

Prediction of Urban Population-Facilities Interactions with Graph Neural Network

Margarita Mishina[1,2(✉)], Stanislav Sobolevsky[1,3,4], Elizaveta Kovtun[6], Alexander Khrulkov[1,2], Alexander Belyi[1], Semen Budennyy[5,6], and Sergey Mityagin[2]

[1] Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Kotlarska 2, Brno 61137, Czech Republic
marg.mished@gmail.com
[2] ITMO University, Birzhevaya line, 14, Saint-Petersburg, Russia
[3] Center for Urban Science and Progress, New York University, 370 Jay Street, Brooklyn 11201, NY, USA
[4] Institute of Law and Technology, Faculty of Law, Masaryk University, Veveri 70, Brno 61180, Czech Republic
[5] Artificial Intelligence Research Institute (AIRI), Moscow, Russia
[6] Sber AI Lab, Moscow, Russia

Abstract. The urban population interacts with service facilities on a daily basis. The information on population-facilities interactions is considered when analyzing the current city organization and revealing gaps in infrastructure at the neighborhood level. However, often this information is limited to several observation areas. The paper presents a new graph-based deep learning approach to reconstruct population-facilities interactions. In the proposed approach, graph attention neural networks learn latent nodes' representation and discover interpretable dependencies in a graph of interactions based on observed data of one part of the city. A novel normalization technique is used to balance doubly-constrained flows between two locations. The experiments show that the proposed approach outperforms classic models in a bipartite graph of population-facilities interactions.

Keywords: urban mobility · graph neural network · flows prediction

1 Introduction

Today, the spatial development of urban infrastructure aims to ensure the normal functioning of a city and the ability to meet the urgent needs of citizens and tourists. Urban infrastructure covers engineered systems, municipal utilities, and various service facilities (kindergartens, schools, polyclinics, and others) [16]. The quality of urban infrastructure, in particular the service facilities system, is determined not only by the number of its elements located in a city but also by the spatial correspondence of their location to the places of citizens'

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
O. Gervasi et al. (Eds.): ICCSA 2023, LNCS 13956, pp. 334–348, 2023.
https://doi.org/10.1007/978-3-031-36805-9_23

concentration (in the areas of residence, leisure, work, and others). The latter reflects the peculiarity of population-facilities interactions prevailing in a city - where and how long citizens must travel to get a specific service.

Given the growing impact of service facilities on citizens' daily life, the issue of spatial population-facilities interactions is of keen interest among government, developers, and researchers [30]. Recent studies consider this issue to examine spatial inequality, which implies the inability of the population living in some parts of a city to access required facilities within a moderate distance [5,9,13]. The solution to this problem lies in facility distribution optimization based on knowledge about population density and population movements between locations of primary activities and preferable facilities in a city [13,15]. One may notice that information about population density in cities is available in the census data of many countries. However, information about intra-city movements is limited with several observation areas and needs modeling in most cases.

Over the last decades, the scientific community has been working on examining the aspects of human mobility and understanding the patterns of preferences the urban population would have for some objects over others [8,17,27]. A breakthrough has been made in this field of study due to publishing in open access a large amount of location-embedded information about people's movement in an urban environment (from census data, mobile phone records, GPS trackers, social networks, and other data sources). Since then, various laws and mathematical models have been derived to formalize people's movement in a city by abstracting it into a straightforward set of regularities applicable to a territory of interest where observed data is unavailable [6].

Among the models that have been proposed, the gravity model [33] and the radiation model [29] have been most commonly used to approximate aggregated population flows between origins and destinations. The models represent two leading schools of thought that explain human mobility patterns differently. The gravity model and its derivatives assume that flows between origin and destination decline as the distance between them increases. Meanwhile, the radiation model and similar ones postulate that the probability of commuting between origin and destination is inversely proportional to the number of opportunities in between (i.e., the number of alternative destinations). Both models are distinguished due to their simplicity and clear interpretability, though they commonly tend to simplify the complex structure of human mobility flows [6]. This limitation can be partly solved via the constrained versions of the models and the calibration of its free parameters based on observed data.

Recently machine learning approaches have been proposed to address the shortcomings of traditional models and improve the ability to reconstruct mobility flows. In [26], the authors train an XGBoost model and a densely connected feed-forward neural network (FNN) to predict migration flows considering a set of externalities. Similarly, the Deep Gravity model introduced in [28] adopts FNN to reconstruct the number of trips between regions of interest. The studies

find that the neural network approach outperforms the state-of-the-art models and offers promising results. The recombination of locations' features in a non-linear way allows to capture more complex dependencies and provides higher generalization capability. This observation seems to indicate that there is room for further development and improvements.

