are not always available for small administrative spatial units that are meaningful for PT route planning. For some cities, official transport zones are available so that they are natural spatial units to represent bus travel demand generation and attraction, e.g. Brisbane Strategic Transport Model (BSTM) zones. Otherwise, one needs to separate a city into multiple units. In Zhao et al. (2022), municipal roads are used to separate the studied city into 1,139 street-level units by a Voronoi Diagram method. This zoning problem is also known as the “Modifiable Areal Unit Problem” (MAUP). Different ways to separate a city into zones may generate fairly different spatial units and demand aggregation. Other than zoning based on Voronoi Diagram, more straightforward and regular zoning methods are discussed in Ghadiri et al. (2019). They test four geometric shapes: square, rhombus, triangle, and hexagon to obtain traffic zones for Khoy City, Iran. Fig. 3.6 uses a simple square method to separate a bus network in Shizuoka City, Japan. Two issues related to the catchment of a bus network should be noted. Firstly, as can be seen in the figure, many bus stops are on or close to the border of separations, and it is arguable which mesh such a stop should be associated with. Secondly, many meshes have no bus stop. However, it does not necessarily mean that the residents in these meshes do not use bus transit. A separation method using irregular shapes based on bus transit accessibility or bus service availability is desired. Ideally, each separated spatial unit should have bus stops. The purpose and evaluation index of zoning can vary from case to case. In the case of demand prediction, it is proper to employ prediction accuracy as an evaluation index on the reasonableness of zoning. It will be interesting if one can further investigate and compare the prediction accuracies resulting from different zoning methods. Insights can be obtained for the open-ended question on the appropriate zoning and spatial unit for origin-destination representation and public transport planning.

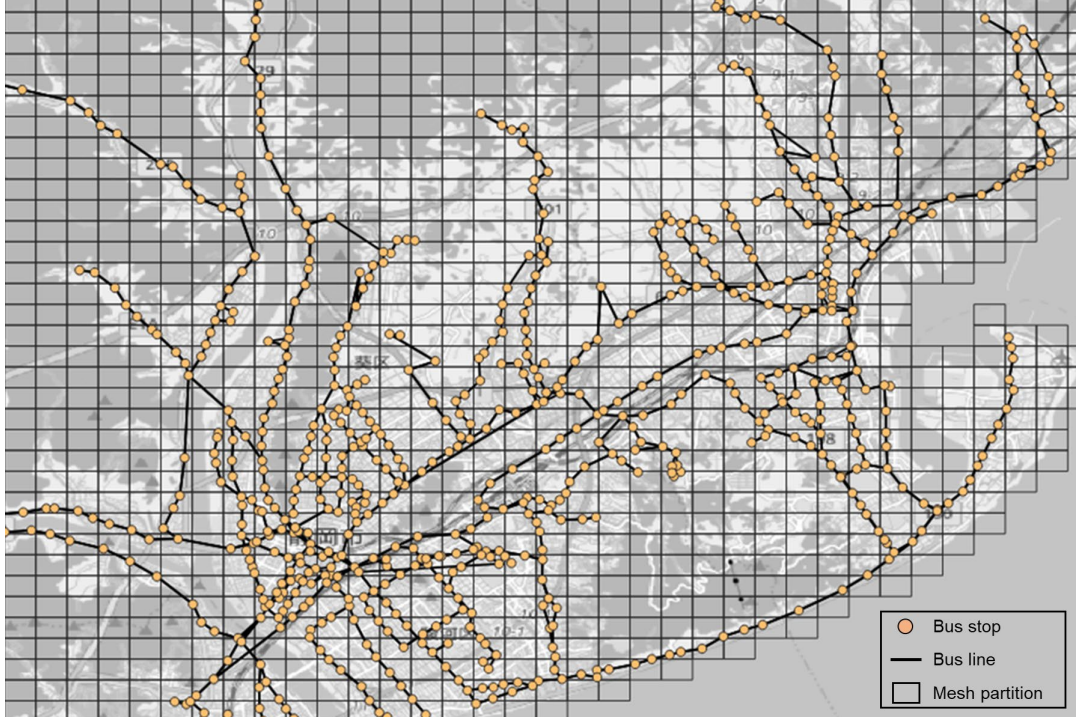


Figure 3.6 A naïve 500m² mesh separation for a bus network in Shizuoka City, Japan, illustrating some of the zone choice problems.

Source: Author provided

After defining the spatial unit for a node, aggregated time-series bus travel demand, which is the fundamental node feature, and the other node features can be obtained. Different from the long-term and “macroscopic” example where GNI and topographical difficulty are used to describe node heterogeneity, land-use attributes are employed in this “microscopic” case. More specifically, the numbers of different types of buildings and Points of Interest (POIs) in a spatial unit are used as node features. The numbers of bus stops, metro stations, enterprises, shopping malls, attractions, schools, and hospitals are used. The richness of node feature inputs regarding facilities is dependent on the digital map services from which the data is collected. For example, by using Google Place API, 96 types of POIs can be collected (Google, https://developers.google.com/maps/documentation/places/web-service/supported_types), and more trivia facilities can be obtained from OpenStreetMap. Extra effort of categorizing or screening POI types to get a meaningful and manageable number of node feature inputs is then required. Some POI-based node features will be time-independent, e.g. the number of historic buildings, whereas others can be further processed into a time-of-day dependent POI, for example considering business hours of shops. This processing is expected to enhance the relationship between POI node features and the bus travel demand for a node features since a traveller will not visit a POI if it is closed.

