Transportation Scenario Planning with Graph Neural Networks

Ana Alice Peregrino
aapp@cin.ufpe.br
Universidade Federal de Pernambuco

Soham Pradhan
spradh8@uic.edu
University of Illinois at Chicago

Zhicheng Liu
zhichengliu@seu.edu.cn
Southeast University

Nivan Ferreira
nivan@cin.ufpe.br
Universidade Federal de Pernambuco

Fabio Miranda
fabiom@uic.edu
University of Illinois at Chicago

ABSTRACT

Providing efficient human mobility services and infrastructure is one of the major concerns of most mid-sized to large cities around the world. A proper understanding of the dynamics of commuting flows is therefore a requisite to better plan urban areas. In this context, an important task is to study hypothetical scenarios in which possible future changes are evaluated. For instance, how the increase in residential units or transportation modes in a neighborhood will change the commuting flows to or from that region? In this paper, we propose to leverage a recently proposed graph neural network model to estimate changes in commuting flows taking into account land use and infrastructure changes. We evaluate our methodology through two real-world case studies set in two large cities from Brazil.

KEYWORDS

Urban planning, transportation, urban data, scenario planning, graph neural networks

1 INTRODUCTION

Cities are complex environments that house most of the world's population; today, 55% of the world's population lives in urban areas, and this is expected to increase to 68% by 2050 [7]. For this reason, an enormous problem faced by governments and urban planners is how to plan for this new surge of people while solving the already challenging scenarios of the present. One of the issues present in nearly all mid-sized and large cities is human mobility. In particular, every day millions of people commute from home to work and limitations in the transportation infrastructure cause not only people to waste their time, but also an increase in pollution and health problems. A proper understanding of the dynamics of commuting flows in a city is therefore a requisite to better plan urban areas. Urban planners, transportation specialists, and city agencies more often than not rely on precedent and data analyzed in isolation to make decisions that can impact or transform a city. With the growing availability of urban data and advances in machine learning, there are new opportunities for data-driven solutions to better support the exploration of possible alternate urban scenarios. Such information can be used, for instance, by urban planners to guide the development of new neighborhoods, or transportation specialist to direct the deployment of new transportation modes.

In this context, the goal of this paper is to study the use of a graph neural network-based commuting flow prediction model to assist experts in the identification of the effects of infrastructure, land use and/or policy changes on commuting flows. Understanding the commuting flow can help answer many what-if questions in the planning stage, such as *"If a new high-rise building is planned for a region, to what regions would people commute to work?"*, and *"How to modify the transportation infrastructure to improve commuting efficiency?"* In order to enable the data-driven scenario planning, we take the first steps in leveraging the Geo-contextual Multitask Embedding Learner (GMEL) model, previously proposed in Liu et al. [15], as our basis model for predicting commuting flows based on geographic information (e.g., infrastructure, land use, transportation). While major cities have the resources to collect and process high-resolution land use data, other cities do not have such capabilities, especially in countries in the Global South. To test the effectiveness of our methods in cities in developing countries, we focus our efforts on two cities in Brazil. Using urban data from these cities, we train the GMEL model to predict flows based on infrastructure and land use changes. Through a set of case studies we show how scenario planning methods based on graph neural networks can reveal important information to transportation experts.

2 RELATED WORK

2.1 Commuting flow prediction

Classical trip distribution works follow a gravity model, first introduced in the 1940s [31], and assumes that the trip volume is proportional to the product of population of origin and destination and is inversely proportional to the distance between origin and destination. Modern extensions of gravity model take into account several factors, such as demographics and land use, to more accurately model attraction [8], but still fall short of properly modeling complex nonlinearities, such as interactions between urban utilities and human mobility. More recent approaches have used radiation model [23, 28], derived from stochastic process considering intervening opportunities. Radiation models are limited by data capacity, using only population distribution and ignoring the growing availability of urban data [1]. Machine learning approaches have also been proposed for trip distribution modeling, including random forest [18, 19, 22]. These machine learning models make use of rich urban data and can better model complex nonlinearities. However, these models ignore the spatial correlations and consider only the characteristics of origin and destination. In our previous work [15], we proposed to use a graph neural network to learn geo-contextual embeddings for commuting trip distribution modeling, achieving better predictive performance when compared against baseline models, such as gradient boosting regression tree, random forest, gravity model, and node2vec.