Considering information on urban population-facilities interactions, the organization of available knowledge in the form of graphs emerges quite naturally. Nodes of a graph could be associated with urban objects, while graph edges are inherently connected with linking routes. When there is a necessity to learn intrinsic patterns of such data structure, Graph Neural Networks (GNNs) take the stage. The benefits that come from leveraging GNNs are tracked in a variety of fields: crystallography [12], fake news detection [7], pose estimation [11]. Similar to the case of the revolutionization of the natural language processing domain with an attention-based mechanism, the incorporation of attention idea into GNNs has led to significant performance enhancements in many tasks. So, there are Graph Attention Networks (GATs) [31] that include self-attentional layers, which implicitly learn the importance of neighboring nodes. This peculiarity is frequently crucial in a range of applied areas, such as recommendation systems [32] or travel time estimation [14]. Noteworthy, graph prediction problems can be broadly divided into three groups: graph-level, node-level, and edge-level tasks. In this work, we focus on the reconstruction of population flows. Thus, the edge-level view, namely edge regression problem formulation, comes to our attention.

To confirm the assumption that GNNs can improve the accuracy of reconstruction population flows between two locations in a city, we designed a new graph-based deep learning approach to modeling population-facilities interactions. The study addresses the inductive problem of predicting the number of people interacting with facilities considering observed demand associated with places of residence, facilities capacity, and travel distance.

The peculiarity of the task is that the population flows between two locations in a city are often doubly constrained, which means that the sum of prediction for all out-going and in-going edges must not exceed the fixed values associated with nodes (namely, the number of people demand facility and facility capacity). The traditional mobility models comply with the constraints using an iterative balancing procedure, which can not be adopted for the neural network models due to the back-propagation mechanism. Previously, this side of the issue has yet to be considered [25,26,28]. To fill the gap, we introduce a one-step normalization technique that substitutes the iterative balancing procedure while using neural networks to predict doubly-constrained population flows.

The study shows that the proposed approach outperforms the doubly constrained gravity model and the model that uses only FNN. The experiments demonstrate that spatial dependencies in a network of interactions learned by GNNs contribute to the accuracy increase in the edge-level prediction task.

2 Materials and Methods

This section is organized the following way: first describes the data used for the experiments (as well as the process of its collection and preparation), then the following subsection clarifies which version of the Gravity Model we use as a baseline, and finally, the last subsection presents the proposed approach.

2.1 Data

The proposed approach operates on a direct bipartite graph in which each edge represents the interaction between two kinds of nodes that stand for places of population activities and facilities. While modeling population-facilities interactions, using smaller location units as urban objects is advantageous instead of grid cells. This makes it possible to reconstruct the interactions on the small distance that constitutes a significant part of the intra-city population movements, although it requires more accurate data. Nodes' features of a graph include the information about origins (places of activities) and destinations (facilities) that may impact the interaction between them, in particular, the population demands concentrated in places of activities (i.e., the number of people who demand a facility) and capacities of facilities. An edge between two nodes represents a population-facility interaction – the number of people concentrated in location i that get the facility's service j per specified unit of time. Each edge is associated with the travel distance between the origin and destination that it connects.

In the study, we perform experiments with the data of the population's daily interactions with kindergartens (as one of the essential social services for a substantial part of the population) in Saint Petersburg, Russia (Fig. 1). According to city-planning regulations, kindergartens are usually located in residential neighborhoods; therefore, we considered residential buildings to be places of primary activity. The graph of interactions consisted of 21 463 origin nodes, 1 278 destination nodes, and 27 429 714 edges ($21\,463 \cdot 1\,278$) from which only 21 742 denoted actual interactions. The nodes' features represented the demand in each residential building and the capacity of each facility. The edge attributes included the number of people interacting with the facility and the travel distance in meters.

The residential buildings and kindergartens' locations were collected from different mapping services (OpenStreetMap [3], GoogleMaps [2], and 2GIS [1]) and merged into one dataset. The same mapping services helped determine each kindergarten's capacity via its standardized building type[1], which could be clearly distinguished on a map or satellite imagery.

The number of people with a demand for kindergartens was specified based on Russian census data [4]. The target group was considered to be people with children aged from 3 to 6 years. Since Russian census data provide the relevant information only for municipal units, the total number of people in the target

[1] In Russian regions, social facilities, such as kindergartens, are usually built on standard projects determining maximum capacity.