There are a number of options for edge features. Zhao et al. (2022) employ a binary edge feature based on whether two spatial units share the boundary, namely, a neighbouring edge feature which is also used in Lai et al. (2022). Similar to Lai et al. (2022), edge features extracted from road networks, public transport networks, and travel costs between two nodes can also be used. Multi-relational GCNs allow a model to include multiple types of edge features. Lai et al. (2022) simply add up these relations for a node, though recent research by Liang et al. (2022) introduces more parameters to make the model learn the contribution of each relation. They conduct a short-term prediction on multimodal travel demand considering multiple transport networks including subway and ride-hailing.

The larger and the more significant demand points are, the more feasible it becomes to avoid zoning questions. In the case of metro and train station demand prediction, often the station can be directly used as a node, as in He et al. (2022). Station-based travel demand and POI numbers within a radius of each station then become node features. Edge features and adjacency matrices are extracted from the metro network topology based on inter-station connectivity, distance, and travel time.

Regarding exogenous global factors, time-of-day, day-of-week, and special event dummies are commonly used. These can help to distinguish demand patterns in commuting/non-commuting hours, on weekdays/weekends, and irregular demand patterns triggered by special events such as festivals. Event dummies can be also used to capture the impacts of COVID-19 regulations on public transportation service provision and demand.

A key difference in model structure between long-term and short-term prediction models is the introduction of temporal layers. Recurrent layers such as those used in LSTM can capture the temporal dependencies of node features and these have been applied successfully in a number of studies (e.g. Ke et al. 2017 on short-term prediction on ride-hailing demand and the metro case by He et al. (2022)). Latest research employs temporal convolutional network (TCN) layers to capture these temporal dependencies. Zhao et al. (2022) use both TCN and LSTM, and Liang et al. (2022) use TCN only.

IV. MODEL INTERPRETATION WITH SHAPLEY ADDITIVE EXPLANATIONS

For some applications, one might accept that deep learning approaches are unexplainable as long as they can make more accurate predictions than simpler, explainable models. However, in the predictions of travel demand and transportation infrastructure, making deep learning models explainable is of great value as understanding the driving forces of travel demand and changes in transportation infrastructure is desired by city planners and transport operators. A theory of “Shapley Additive Explanations” (SHAP), first proposed by Lundberg and Lee (2017), sheds light on the possibility of explainable deep learning. SHAP adapts the original concept of Shapley values in game theory to machine learning prediction models. Each input feature value of the instance is equivalent to a player in the game and the output from the prediction model is the pay-out. SHAP values then can decompose the prediction output and attribute the contribution to each input feature in a particular prediction made by various types of models such as tree-based machine learning models and deep learning models. Though SHAP values are increasingly used in a wide range of applications, from machine learning approaches e.g. gradient boosting regression through to deep learning approaches (Feng et al. 2021; Zhang et al. 2020), the application within the context of a network structure is still lacking. A quantitative method to explain GCNs by decomposing an output node feature value and distributing the contributions to each input node feature is discussed in this chapter. The method is based on *DeepExplainer* (Lundberg & Lee, 2017) which in turn is based on *DeepLIFT*. It is a fast approximation method of the contribution of each feature by decomposing the output from a neural network for a certain input by backpropagation and explaining the difference between the output and a “reference output” obtained by using a reference input in the original model (Shrikumar et al., 2017). For example, in image recognition models, *DeepExplainer* can find the contribution of pixels that increase or decrease the model’s output, by using large numbers of images as reference input. However, *DeepExplainer* is not applicable for handling deep learning models with graph-structured data due to the inadequacy of reference graphs. Therefore, a modified *DeepExplainer* called *GraphExplainer* which generates reference inputs by random permutation of the original input is developed.

The pseudocode of *GraphExplainer* is provided below. In this algorithm, 100 reference node feature matrices (i.e. $Q = 100$) are generated. Each reference matrix randomizes the index of the elements of each input vector. The reference matrices are used to obtain the “reference output” Y_{ref} . It is the expected output regardless of the input features. All the nodes in the network are defined as a target. Based on these, matrices $grad(X_{ref})$ representing the gradient of $selected_{ref}$ with respect to $\mathbf{X}_{ref}$, are obtained. The SHAP value is obtained by the mean change in the output considering the distance from the reference solution.

Algorithm: *GraphExplainer*

Input: ind_{target} (target node index), $\mathbf{A}$ (adjacency matrix), $\mathbf{X}$ (node feature matrix)

Get model output: $Y = model(\mathbf{A}, \mathbf{X})$

for ind in ind_{target} do

 Generate Q randomized index lists $\{O_1, O_2, \dots, O_Q\}$

 Initialization: $\phi = 0$, $diff_{shap} = 0$, $diff = 0$

 Get target output: $selected = Y(ind)$

 For j in $\{1, 2, \dots, Q\}$ do

 Generate reference node feature matrix: $\mathbf{X}_{ref} = \mathbf{X}(O_j)$

 Get reference output: $Y_{ref} = model(\mathbf{A}, \mathbf{X}_{ref})$

 Get reference target output: $selected_{ref} = Y_{ref}(ind)$

 Compute the gradient of $selected_{ref}$ w.r.t all variables: $selected_{ref}.backward()$

$\phi = \phi + grad(X_{ref}) * (X - X_{ref})$, where $grad(X_{ref})$ is the gradient obtained above w.r.t variable X_{ref}