2.2 Graph representation learning

Graph representation learning aims at learning low-dimensional features (i.e., embeddings) for each node in a graph, preserving both the graph structure and node attributes. This approach allows the embeddings to be used in a myriad of analytical tasks, such as community detection [14], traffic prediction [26] and graph isomorphism [27]. Graph neural networks (GNN) provides powerful graph embedding capabilities [27]. Different approaches include graph convolution neural network (GCN), based on the notion of convolution on graphs [3], general inductive framework that leverages node attributes to generate node embeddings in a message-passing way [10], and graph attention networks that leverage self-attention mechanisms to allow messages passed by neighbors to be aggregated with different weights [24]. Successful applications include traffic prediction [30], recommender system [29] and drug discovery [11]. In this work, we use our previous work on graph attention network with a modified attention mechanism so that the model can capture the spatial correlations and enable the construction of alternate scenarios [15].

2.3 Scenario planning

The ability to plan for different scenarios and consider different outcomes is important in several domains, including urban planning and transportation. At its core, scenario planning allows planners to analyze future outcomes based on present-day decisions [5]. Traditional approaches usually use regression analysis [4], travel forecasting models, or economic models [2]. More recently, machine learning approaches have been proposed to best guide stakeholders on how to best plan for future growth, while taking into account environmental considerations [12], or how to best calculate land-use configuration given surround spatial contexts [25]. Simultaneously, given that transparency, expert feedback and community participation is an increasingly important topic in scenario planning, different proposals have considered a human-in-the-loop approach to foster the involvement of stakeholders [9, 16, 17].

3 METHODOLOGY

In this work, we leverage our previously proposed graph neural network model, called Geo-contextual Multitask Embedding Learner (GMEL) [15]. Next, we briefly describe GMEL and also how we use the model for scenario planning in two cities in Brazil.

3.1 GMEL

GMEL is a graph neural network model for commuting trip distribution modeling. Basically, the model consists of two components: a geo-contextual multitask embedding learner and a flow predictor. The learner was designed to capture the spatial correlations from geographic neighborhoods. The model utilizes graph attention network (GAT) to encode the spatial dependencies into an embedding space. To disentangle the origin and destination characteristics that are hidden in the infrastructure and land use data, GMEL employs two separate GATs to encode the geographic contextual information into two different embedding spaces. GMEL employs multitask learning framework which imposes stronger restrictions forcing the embeddings to encapsulate effective representations for flow prediction. The second component, the flow predictor, employs

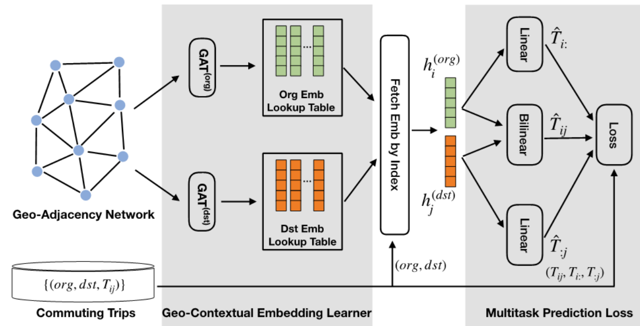

Figure 1: GMEL architecture.

a gradient boosting machine (GBM) as the regression model to predict commuting flows. GBM iteratively evaluate the largest information gain of features, automatically selecting and combining useful numerical features to fit the targets.

Figure 1 presents an overview of the GMEL architecture. The model makes use of two datasets: 1) a set of origin-destination commuting trips, where each trip is composed of an origin, a destination, and number of commuters traveling from origin to destination. Origin and destinations are aggregated at a census tract level, and commuting flows are yearly static values, i.e., number of people that report living in an origin census tract and working at a destination census tract. 2) a geo-adjacency network, an undirected weighted graph with the following properties: geographic units (i.e., census tracts) as nodes, with an associated set of urban indicators, and the weight of edges describe the strenght of correlations between units (i.e., travel distance, trip duration).