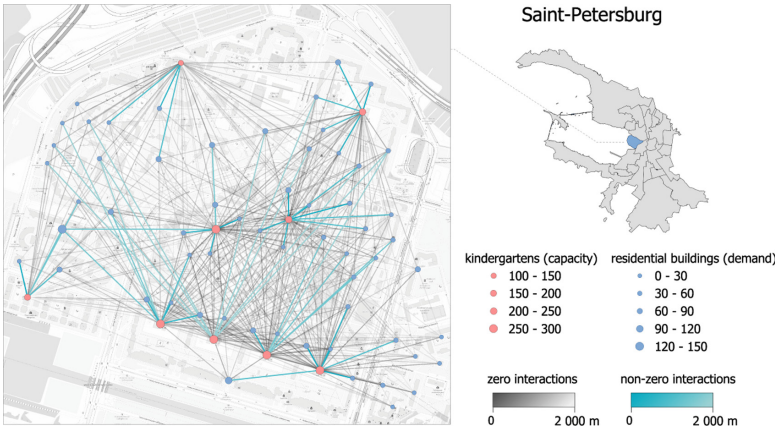

Fig. 1. A part of the bipartite graph representing the population's daily interactions with kindergartens (Saint-Petersburg).

group was distributed between residential buildings in the corresponding municipal unit, as described in the study [20].

Travel distances between residential buildings and kindergartens were measured as the shortest paths on the city road network consisting of driveways and walkways. The city road network was obtained from OpenStreetMap and pre-processed to eliminate topological errors [23]. Identifying the shortest paths between a large number of objects on a road network is often computationally expensive and, therefore, might be simplified to calculating Euclidian distances. However, one should keep in mind that only road network distances consider the urban environment's barriers that impede citizens' movement and thus provide more precise measurements, which can be crucial in some cases.

Unlike other urban data described above, data on actual population-facilities interactions is not publicly available. However, this is the information that stakeholders (represented by the government and some types of corporations) usually own. To conduct the experiments without real data, we modeled the population-facilities interactions based on the two statements. Given that kindergartens is a critical daily-use social infrastructure for a significant part of the urban population, we assumed the following:

1. All citizens prefer applying to the nearest kindergarten to meet the demand.
2. The system of interactions is in balance so that the maximum number of people is distributed between all facilities within available capacities.

We formed the proposed assumptions as an optimization function to distribute the citizens between available facilities. The system of linear inequalities implied the double constraints on the sum of out-going and in-going flows. The constraints reflect that the total number of interactions originating from the residential building i must not exceed the demand O_i (assuming that a citizen can

apply to only one facility), and the total number of interactions with the facility j must not be greater than its capacity D_j (not to cause facility's overloading).

$$\begin{aligned} &\text{maximize} \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{y_{ij}}{d_{ij}} \\ &\sum_{j=1}^{n} y_{ij} \leq O_i, i = 1..m \\ &\sum_{i=1}^{m} y_{ij} \leq D_j, j = 1..n \end{aligned} \tag{1}$$

where m and n denote the numbers of residential buildings and facilities respectively; O_i is the demand accumulated in the building i (origin); D_j is a capacity of facility j (destination); y_{ij} indicates the number of people living in the building i that get a service of the facility j traveling the distance d_{ij}.

Solving this system via integer linear programming, we obtained information on population-facilities interactions throughout the city. One may notice that the data obtained this way can differ from the real one. However, they provide a solid opportunity to investigate the performance of the proposed approach, which might encourage further research with more accurate data.

2.2 Gravity Model

The Gravity Model broadly applies to model spatial interactions in many domains. In most studies on human mobility, the Gravity Model is considered a fundamental baseline. The basic idea of the model is that the number of interactions between two locations is directly proportional to their sizes (in terms of population and other quantitative characteristics) and inversely proportional to the distance between them. In addition, the singly- and doubly-constrained model versions fix the total number of people originating from a location i or/and arriving at a location j. The choice of a model version depends on the amount of information available. In the current study, the demand accumulated in a residential building and the capacity of a facility naturally constrain the sum of out-going and in-going flows in the system of spatial interactions (Eq. 1). Therefore, we employ the Doubly-Constrained Gravity Model (DCGM) as a primary baseline.

$$\begin{aligned} \widehat{y}_{i,j} &= A_i B_j O_i D_j f(d_{ij}) \\ A_i = \frac{1}{\sum_j B_j D_j f(d_{ij})}; B_j &= \frac{1}{\sum_i A_i O_i f(d_{ij})} \end{aligned} \tag{2}$$

where $\widehat{y}_{i,j}$ denotes the predicted number of interactions between building i and facility j; $f(d_{ij})$ indicates a distance decay function which commonly modeled with a power-law d_{ij}^{β} or an exponential form $\exp(\beta d_{ij})$, β stands for a free parameter; A_i is a vector of the origin balancing factors; B_j is a vector of the destination balancing factors.