$diff = diff + selected - selected_{ref}$

$diff = diff / Q$

$\phi_{ind} = \phi / Q$

$diff_{shap} = sum(\phi_{ind})$

Output: $\phi_{ind} \in \mathbb{R}^{N \times F}$

This *GraphExplainer* has been tested by Lai et al. (2022). With the mathematical definition and some results in the next section, this chapter aims to provide more precise explanations of this theory of explainable deep learning with SHAP values. Let $\phi_{x_{i^*, k^*, t}^{y_{i, k, t+\Delta t}}}$ denote a specific SHAP value. A specific predicted future node feature $y_{i, k, t+\Delta t}$ can be decomposed into $N \times F^{in}$ SHAP values according to the number of input node feature at time t and the specific reference output $\tilde{y}_{i, k, t+\Delta t}$, so that $x_{i^* \in \{1, 2, \dots, 2833\}, k^* \in \{1, 2, 3, 4\}, t}$ indexes each SHAP value and $\phi_{x_{i^*, k^*, t}^{y_{i, k, t+\Delta t}}} \in \mathbb{R}^{N \times F^{in}}$. $\phi_{x_{i^*, k^*, t}^{y_{i, k, t+\Delta t}}}$ measures the contribution made by the k^{th} feature of node i^* at time t to the prediction of the k^{th} feature of node i at time $t + \Delta t$. According to the additive nature of SHAP, the predicted value for the k^{th} feature of node i at time t is the summation of related SHAPs and the baseline as in (2). It is important to note that i^* can take the same value as i . In this case, $\phi_{x_{i, k^*, t}^{y_{i, k, t+\Delta t}}}$ is termed “self-SHAP” and defines the contribution made by the k^{th} feature of node i^* itself at time t . Hence, in particular in slowly changing networks, it is expected to have a larger magnitude than the SHAPs of other nodes, which we term “influential-SHAP”. For better interpretability of the results, we distinguish the sum of self-SHAPs and the sum of influential-SHAPs in (3). The first term on the right of (3) is the sum of the four self-SHAPs of the node features, and the second term is the sum of influential-SHAPs of the features of other nodes. It should be noted that producing SHAPs is time-consuming. Possessing a good Graphics Processing Unit (GPU) is essential, but even then it takes around 80 seconds for a computer equipped with GeForce RTX 3060 Ti to produce $\phi_{y_{i, k, t}} \in \mathbb{R}^{2833 \times 4}$ for one prediction value $y_{i, k, t}$ so that it requires more than 60 hours to obtain all the SHAPs for predicting a specific vector $\mathbf{y}_{k, t}$.

$$y_{i, k, t} = \sum_{i^*=1}^N \sum_{k^*=1}^F \phi_{x_{i^*, k^*, t}^{y_{i, k, t+\Delta t}}} + \tilde{y}_{i, k, t+\Delta t} \quad (2)$$

$$y_{i, k, t} = \sum_{k^*=1}^F \phi_{x_{i, k^*, t}^{y_{i, k, t+\Delta t}}} + \sum_{i^* \neq i} \sum_{k^*=1}^F \phi_{x_{i^*, k^*, t}^{y_{i, k, t+\Delta t}}} + \tilde{y}_{i, k, t+\Delta t} \quad (3)$$

V. APPLICATIONS AND SHAP INTERPRETATIONS, SWITZERLAND EXAMPLE

A. Prediction Accuracy

Prediction accuracy is usually the strength of deep learning approaches. This section showcases the SHAP interpretation based on results obtained from the long-term prediction application in Switzerland. Before that, attention is called for an accuracy trap. Recall that from the six pairs [1910,1930], [1930,1950], [1950,1970], [1960,1980], [1970,1990], [1980,2000], two split patterns are designed. In forward prediction, the former four pairs are used as training data and the latter two are used as testing data. In backward prediction, the latter four pairs are used as training data and the former two are used as testing data. Fig. 5.1 shows some prediction results. Railway network density is excluded from the result presentation because the railway network did not change much within the studied timeline. In this Switzerland example, prediction accuracy in terms of R^2 , which is the square of the correlation coefficient between observations and predictions, is acceptable for all cases.