GMEL was originally proposed using an extensive land-use dataset from New York City, called Property Land Use Tax Lot Output (PLUTO). The dataset describes, for each year between 2008 and 2021, the land use information of the city, at a lot level. This enabled us to perform initial *what-if* scenario explorations. In one specific case study, we showed how we could leverage a GMEL model that was trained using the PLUTO data set for one specific year (2013) and predict flows considering the modified land use of a subsequent year (2015), highlighting how GMEL can guide urban planners and policy makers to make informed decisions when it comes to new urban development scenarios.

3.2 Scenario planning with GMEL

In this paper, we further explore the possibilities of using GMEL for *what-if* scenario planning. We focus our efforts on two Brazilian cities: Recife, the oldest capital city of Brazil, the 3rd most populous city in the Northeast region of Brazil and the 9th most populous in the country; Curitiba is the most populous city in the South region of Brazil, and the 8th most populous city in the country. Curitiba in particular is known for its innovative urban planning initiatives, targeted at improving public transportation accessibility and promote housing development.

Recife and Curitiba are among the fastest growing cities in Brazil [6], creating the need to have the right methods in place to allow stakeholders to better plan urban interventions. This need was recently highlighted by a report from the city of Recife, that

Table 1: Summary of urban indicators for Recife and Curitiba.

Category	Recife		Curitiba	
	# Feat.	Content	# Feat.	Content
Infrastructure	30	No. of different types of buildings (12), density of res. units (4), no. of buildings per built year (11), no. of bike stations (1), perimeter of bike lanes (1), perimeter of bus lanes (1)	12	Perimeter of bike lanes (1), no. of lots per zone (11)
Land use	14	Land area ratio of retail/office (12), floor area ratio (2)	11	Land area ratio of zones (11)
Speciality	2	Whether or not the urban geographic unit contains landmarks (1), no. of cultural routes (1)	1	Number of lots contained in a historic area (1)

states the *"need to develop analytical tools that allow the creation of scenarios to capture specific changes in certain areas of the city, and to allow the assessment of scenarios related to the implementation of new transportation infrastructure."* At the core of our proposal is the ability to use a graph neural network (i.e., GMEL) to predict changes in commuting flows, given changes in the land use and built environment. We follow a set of steps that allow us to train and validate a model, and then use this model to predict flows in a different scenario. Our steps can be summarized as follows:

Model training: we initially train a GMEL model for each city, considering land-use and commuting flow data from Recife and Curitiba. The model is trained using stochastic gradient descent in an end-to-end manner. With the embeddings from the trained GMEL, a GBRT is trained as flow predictor based on the concatenation of origin-destination embeddings and travel distances to predict the commuting flow. The model is then tested using a holdout set.

Scenario planning: after the model is trained and tested, we change the urban indicators of the city, following plausible urban modification scenarios currently being proposed in the two cities. We update the geo-adjacency network as to follow these modifications, and use the previously trained models to generate *new* embeddings for the modified network. The flow prediction then uses these embeddings to predict new commuting flows.

4 MODEL TRAINING AND RESULTS

This section describes the input generation process for the GMEL model and the experimental results for both Recife and Curitiba.

4.1 Data description

To train and validate the models used for scenario planning, we used open data sets from Recife and Curitiba. In both cases, we used the citys' 2020 census tracts as the *geographic units*. To measure the travel distance between (the centroids of) the census tracts, the Open Source Routing Machine (OSRM) was used, as described in [15]. Next, we detail the data sets and their sources. The information is also summarized in Table 1.

The first city studied in this paper is the city of Recife, capital of the state of Pernambuco in the Northeast cost of Brazil. For this analysis we used the commuting flow data set obtained from a survey carried out by the city's administration in the year of 2018. This survey captured data on typical commuting movements performed by the population that resides, works, studies or seeks services in the metropolitan region of the city. As urban indicators, we used information about individual lots, as well as indicators that show the presence of special preservation buildings, cultural routes, bicycle stations, bike lanes and exclusive bus lanes, all available on Recife's open data portal [21]. After joining flow and the indicators data sets, we ended up with 1,347 census tracts that cover the vast majority of the city. The commuting flows were aggregated into geographic unit level flows, resulting in 23,336 commuters and 15,945 pairs of origin-destination trips divided in train (60%), test (20%) and validation (20%) data sets.