Most often, DCGM implements three steps: model calibration, estimation of interactions probabilities, and balancing the origin-destination (OD) matrix. The calibration aims to tune a free model parameter β based on observed data, more commonly via Poisson regression and Generalized Linear Modeling frameworks [18,24]. In the second step, the probability of interactions between each pair of locations is computed by Eq. 2 with the adjusted parameter β and unit vectors A_i, B_j. Finally, the obtained OD matrix is balanced by calibrating vectors A_i, B_j with the Iterative Proportional Fitting procedure [10].

The current study performed the experiments with the R package TDLM[1], which implements all three steps of DCGM [21]. Applying the gravity model to the dataset described in the previous section, we spotted a poor generalization ability of the model in this particular case which can be justified by a large size of interactions network [19]. To overcome this obstacle, we additionally tested the local version of the gravity model introduced in [22], which implies applying the model not to the whole graph but to the interactions within a truncation radius r. We slightly modified the model suggesting its iterative application to the graph with the increase of the truncation radius by Δr and recalculation of the residual fixed sum of in-going (out-going) flows on each step (Fig. 2).

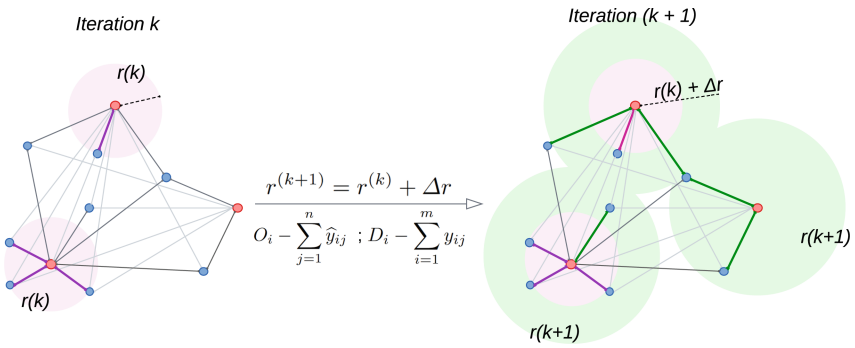

Fig. 2. The application of the Local Doubly-Constrained Gravity Model (LDCGM) to the interactions network. On each iteration k, LDCGM is applied to the edges with the distances $r^k \leq d_{ij} < r^k + \Delta r$; then for each node in a graph, the demand O_i and the capacity D_j are recalculated by subtracting the sum of predicted in-going (out-going) flows so that on the next iteration, only residual values (that have not been distributed within the truncation radius) are included in calculations. The first iteration starts from the state $r^{(0)} = 0$ and $r^{(1)} = r^{(0)} + \Delta r$, and the last iteration finishes when the longest edge in a graph is reached, or the demands (capacities) are fully distributed.

[1] https://epivec.github.io/TDLM/.

2.3 The Proposed Approach

The neural network approach performs non-linear recombination of nodes' and edges' features in a graph, which detects intrinsic patterns and improves model performance. Still, the advantage of neural network models is often accompanied by high computational complexity. The large size of a densely connected network further complicates model training and prediction.

The densely connected network of spatial interactions might include millions of edges. However, the edges are heavily zero-inflated – most do not represent any actual interaction (flow) between pair of locations. This peculiarity makes the dataset large and imbalanced; both can adversely affect the training process. To address this problem, we employed the undersampling technique to prepare the training dataset. The technique suggests decreasing the number of negative (zero) edges to $n_p f$, where n_p denotes the number of positive (non-zero) edges and f indicates a hyperparameter adjusting the ratio between positive and negative edges. Compared to the study [26], which describes a similar undersampling technique, we additionally clarify that the negative edges should be sampled with the same distance distribution as the positive edges. This clarification might be crucial if the positive edges have long-tail distance distribution.

In contrast to the training process, testing must be performed on all edges of a densely connected network to avoid adjusting model predictions on knowledge about actual interactions. Given the size of the interactions network, the prediction on a full graph would have required a substantial amount of computational resources. To tackle this issue, we employed a local approach to the model prediction described in the previous section (Fig. 2), considering the step of the truncation radius Δr to be a hyperparameter.

The following subsection describes a four-step model architecture to predict population-facilities interactions in a city (Fig. 3). The model implements a novel doubly-constrained normalization technique that substitutes the iterative balancing procedure that can not be used in the back-propagation mechanism of neural networks.