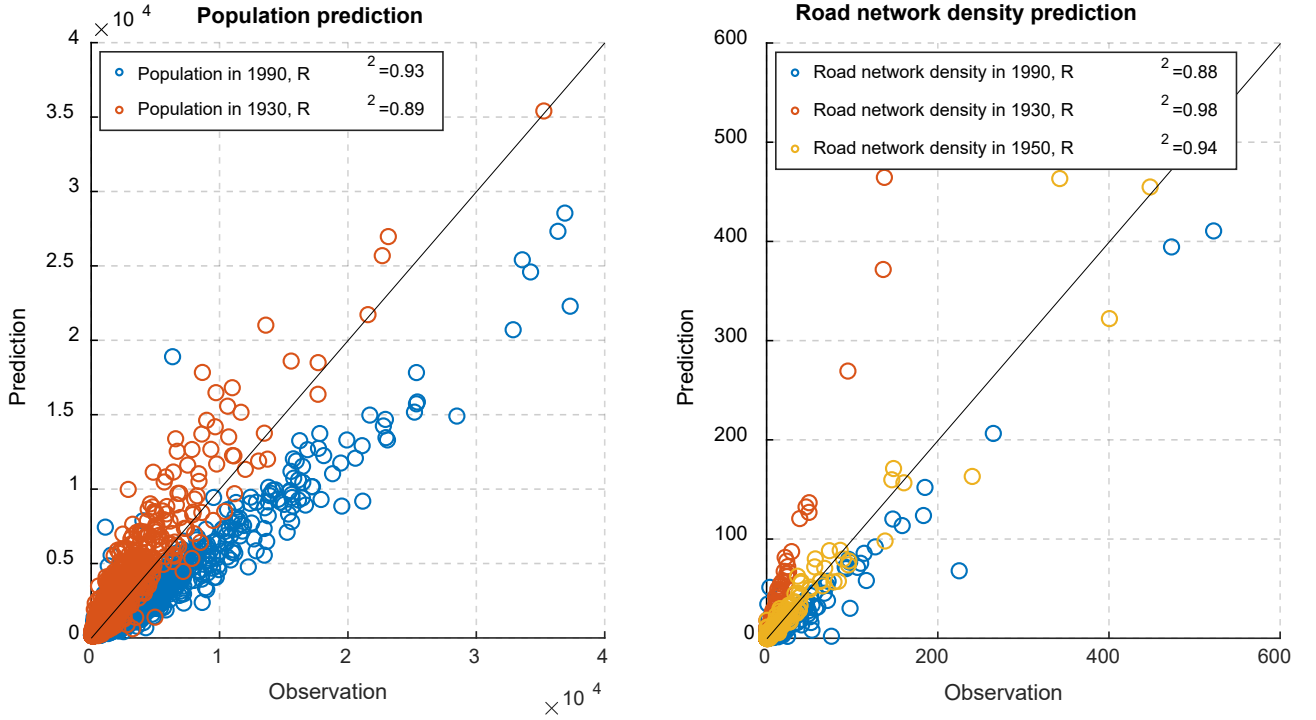


Figure 5.1 Prediction accuracy by GCN

Source: Author provided

Nevertheless, obvious over- and under-estimations can be observed. The population in 1990 is predicted by forward prediction using the data in 1970 as input and the GCN model trained by the data in the early 20th century, and the population in 1930 is predicted backward prediction by using the data in 1910 and the GCN model trained by the data in late 20th century. With this, it is not difficult to understand the over- and under-estimations. The model training captures the mean increase rate of population and road network density and expects that the development of population and road network density can maintain this speed for the testing datasets, ignoring the fact that neither population nor road network density increases in a linear way. Consequently, learning the population increase pattern from 1910 to 1980 makes the model underestimate the population in 1990 as the population in Switzerland experienced a steeper increase from 1950 to 1990 than from 1910 to 1950. Reversely, learning from the period having a steeper population increase slightly overestimated the population in 1930. For road network density, the road network rapidly developed after the 1930s in Switzerland and this high rate was maintained till the 1970s. The model therefore significantly overestimates the road network density in 1930 by applying the high development rate learned from 1950 to 2000, and fits the road network density in 1950 well with this appropriate development rate. By viewing this shortcoming of not considering temporal relations in this long-term prediction GCN structure, the introduction of temporal deep learning layers appears to be necessary. However, as discussed before,

the LSTM and TCG layers used for short-term predictions are not suitable for long-term predictions due to the limited number of time points for which data are available. Furthermore, the time step for which data are available is inconsistent. This issue forces the time step for learning to be long, e.g. 20 years in this case so that each time step reflects rather a general “era” than a specific year. In fact, in our example, the temporal relation coincides with development stages and unique events e.g. steeper development speed after 1950 receiving the after-war effect of World War II. In addition, saturation issues of population and infrastructure development also need to be taken into account. It is essential to identify the turning points for fast growth or saturation in the long-term prediction of population, travel demand, and infrastructure density. Proper temporal deep learning layers are worth attention and further exploration.

B. SHAP Interpretation

1) Self-SHAPs

SHAP values are calculated to interpret the prediction results. Self-SHAPs are found to be more influential than influential SHAPs and strong correlations can be observed between self-SHAPs and self-features. Self-features are the input data in the previous time step and self-SHAPs are the contribution to the prediction results by self-features in the target time step. The relationship between the self-SHAP per feature of a target node and the specific feature instance of the specific node in the prediction of the population is shown in Fig. 5.2. In mathematical terms, it is the relationship between $\phi_{x_{i,k^*,t}}^{y_{i,1,t+\Delta t}}$ and $x_{i,k^*,t}$. Each point in this figure represents a node i and each subplot denotes a node feature type k^* . Fig 5.2(a) illustrates a clear linear correlation that a larger population results in a larger contribution of a population feature to future population prediction of a node. On the contrary, a higher density of railway infrastructure or better accessibility to railway of a municipality may cause a reduction in future population. Taking Figs 5.2(a) and (b) together, the linear correlation between self-SHAPs and self-features of population and railway, either positive or negative, was found stronger in the early 20th century. It shows that in the late 20th century, the future population of a municipality was less determined by its own population and less decreased by its own railway network density. This might indicate an increasing tendency to move away from municipalities with a large population and a decreasing tendency to move away from municipalities that have a well-developed railway network. In Fig. 5.2(c), the trend of how road network density influenced future population changed drastically at the beginning and end of the 20th century. In the beginning, a high road network density tended to increase population sharply while it failed to increase and even reduced population in the late 20th century. Taken together, the findings illustrate that nowadays residents are more likely to move out of the municipalities with better accessibility as it becomes easier to commute from other places. The main factor driving the population to move out has shifted from railway to road network. So far, the sensitivity of prediction output to data input in this GCN example shows that little node heterogeneity exists in the population proportion predicted by lagged population and network density. One therefore may consider using linear regression to capture these relationships between the target node population and lagged node features. The analysis on difficulty further shows the strength of GCN that can capture node heterogeneity. It is interesting that the impact of lagged difficulty on population shows two diverging trends in the early 20th century and one of them disappears in the late 20th century. The municipalities presenting a more obvious negative correlation between self-SHAP of difficulty and lagged difficulty in the early 20th century are found generally close to railway stations. A reasonable hypothesis could be, that favourable accessibility to railway might give the residents in a municipality having unfavourable difficulty more opportunities and freedom to immigrate elsewhere. The desire for immigration vanished in the late 20th century in that the immigration had been completed and the inequality in difficulty became less pronounced. These interpretations generated by SHAPs shed light on the drivers of regional development, helping planners to understand the reasons behind the increase in population, transport demand, and transportation infrastructure.