The second city considered is Curitiba, capital of the state of Paraná in the South region of Brazil. The commuting flow data set used for the analysis was obtained from a survey carried out by the city's administration in 2017, mapping commuting patterns in the metropolitan region of the city. As urban indicators, we used individual lots aggregated by zone, according to the new zoning legislation, established in 2020, as well as the distribution of bike lanes in the city, data made available by Curitiba's urban planning and research institute [20]. After joining commuting flows and indicators, we ended up with 2,200 census tracts, covering the vast majority of the city. The commuting data resulted in 45,365 commuters and 32,988 pairs of origin-destination trips, divided in the same way as previously described.

In our case studies, we modify the previously mentioned urban indicators as to effectively predict new flows when considering alternate scenario plans.

4.2 Performance analysis

We evaluated the performance of the GMEL model in both Recife and Curitiba. To measure the prediction performance, we adopted three evaluation metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Common Part of Commuters (CPC). For each pair i, j of nodes, and predicted flow value $\hat{T}_{i,j}$ and groundtruth flow value $T_{i,j}$, we have RMSE $= \sqrt{\frac{1}{|T|}\sum_{i,j}(\hat{T}_{ij} - T_{ij})^2}$, MAE $= \frac{1}{|T|}\sum_{i,j}|\hat{T}_{ij} - T_{ij}|$, and CPC $= \frac{2\sum_{ij} min(\hat{T}_{ij}, T_{ij})}{\sum_{ij}\hat{T}_{ij} + \sum_{ij} T_{ij}}$.

RMSE and MAE are widely used as evaluation metrics for regression problems. CPC is widely used in trip distribution modeling [13, 22], and it measures the agreement between predicted value and target value; CPC is 0 when no agreement is found, and it is 1 when the two are identical.

The model's performance for the two cities was quite similar (Table 2), with the RMSE being the metric with the most significant difference, although still small. This can be attributed to the difference in the number of commuting flows and census tracts used in Curitiba and Recife. Given that this number was greater in Curitiba, a lower RMSE was expected.

Table 2: Performance on test set.

City	RMSE	MAE	CPC*
Recife	1.43	0.73	0.73
Curitiba	1.08	0.70	0.73

* Higher is better

5 CASE STUDIES

Following our methodology, we now present two initial alternate scenario plans for the city of Recife. Both of the case studies highlight the importance of developing tools and frameworks that allow the creation of scenarios to assess the impact of land use changes.

5.1 New bike lanes in Recife

In recent years, there has been a growing movement in which citizens organize themselves to think about alternatives that go beyond motor vehicles. Bicycles are means of transport for small and medium distances that bring benefits to health and the environment. In some cities around the world, bicycles are already a widely used mode of transportation for a significant portion of the population, while in Recife it is mostly used by a small portion of the population. While there has been an increasing demand for new cycle paths around the city, an important question that should be answered before actually creating this new paths is whether this infrastructure will effectively generate an increase in the flow of cyclists in the region. To show how to apply our methodology in this scenario, we chose four major avenues that connect important points in Recife. These avenues have a large daily flow of vehicles, however currently they do not have any type of cycling infrastructure, posing a risk to cyclists who need to travel around them to reach their destination. We distributed 24 km of cycle paths between these four avenues, mapped to the zones where the avenues are located (Figure 2(a)). To analyze the changes caused by this modifications we looked at the flows among units whose geographical centers are up to 2 km away from the centers of units that received the new cycle paths. For each flow we computed the relative change as a way to assess the amount of flow change in these areas. The results show an average increase of 13% (and 0.59 of standard deviation) of flows among those units (Figure 2(b,c)). While our model did not use any data on the modality of transportation, we hypothesize that these changes could represent an increase of biking in these areas.

5.2 New high-rise buildings in Recife

Building large projects in a city is an activity that is always accompanied by a series of impacts, whether on the landscape, on the region's economy and/or on the flow of people moving from one point to another. It is therefore of great importance for urban planners to first estimate these impacts on the city's infrastructure in order to anticipate actions to fulfill new needs or solve possible upcoming problems. As a way to estimate the impact on the flow of people due to a large construction in downtown Recife, we used the real-world case of the construction of 13 towers, with sizes ranging from 13 to 44 floors, on the former historical site of the José Estelita Pier (Figure 3(a)). We incorporated all the buildings into the model, adding information of features related to the number of different types of buildings, density, buildings in each build year interval,