Model Architecture. First, two-layer multi-head GAT operates on a directed bipartite graph to learn the latent nodes' representations considering the relative importance of the neighbors' features. Each GAT layer performs two convolutions with different message-passing directions (from origin to destination and destination to origin) to learn representations of nodes of each type in the directed graph. Multi-head attention mechanism executes five independent transformations. As recommended in the paper [31], the first layer of GAT concatenates new nodes' embeddings, while the second layer averages them. On each layer, GAT employs 128 neurons, 5 attention heads, the LeakyRelu activation function, a dropout technique (with a probability equal to 0.3), and layer normalization.

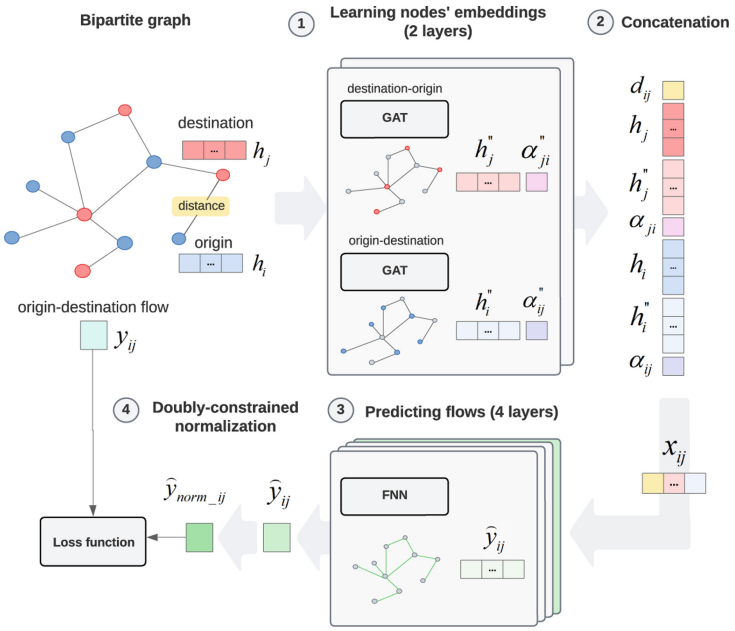

Fig. 3. Model architecture of the proposed approach. A directed bipartite graph is treated as input to the multi-head GAT, which learns attention coefficients and nodes' embeddings. For every edge, an input vector is obtained by concatenating the travel distance between nodes d_{ij}, their input features h_i, h_j, learned embeddings h_i'', h_j'' and the averages of attention coefficients $\overline{\alpha}_{ij}''$, $\overline{\alpha}_{ji}''$. Edges' input vectors are fed to the FNN to predict the number of interactions which are subsequently normalized with respect to the sum of nodes' in-going and out-going flows

The nodes' input features for learning nodes' embeddings include the information about origins and destinations that impact the interaction between them, specifically, the node's type (origin or destination) and either the population demands (for origin nodes) or capacity (for destination nodes).

$$h_i' = \Big\Vert_{k=1}^{K=5} LeakyReLU\left(\sum_{j\in\mathcal{N}(i)} {\alpha'}_{ij}^{k} {W'}^{k} h_j\right)$$
$$h_i'' = LeakyReLU\left(\frac{1}{K}\sum_{k=1}^{K=5}\sum_{j\in\mathcal{N}(i)} {\alpha''}_{ij}^{k} {W''}^{k} h_j'\right) \tag{3}$$

where h_j is an input feature vector of node j, ${\alpha'}_{ij}^{k}$ and ${\alpha''}_{ij}^{k}$ are attention coefficients computed by the k-th attention head for nodes i and j; ${W'}_k$ and ${W''}_k$ are the linear transformation's weight matrices; $\mathcal{N}(i)$ is the neighborhood of node i; h_i' and h_i'' are embeddings on the first and the second layers respectively.

Second, for every edge in a graph, the input vector x_{ij} is formed by concatenating the travel distance between nodes that it connects d_{ij}, their input feature vectors h_i, h_j, learned embeddings h''_i, h''_j and the averages of attention coefficients $\overline{\alpha}''_{ij}$, $\overline{\alpha}''_{ji}$.

$$x_{ij} = \left(d_{ij}, h_i, h_j, h''_i, h''_j, \overline{\alpha}''_{ij}, \overline{\alpha}''_{ji}\right) \tag{4}$$

As a third step, the obtained vectors are fed to FNN. The architecture of FNN includes three hidden linear layers of dimensions 128 with the LeakyRelu function, dropout technique (with a probability equal to 0.3), and layer normalization. The last one-dimensional layer makes final predictions $\widehat{y}_{ij}$. In the last layer, ReLU function limits the output $\widehat{y}_{ij}$ with the interval $[0, \infty)$ as the number of interactions assumes non-negative values.