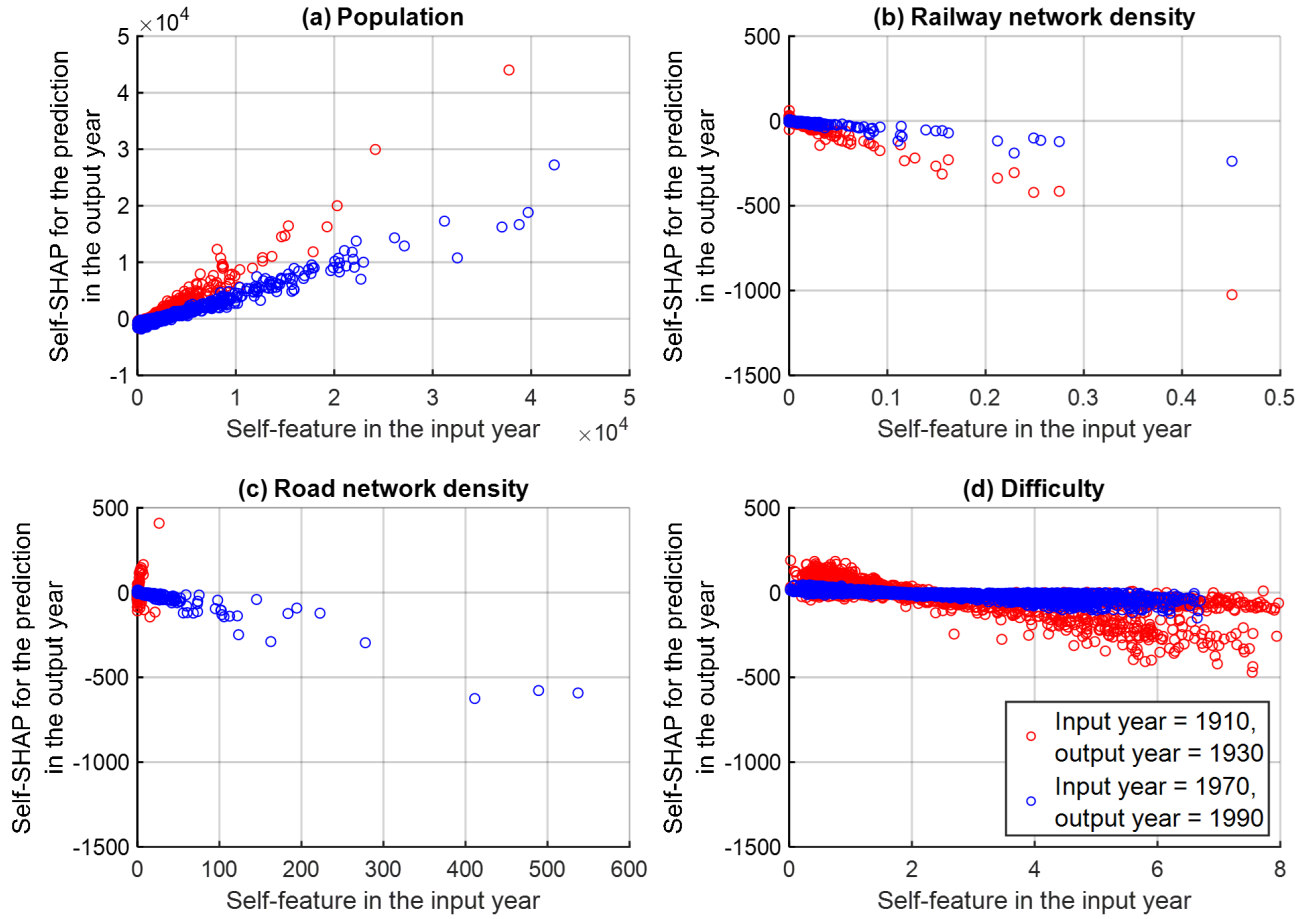


Figure 5.2 The relationship between self-SHAP values and self-feature values in the prediction of population

Source: Author provided

2) Influential-SHAPs

Recall from Eq. (3) that prediction can be divided into self-SHAPs and influential SHAPs. As shown in Fig. 5.3(a), the *GraphExplainer* algorithm can decompose the predicted feature of a target node into the proportion claimed by its own lagged node features and the proportion contributed by the node features of other nodes, considering network topology structure. It can be seen in Fig. 5.3(a) that self-SHAPs dominate the prediction for the target node, although the interpretations obtained from influential-SHAPs are also noteworthy and will be discussed in this section.