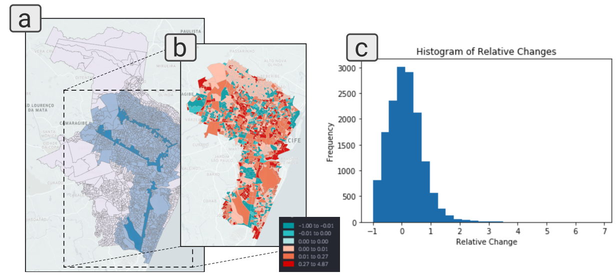

Figure 2: Alternate land-use scenario considering new bike lanes: (a) census tracts new bike lanes (dark blue), and tracts with centroids within 2 km of distance (light blue); (b) commuting flow differences, considering the new scenario; (c) histogram with commuting flow differences.

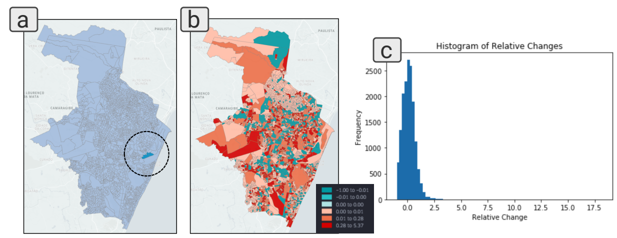

Figure 3: Alternate land-use scenario considering new buildings: (a) census tracts with new buildings (dark blue); (b) commuting flow differences, considering the new scenario; (c) histogram with commuting flow differences.

land/area ratio and floor/area ratio. We performed a similar analysis as in Section 5.1 in order to analyze relative changes in commuting flows due to this urban indicator modifications. The results show an average increase of 12,5% (0.605 std. deviation) of flows among those units (Figure 3(b,c)). Notice that the model predicts relative changes thought the city. However, the most significant changes happen in units close to the new project, or in high-density units that already have large commuting flow volumes. These flows are predicted to increase by at least 25%.

6 CONCLUSIONS

Our goal in this paper was to analyze the behavior of flows on proposed *what-if* scenarios using the GMEL model, evaluating the model's performance with available urban data from two different cities in Brazil. Our initial results show that the model could be used as a tool to assist urban planners and transportation researchers in the decision-making process, with good performance even when considering cities in developing countries. There are, however, challenges regarding data quality that need to be addressed. As the majority of data-driven approaches, GMEL is heavily dependent on the quality of the data sets used, therefore it is necessary an effective involvement of public administration in providing accurate information to be used in the model. In future work, we plan to further evaluate our model, and extend it to take into account multimodal commuting flows, which will allow us to answer more specific questions regarding people's movement in the city considering the type of transportation used.

REFERENCES

[1] Luciano Barbosa, Kien Pham, Claudio Silva, Marcos R. Vieira, and Juliana Freire. 2014. Structured Open Urban Data: Understanding the Landscape. *Big Data* 2, 3 (2014), 144–154.

[2] Keith Bartholomew. 2007. Land use-transportation scenario planning: promise and reality. *Transportation* 34, 4 (2007), 397–412.

[3] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. 2014. Spectral networks and locally connected networks on graphs. *Proceedings of International Conference on Learning Representations* (2014).

[4] Arnab Chakraborty, Nikhil Kaza, Gerrit-Jan Knaap, and Brian Deal. 2011. Robust Plans and Contingent Plans. *Journal of the American Planning Association* 77, 3 (2011), 251–266.

[5] Arnab Chakraborty and Andrew McMillan. 2015. Scenario Planning for Urban Planners: Toward a Practitioner's Guide. *Journal of the American Planning Association* 81, 1 (2015), 18–29.

[6] CityMayors Statistics. 2015. The world's fastest growing cities and urban areas from 2006 to 2020. http://www.citymayors.com/statistics/urban_growth1.html.

[7] Department of Economic and Social Affairs, United Nations. 2018. 68% of the world population projected to live in urban areas by 2050, says UN. https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html.

[8] Sven Erlander and Neil F Stewart. 1990. *The gravity model in transportation analysis: theory and extensions*. Vol. 3. Vsp.