Finally, the predicted values $\widehat{y}_{ij}$ are normalized to comply with the double constraints on the sum of out-going and in-going flows:

$$\widehat{y}_{norm_{ij}} = \min\left(\frac{\widehat{y}_{ij} O_i}{\sum\limits_{j \in \mathcal{N}(i)} \widehat{y}_{ij}}, \frac{\widehat{y}_{ij} D_j}{\sum\limits_{i \in \mathcal{N}(j)} \widehat{y}_{ij}}\right) \tag{5}$$

The loss function for the model is Weighted Mean Square Error (WMSE). The weight $\log(d_{ij})$ assigns higher importance to prediction errors for edges of greater distance, which are a minority in the training dataset with long-tail distance distribution.

$$Loss = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n} \left((y_{ij} - \widehat{y}_{norm_{ij}})^2 \log(d_{ij})\right) \tag{6}$$

where N is the number of predictions (equal to the number of edges in the training dataset represented actual and undersampled zero interactions); y_{ij} indicates the observed number of interactions, while $\widehat{y}_{norm_{ij}}$ denotes the normalized predicted number of interactions for edge ij; d_{ij} is the travel distance between nodes i and j.

3 Results and Discussion

The described models were evaluated with a cross-validation technique on subgraphs obtained by splitting the dataset into parts related to the central, southern, northern, and eastern parts of the city. The cross-validation was repeated four times, with each subgraph used exactly once as the test data. In each subgraph, nodes' features comprised a node type (residential building or facility) and either capacity or demand associated with a node. The edge attributes

contained the distance and the number of interactions (as a target value) between the nodes. Table 1 represents the number of elements (nodes and edges) included in each subgraph.

Table 1. The size of subgraph for each part of the city.

Part of the city	Origins	Destination	Edges
Central	5456	381	2 078 736 (5 731)*
Southern	6490	250	1 622 500 (6 599)
Northern	6092	413	2 515 996 (6 365)
Eastern	2901	231	670 131 (3 047)

*The value in the brackets is the number of (actual) non-zero interactions.

DCGM was calibrated and tested with power law (pow) and exponential (exp) distance decay functions by the R package TDLM. In a series of experiments, we observed a poor generalization ability of traditional DCGM on an extensive network of interactions. Specifically, DCGM tended to underestimate the number of interactions heavily. To minimize this downside of the traditional model, we additionally tested the local version of DCGM (LDCGM) with fitted hyperparameters β and Δr and found a better performance. However, both models appeared to share the identical problem - in the large system of interactions, it is practically impossible to reach a convergence during the iterative balancing of OD matrices which means that either the sum of out-going flows or in-going flows will exceed the fixed values.

The proposed graph-based deep learning approach significantly improved performance on each out-of-sample data compared to both DCGM and LDCGM models. Moreover, the suggested normalization technique (used to substitute the iterative balancing procedure) managed to consider the double constraints on the sum of in-going and out-going flows in OD matrix. To reduce the computational complexity of the neural network model, we employed the undersampling technique for model training and the local approach for predicting. Two hyperparameters f and Δr were fitted based on training data.

Further, we noticed that minor feature engineering positively impacts prediction accuracy. We extended nodes' feature vectors with information on the average distance of out-going (in-going) edges and the total demand (capacity) of connected nodes. The additional engineered features included in convolutions brought a noticeable increase in accuracy metrics.

Finally, to confirm our assumptions that GNNs embeddings could improve the accuracy of reconstruction population flows, we evaluated the interactions predicted only by FNN. The architecture of FNN was the same as described in the proposed approach, except that the input layer consisted of edge vectors obtained by concatenation only nodes' input features and travel distance between them. The experiments showed that four-layer FNN outperformed DCGM and

LDCGM models, although the accuracy of predictions was considerably lower than using the graph-based deep learning approach that adopts GAT.

To evaluate the models' performance, we computed metrics most commonly used in the regression tasks – MAE, MSE, and R-squared. Table 2 represents out-of-sample model performance for the eastern part of the city. The results for the rest parts of the city are placed in the Supplementary Materials.

Table 2. Performance on out-of-sample data for eastern part.

Models	MAE	MSE	R-squared
DCGM (pow)	14.456	468.490	0.114
DCGM (exp)	14.472	469.144	0.112
LDCGM (pow)	6.215	209.031	0.417
LDCGM (exp)	5.828	193.848	0.469
FNN	6.677	148.136	0.653
GAT + FNN (ours)	5.003	121.045	0.698
GAT ext. features + FNN (ours)	**4.719**	**113.341**	**0.719**

pow/exp - power law/exponential distance decay function.
ext. features - extended nodes' features included in convolutions.