Two municipalities, La Chaux-de-Fonds and Le Locle in the canton of Neuchâtel, are selected to illustrate the interpretability of SHAP results. These two municipalities are selected as they have close relationships and coherent urban planning by the virtue of their watchmaking industry as early as in the 18th century. The industry flourished before the 1930s crisis and after the war. Later it has been influenced by the arrival of quartz crystal and electronic techniques but then the industry stabilized since the 1970s (ICOMOS, 2009).

Figs. 5.3(c) and (d) show that the population of both municipalities can be predicted to a large degree by their lagged population, though the dependency between the current and future population becomes weaker in the late 20th century. A negative contribution from the population of La Chaux-de-Fonds to the future population in Le Locle can be found in Fig. 5.3(d). This is an indication of the coherent development of the watchmaking industry in these two municipalities, a labor movement between them can be further inferred.

The dependencies of the railway transport development in La Chaux-de-Fonds on other municipalities in the 20th century are illustrated in Figs. 5.3(e) and (f). As expected, difficulty is the dominant influential factor of railway development. The railway network development in La Chaux-de-Fonds was influenced negatively by its own difficulty and Le Locle's difficulty, though much

less, in both early and late 20th century. In late 20th century, the railway development in La Chaux-de-Fonds was further limited by its own railway density, which indicates the saturation of its railway development. Different from the railway network, the future road network density of La Chaux-de-Fonds was highly but negatively dependent on its own population and almost independent of difficulty, as can be observed in Figs. 5.3(g) and (h). The road network development in La Chaux-de-Fonds appears to be supported by the road network density of Le Locle but slowed down by its own road network density (saturation effect). One can hence interpret this that the strengthening of the inter-city connections has been a motivation of developing the road network of La Chaux-de-Fonds, especially in the late 20th century. Finally, Figs 5.3(h) shows a competition effect between the road and railway network in the late 20th century in La Chaux-de-Fonds. The figure suggests that the future road network density is negatively influenced by the railway network density.