[9] Nivan Ferreira, Marcos Lage, Harish Doraiswamy, Huy Vo, Luc Wilson, Heidi Werner, Muchan Park, and Cláudio Silva. 2015. Urbane: A 3D framework to support data driven decision making in urban development. In *2015 IEEE Conference on Visual Analytics Science and Technology, VAST 2015 - Proceedings*. Institute of Electrical and Electronics Engineers Inc., 97–104.

[10] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., 1024–1034.

[11] Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. 2016. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design* 30, 8 (2016), 595–608.

[12] Youjung Kim and Galen Newman. 2020. Advancing scenario planning through integrating urban growth prediction with future flood risk models. *Computers, Environment and Urban Systems* 82 (2020), 101498.

[13] Maxime Lenormand, Aleix Bassolas, and José J. Ramasco. 2016. Systematic comparison of trip distribution laws and models. *Journal of Transport Geography* 51 (Feb. 2016), 158–169.

[14] Ye Li, Chaofeng Sha, Xin Huang, and Yanchun Zhang. 2018. Community detection in attributed graphs: An embedding approach. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[15] Zhicheng Liu, Fabio Miranda, Weiting Xiong, Junyan Yang, Qiao Wang, and Claudio T. Silva. 2020. Learning Geo-Contextual Embeddings for Commuting Flow Prediction. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*.

[16] F. Miranda, H. Doraiswamy, M. Lage, L. Wilson, M. Hsieh, and C. T. Silva. 2019. Shadow Accrual Maps: Efficient Accumulation of City-Scale Shadows Over Time. *IEEE Transactions on Visualization and Computer Graphics* 25, 3 (March 2019), 1559–1574.

[17] Thomas Ortner, Johannes Sorger, Harald Steinlechner, Gerd Hesina, Harald Piringer, and Eduard Gröller. 2016. Vis-a-ware: Integrating spatial and non-spatial visualization for visibility-aware urban planning. *IEEE Transactions on Visualization and Computer Graphics* 23, 2 (2016), 1139–1151.

[18] Nastaran Pourebrahim, Selima Sultana, Amirreza Niakanlahiji, and Jean-Claude Thill. 2019. Trip distribution modeling with Twitter data. *Computers, Environment and Urban Systems* 77 (Sept. 2019), 101354.

[19] Nastaran Pourebrahim, Selima Sultana, Jean-Claude Thill, and Somya Mohanty. 2018. Enhancing Trip Distribution Prediction with Twitter Data: Comparison of Neural Network and Gravity Models. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. ACM, 5–8.

[20] Prefeitura de Curitiba. 2021. Instituto de Pesquisa e Planejamento Urbano de Curitiba. https://ippuc.org.br.

[21] Prefeitura do Recife. 2021. Portal de Dados Abertos do Recife. http://dados.recife.pe.gov.br/.

[22] Caleb Robinson and Bistra Dilkina. 2018. A Machine Learning Approach to Modeling Human Migration. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies* (Menlo Park and San Jose, CA, USA) *(COMPASS '18)*. ACM, New York, NY, USA, Article 30, 8 pages.

[23] Filippo Simini, Marta C. González, Amos Maritan, and Albert-László Barabási. 2012. A universal model for mobility and migration patterns. *Nature* 484, 7392 (April 2012), 96–100.

[24] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. *International Conference on Learning Representations* (2018).

[25] Dongjie Wang, Yanjie Fu, Pengyang Wang, Bo Huang, and Chang-Tien Lu. 2020. Reimagining City Configuration: Automated Urban Planning via Adversarial Learning. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*. 497–506.

[26] Hongjian Wang and Zhenhui Li. 2017. Region Representation Learning via Mobility Flow. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (Singapore, Singapore) *(CIKM '17)*. ACM, New York, NY, USA, 237–246.

[27] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How powerful are graph neural networks?. In *Proceedings of International Conference on Learning Representations*.

[28] Xiao-Yong Yan, Chen Zhao, Ying Fan, Zengru Di, and Wen-Xu Wang. 2014. Universal predictability of mobility patterns in cities. *Journal of The Royal Society Interface* 11, 100 (Nov. 2014), 20140834.

[29] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 974–983.

[30] Byeonghyeop Yu, Yongjin Lee, and Keemin Sohn. 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 (2020), 189–204.

[31] George Kingsley Zipf. 1946. The P1 P2/D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review* 11, 6 (1946), 677–686.