4 Conclusion

The paper presents a new graph-based deep learning approach to modeling the population-facilities interactions. The proposed approach shows a significantly better performance in predicting doubly-constrained mobility flows than the classic Gravity Model used as a baseline. The one-step normalization technique effectively substitutes the iterative balancing procedure of OD matrix and complies with double constraints on in-going and out-going flows that often take place in the task formulation considered. The results obtained with partly modeled data provide reasonable grounds to exert efforts in collecting factual information on population-facilities interactions prevailing in a city and further research.

The experiments demonstrate that GAT significantly contributes to increased accuracy in edge regression tasks due to the attention mechanism that learns the relative importance of the neighbor's features. Moreover, based on the model's performance on out-of-sample data, GAT indicates transferability across graphs related to different parts of the city. This observation makes one think about the ability of GNNs to be transferred across graphs associated with the different cities to reconstruct population-facilities interactions. Moving forward, the transferability property of GNNs could be broadly applied to solve the problem of the incomplete availability of observed data in a city by replacing them with accurate predictions.

5 Supplementary Materials

The results of the models' evaluation for central, southern, northern, and eastern parts of the city that were obtained independently by k-fold cross-validation technique (the number of folds k is equal to 4).

Table 3. Performance on out-of-sample data.

Models	MAE	MSE	R-squared
Central part			
DCGM (pow)	14.333	489.973	0.109
DCGM (exp)	14.098	478.502	0.127
LDCGM (pow)	6.383	208.078	0.377
LDCGM (exp)	6.162	197.455	0.411
NN	6.952	173.660	0.614
GAT + FNN (ours)	6.270	167.531	0.647
GAT ext. features + FNN (ours)	**5.418**	**151.024**	**0.669**
Northern part			
DCGM (pow)	5.774	86.26	0.028
DCGM (exp)	5.791	86.06	0.030
LDCGM (pow)	2.748	37.178	0.305
LDCGM (exp)	2.625	34.435	0.366
NN	3.312	34.332	0.524
GAT + FNN (ours)	**2.253**	29.995	0.531
GAT ext. features + FNN (ours)	2.333	**28.045**	**0.592**
Southern part			
DCGM (pow)	12.336	350.747	0.134
DCGM (exp)	0.158	339.042	12.122
LDCGM (pow)	4.677	115.026	0.520
LDCGM (exp)	4.471	107.402	0.555
NN	5.66	102.783	0.690
GAT + FNN (ours)	**4.318**	**97.125**	0.694
GAT ext. features + FNN (ours)	4.829	97.240	**0.695**
Eastern part			
DCGM (pow)	14.456	468.490	0.114
DCGM (exp)	14.472	469.144	0.112
LDCGM (pow)	6.215	209.031	0.417
LDCGM (exp)	5.828	193.848	0.469
FNN	6.677	148.136	0.653
GAT + FNN (ours)	5.003	121.045	0.698
GAT ext. features + FNN (ours)	**4.719**	**113.341**	**0.719**

pow/exp - power law/exponential distance decay function.
ext. features - extended nodes' features included in convolutions.

Acknowledgments. This research is financially supported by the Russian Science Foundation, Agreement 17-71-30029 (https://rscf.ru/en/project/17-71-30029/), with co-financing of Bank Saint-Petersburg.