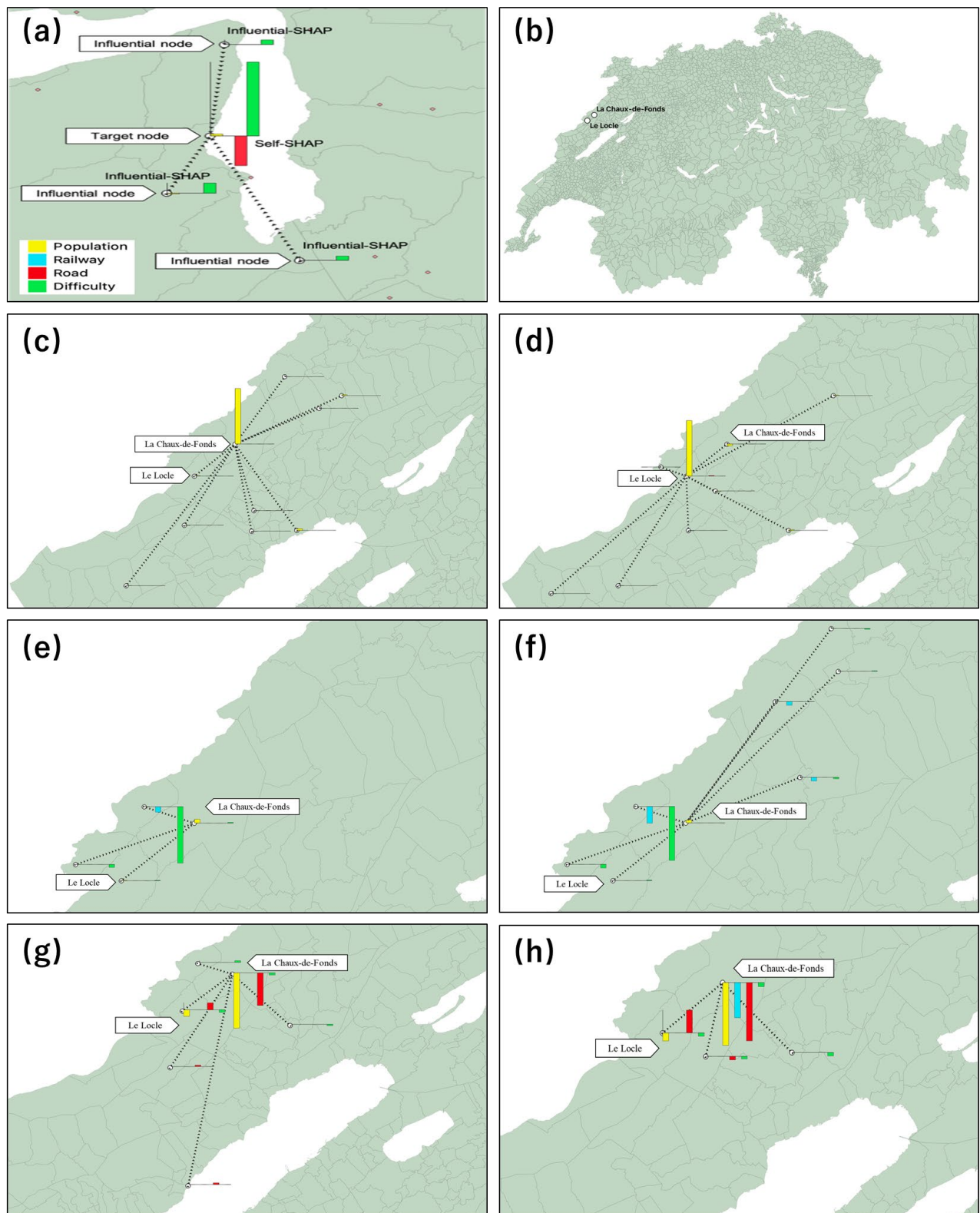


Figure. 5.3 Geographically illustrating the contribution of the node features of different municipalities to the prediction of a target node by SHAPs. (a) An example illustration with a legend (b) Location of case study municipalities: La Chaux-de-Fonds and Le Locle (c) Target: La Chaux-de-Fonds, population, 1930 (d) Le Locle, population, 1930 (e) Target: La Chaux-de-Fonds, railway density, 1930 (f) La Chaux-de-Fonds, railway density, 1990 (g) La Chaux-de-Fonds, road density, 1930 (h) La Chaux-de-Fonds, road density, 1990

Source : Author provided

VI. CONCLUSIONS

Graph-based deep learning approaches are drawing increasing attention from both academic researchers and industrial practitioners, and they have been applied to solve a number of problems in public transport planning and operation, in particular for short-term travel demand prediction. As the deep learning approaches are advancing rapidly, this chapter attempts to bridge the gap between state-of-art GCN studies and those who are interested in GCN approaches but with limited background knowledge. This chapter provides a general framework that accommodates research problems of different scopes ranging from long-term to short-term predictions and from bus travel demand to metro travel demand. In summary, the main appeal of GCN approaches is to use a graph to describe the studied problem and the information available. A graph has two essential components: node and edge. The scale of the studied area can vary from country to city. Nodes and edges are naturally available in transportation networks, e.g. a railway station as a node and a railway link as an edge for a railway network. However, more common cases require extra effort to separate an area into spatial units and take each as a node. Then edge features can be aggregated based on this spatial unit. Several examples using natural station nodes or tailored spatial unit nodes are discussed. The chapter discussed and gave examples for choosing suitable node features considering the spatial scale for which estimations and/or predictions are made. In the prediction of municipality population, city-level population and infrastructure density are used to characterise a node. Land-use attributes of a spatial unit such as the number of different types of POIs, that are increasingly available from sources such as OpenStreetMap, are more suitable for “microscopic problems” of bus travel demand prediction. It needs to be emphasised that this chapter partly could not remain general with respect to zone size and node feature choice in a GCN framework, as it is a trial-and-error process based on testing prediction accuracy.

The majority of prediction approaches using GCNs include the lagged status of the node feature to predict its future status as one of the data inputs. This limits the applications of GCN-based approaches. For future directions, the potential of GCNs for demand estimation is worth exploring. The discrepancy between demand estimation and prediction mainly relies on whether the demand information is included in the model input. Estimating travel demand size for an area from its features excluding demand attributes, e.g. bus travel demand estimation from merely POI distributions, is expected to be solved by GCNs. An interesting related work by Koca et al (2021) estimates taxi origin-destination demand in New York from a graph adapted from OpenStreetMap data. A similar OD estimation problem is studied by Rong et al. (2021), using graph-based deep learning, POI data, and population distribution. A joint task of estimation and prediction can be explored using GCNs and it will be interesting to investigate to which degree the accuracy will be downgraded compared to a sole prediction task.

A further key discussion of this chapter has been the recent effort to make deep learning approaches explainable through SHAP values. As a player can claim the reward according to the contribution in game theory, a node feature input can claim its contribution to the target prediction result by SHAP. A new algorithm to interpret a GCN model and related results was provided in detail. A prediction result is decomposed into pieces in line with the number of input features, and this decomposed value is called SHAP value. To obtain more interpretability, SHAP values of a GCN model can be distinguished either as self-SHAPs or influential-SHAPs. This can better classify the interactions between nodes of a network structure. Valuable insights can be derived on whether the demand of a spatial unit is self-driven or driven by other units close to or having good connections to this unit. This is important for understanding spatial interactions and correlations and helps to make deep learning more explainable and helpful for public transport planning and policy-making.

It can be foreseen that complications of GCN layer structures will continue to be explored by the research community. This is identified as a research direction focusing on incremental accuracy. Given that the existing complicated GCN structures containing temporal and spatial convolutional layers have already generated fairly high accuracy (95% or above for some complex multimodal demand prediction problems), the room for improvement is limited. A more exciting research direction is hence to further explore approaches that improve the interpretability of GCN. Though the chapter showed the potential of SHAP-based theory, it also has several limitations. First, SHAP values are dependent on prediction results and are not suitable to describe the relationship between input changes and output changes as the relationship may vary from two conducts of prediction using the same model. Second, it is difficult to define reference values for a graph. Last but not least, the computation of SHAP values is time-consuming. These limitations of SHAP-based methods are expected to be addressed and new theories to be developed in the future.

REFERENCES

- Ceder, A. (2016). *Public transit planning and operation: Modeling, practice and behavior*. CRC press.
- Feng, D. C., Wang, W. J., Mangalathu, S., & Taciroglu, E. (2021). Interpretable XGBoost-SHAP machine-learning model for shear strength prediction of squat RC walls. *Journal of Structural Engineering*, 147(11), 04021173.
- Gentile, G., & Nökel, K. (2016). *Modelling public transport passenger flows in the era of intelligent transport systems*. Springer Nature.
- Ghadiri, M., Rassafi, A. A., & Mirbaha, B. (2019). The effects of traffic zoning with regular geometric shapes on the precision of trip production models. *Journal of Transport Geography*, 78, 150-159.

- Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H. (2019, July). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33, No. 01, 922-929.
- Hart, S. (1989). Shapley value. In *Game theory* (pp. 210-216). Palgrave Macmillan.
- He, Y., Li, L., Zhu, X., & Tsui, K. L. (2022). Multi-Graph Convolutional-Recurrent Neural Network (MGC-RNN) for Short-Term Forecasting of Transit Passenger Flow. *IEEE Transactions on Intelligent Transportation Systems*, 1-20.
- Hussain, E., Bhaskar, A., & Chung, E. (2021). Transit OD matrix estimation using smartcard data: Recent developments and future research challenges. *Transportation Research Part C: Emerging Technologies*, 125, 103044.
- ICOMOS. (2009). La Chaux-de-Fonds/Le Locle (Switzerland) No 1302. UNESCO, Advisory Body Evaluation. Retrieved from <<https://whc.unesco.org/document/152298>>.
- Ke, J., Zheng, H., Yang, H., & Chen, X. M. (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation research part C: Emerging technologies*, 85, 591-608.
- Ke, J., Qin, X., Yang, H., Zheng, Z., Zhu, Z., & Ye, J. (2021). Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network. *Transportation Research Part C: Emerging Technologies*, 122, 102858.
- Koca, D., Schmöcker, J.-D. and Fukuda, K. (2021). Origin-destination matrix estimation by deep learning using maps with New York case study. 7th International IEEE Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). June 16-18, Crete, Greece. Published in IEE Xplore. doi:10.1109/MT-ITS49943.2021.9529275.
- Lai, Y., Sun, W., Schmöcker, J.D., Fukuda, K., & Axhausen K.W. (2022). Explaining a century of Swiss regional development by deep learning and SHAP values. *Environment and Planning B: Urban Analytics and City Science*, August 2022. doi:10.1177/23998083221116895.
- Lee, K., & Rhee, W. (2022). DDP-GCN: Multi-graph convolutional network for spatiotemporal traffic forecasting. *Transportation Research Part C: Emerging Technologies*, 134, 103466.
- Liang, Y., Huang, G., & Zhao, Z. (2022). Joint demand prediction for multimodal systems: A multi-task multi-relational spatiotemporal graph neural network approach. *Transportation Research Part C: Emerging Technologies*, 140, 103731.
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in neural information processing systems*, 30.
- Ortúzar, J., & Willumsen, L. G. (2011). *Modelling transport*. John Wiley & sons.
- Rong, C., Li, T., Feng, J., Yang, H., Geng, L., & Li, Y. (2021). Inferring Origin-Destination Flows from Population Distribution. *IEEE Transactions on Knowledge and Data Engineering*. doi: 10.1109/TKDE.2021.3075928.
- Ryu, S., Lim, J., Hong, S. H., & Kim, W. Y. (2018). Deeply learning molecular structure-property relationships using attention-and gate-augmented graph convolutional network. arXiv preprint arXiv:1805.10988.
- Shapley, L. (1953). Quota solutions op n-person games. Retrieved from <<https://apps.dtic.mil/sti/citations/AD0604085>>.
- Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning Important Features Through Propagating Activation Differences. 34th International Conference on Machine learning. PMLR 70:3145-3153.
- Sun, W., Schmöcker, J.D. (2021). Demand estimation for public transport network planning. In C. Mulley, J. Nelson, & S. Ison (Eds.), *The Routledge Handbook of Public Transport* (pp. 289–305). Routledge.
- Vongvanich, T., Sun, W., & Schmöcker, J.D. (2022). Explaining station demand patterns using Google Popular Times data. *The 15th International Conference on Advanced Systems in Public Transport (CASPT2022) and 8th TransitData conference*.
- Vuchic, V. R. (2017). *Urban transit: operations, planning, and economics*. John Wiley & Sons.
- Wu, L., Sun, P., Hong, R., Fu, Y., Wang, X., & Wang, M. (2018). SocialGCN: An efficient graph convolutional network based model for social recommendation. arXiv preprint arXiv:1811.02815.
- Yu, B., Yin, H., & Zhu, Z. (2017). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875.
- Yu, B., Lee, Y., & Sohn, K. (2020). Forecasting road traffic speeds by considering area-wide spatio-temporal dependencies based on a graph convolutional neural network (GCN). *Transportation research part C: emerging technologies*, 114, 189-204.
- Zhang, K., Xu, P., & Zhang, J. (2020). Explainable AI in deep reinforcement learning models: A SHAP method applied in power system emergency control. In *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, 711-716, IEEE.
- Zhao, T., Huang, Z., Tu, W., He, B., Cao, R., Cao, J., & Li, M. (2022). Coupling graph deep learning and spatial-temporal influence of built environment for short-term bus travel demand prediction. *Computers, Environment and Urban Systems*, 94, 101776.
- Zhuang, D., Wang, S., Koutsopoulos, H. N., & Zhao, J. (2022). Uncertainty Quantification of Sparse Travel Demand Prediction with Spatial-Temporal Graph Neural Networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD' 22)*, 4639–4647. <https://doi.org/10.1145/3534678.3539093>.