References

1. 2GIS. City information service. www.2gis.ru
2. Google Maps. Satellite image. https://www.google.com.sg/maps/
3. Map data from OpenStreetMap. https://www.openstreetmap.org/copyright
4. Rosstat. https://rosstat.gov.ru/
5. Ashik, F.R., Mim, S.A., Neema, M.N.: Towards vertical spatial equity of urban facilities: An integration of spatial and aspatial accessibility. J. Urban Manag. **9**(1), 77–92 (2020)
6. Barbosa, H., et al.: Human mobility: Models and applications. Phys. Rep. **734**, 1–74 (2018)
7. Benamira, A., Devillers, B., Lesot, E., Ray, A.K., Saadi, M., Malliaros, F.D.: Semi-supervised learning and graph neural networks for fake news detection. In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 568–569 (2019)
8. Calabrese, F., Di Lorenzo, G., Ratti, C.: Human mobility prediction based on individual and collective geographical preferences. In: 13th International IEEE Conference On Intelligent Transportation Systems, pp. 312–317. IEEE (2010)
9. Dadashpoor, H., Rostami, F., Alizadeh, B.: Is inequality in the distribution of urban facilities inequitable? exploring a method for identifying spatial inequity in an iranian city. Cities **52**, 159–172 (2016)
10. Deming, W.E., Stephan, F.F.: On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Ann. Math. Stat. **11**(4), 427–444 (1940)
11. Doosti, B., Naha, S., Mirbagheri, M., Crandall, D.J.: Hope-net: A graph-based model for hand-object pose estimation. In: Proceedings of the IEEE/CVF Conference On Computer Vision And Pattern Recognition, pp. 6608–6617 (2020)
12. Eremin, R.A., Humonen, I.S., Zolotarev, P.N., Medrish, I.V., Zhukov, L.E., Budennyy, S.A.: Hybrid dft/data-driven approach for searching for new quasicrystal approximants in sc-x (x= rh, pd, ir, pt) systems. Crystal Growth Design **22**(7), 4570–4581 (2022)
13. Fan, C., Jiang, X., Lee, R., Mostafavi, A.: Equality of access and resilience in urban population-facility networks. npj Urban Sustainability **2**(1), 9 (2022)
14. Fang, X., Huang, J., Wang, F., Zeng, L., Liang, H., Wang, H.: Constgat: Contextual spatial-temporal graph attention network for travel time estimation at baidu maps. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2697–2705 (2020)
15. Farahani, R.Z., Fallah, S., Ruiz, R., Hosseini, S., Asgari, N.: Or models in urban service facility location: A critical review of applications and future developments. Eur. J. Oper. Res. **276**(1), 1–27 (2019)
16. Ferrer, A.L.C., Thome, A.M.T., Scavarda, A.J.: Sustainable urban infrastructure: A review. Resour. Conserv. Recycl. **128**, 360–372 (2018)
17. Grauwin, S., et al.: Identifying and modeling the structural discontinuities of human interactions. Sci. Rep. **7**(1), 46677 (2017)

18. Griffith, D.A., Fischer, M.M.: Constrained variants of the gravity model and spatial dependence: model specification and estimation issues. Springer (2016). https://doi.org/10.1007/978-3-319-30196-9_3
19. Hsu, C., Fan, C., Mostafavi, A.: Limitations of gravity models in predicting fine-scale spatial-temporal urban mobility networks. arXiv preprint arXiv:2109.03873 (2021)
20. Kontsevik, G., Sokol, A., Bogomolov, Y., Evstigneev, V.P., Mityagin, S.A.: Modeling the citizens' settlement in residential buildings. Procedia Comput. Sci. **212**, 51–63 (2022)
21. Lenormand, M., Bassolas, A., Ramasco, J.J.: Systematic comparison of trip distribution laws and models. J. Transp. Geogr. **51**, 158–169 (2016)
22. Li, Z., Ren, T., Ma, X., Liu, S., Zhang, Y., Zhou, T.: Identifying influential spreaders by gravity model. Sci. Rep. **9**(1), 8387 (2019)
23. Mishina, M., Khrulkov, A., Solovieva, V., Tupikina, L., Mityagin, S.: Method of intermodal accessibility graph construction. Proc. Comput. Sci. **212**, 42–50 (2022)
24. Oshan, T.M.: A primer for working with the spatial interaction modeling (spint) module in the python spatial analysis library (pysal). Region **3**(2), R11–R23 (2016)
25. Peregrino, A.A., Pradhan, S., Liu, Z., Ferreira, N., Miranda, F.: Transportation scenario planning with graph neural networks. arXiv preprint arXiv:2110.13202 (2021)
26. Robinson, C., Dilkina, B.: A machine learning approach to modeling human migration. In: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, pp. 1–8 (2018)
27. Schläpfer, M., et al.: The universal visitation law of human mobility. Nature **593**(7860), 522–527 (2021)
28. Simini, F., Barlacchi, G., Luca, M., Pappalardo, L.: A deep gravity model for mobility flows generation. Nat. Commun. **12**(1), 6576 (2021)
29. Simini, F., González, M.C., Maritan, A., Barabási, A.L.: A universal model for mobility and migration patterns. Nature **484**(7392), 96–100 (2012)
30. Temeljotov Salaj, A., Lindkvist, C.M.: Urban facility management. Facilities **39**(7/8), 525–537 (2021)
31. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
32. Wang, X., He, X., Cao, Y., Liu, M., Chua, T.S.: Kgat: Knowledge graph attention network for recommendation. In: Proceedings of the 25th ACM SIGKDD International Conference On Knowledge Discovery & Data Mining, pp. 950–958 (2019)
33. Zipf, G.K.: The p 1 p 2/d hypothesis: on the intercity movement of persons. Am. Sociol. Rev. **11**(6), 677–686 (1946)