Attentional Multi-graph Convolutional Network for Regional Economy Prediction with Open Migration Data

Fengli Xu
fenglixu2020@hotmail.com
Tsinghua University
Beijing, China

Yong Li*
liyong07@tsinghu.edu.cn
Tsinghua University
Beijing, China

Shusheng Xu
xssstory@gmail.com
Tsinghua University
Beijing, China

ABSTRACT

We study the problem of predicting regional economy of U.S. counties with open migration data collected from U.S. Internal Revenue Service (IRS) records. To capture the complicated correlations between them, we design a novel Attentional Multi-graph Convolutional Network (AMCN), which models the migration behavior as a multi-graph with different types of edges denoting the migration flows collected from heterogeneous sources of different years and different demographics. AMCN extracts high quality feature from the migration multi-graph by first applying customized aggregator functions on the induced subgraphs, and then fusing the aggregated features with a higher-order attentional aggregator function. In addition, we address the data sparsity problem with an important neighbor discovery algorithm that can automatically supplement important neighbors that are absent in the empirical data. Experiment results show our AMCN model significantly outperforms all baselines in terms of reducing the relative mean square error by 43.8% against the classic regression model and by 12.7% against the state-of-the-art deep learning baselines. In-depth model analysis shows our proposed AMCN model reveals insightful correlations between regional economy and migration data.

CCS CONCEPTS

• **Computing methodologies → Learning latent representations**; • **Information systems → Data mining**; • **Applied computing** → *Economics*.

KEYWORDS

Representation learning, mobility network, economy prediction.

ACM Reference Format:
Fengli Xu, Yong Li, and Shusheng Xu. 2020. Attentional Multi-graph Convolutional Network for Regional Economy Prediction with Open Migration Data. In *Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining USB Stick (KDD '20), August 23–27, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3394486.3403273

*This is the corresponding author

KDD '20, August 23–27, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7998-4/20/08…$15.00
https://doi.org/10.1145/3394486.3403273

1 INTRODUCTION

Uncovering the correlations between population migration and regional economy has been a long-standing research problem [1–3], which carries profound implications for policy making [4] and city planning [5]. Survey shows most of economists believe migration replenishes the labor market of the destination regions [6] and provides cash flow for the origination regions through remittances [7], and therefore can benefits the economy of both the destination and origination regions [4, 6]. However, empirical studies often show very mixed results (negative, positive or no impact) in real-world scenarios [8–10]. On the other hand, recent years have witnessed a proliferation of large-scale migration datasets collected from heterogeneous sources, which provides a unique angle to investigate this problem with a data-driven approach. In this paper, we aim to design a novel model for regional economy prediction based on the open migration datasets.

The main challenges of our research are three-fold. *First*, the correlations between regional economy and migration data are inherently complicated. For example, empirical studies on migration's impact yield highly volatile results that has no strong statistical support for theoretical predictions [4]. It indicates the proposed model needs to be expressive to capture complex correlations. *Second*, although including migration data from heterogeneous sources might benefit the prediction, it also poses challenges to the model's flexibility, which is required to jointly model the heterogeneous migration datasets, e.g., of different demographics and in different years. *Third*, the distribution of migration data is often significantly biases to the highly populated regions. Therefore, it is non-trivial to achieve accurate prediction for the less populated regions due to data sparsity.

To address these problems, we propose a novel predictive model, i.e., Attentional Multi-graph Convolutional Network (AMCN). The key idea is to model the heterogeneous migration datasets as a migration multi-graph, where each node represents a region and there are multiple edges between two nodes denoting the heterogeneous types migration flows. The proposed AMCN model can be broken down into three key components, which address the main challenges accordingly. *First*, to capture the complicated correlations, we extend the newly emergent graph convolutional networks (GCN) to model the heterogeneous migration multi-graph [11, 12]. Specifically, it uses customized aggregator functions to collect feature from the induced subgraphs of different migration flows, which allows it to simultaneously capture the structure of heterogeneous migration graph and region attribute. *Second*, to differentiate the semantics of heterogeneous migration flows, we design a higher-order aggregator function that leverages the attention mechanism to fuse the feature aggregated from the different induced subgraphs [13].

It allows the model to identify the most important features by dynamically assigning different attention weights to the aggregated features. *Third*, we design an important neighbor discovery algorithm to address the data sparsity issue in less populated regions. To be specific, by drawing strength from both theoretical migration model and deep learning technique, it accurately supplements the important neighbors to the less connected nodes in the empirical migration graph.

To evaluate the effectiveness of our proposed model, we leverage the public available U.S. Internal Revenue Service (IRS) records to extract a large-scale county-to-county migration data [1]. In addition, we use the county's GDP per captia statistics released by U.S. Bureau of Economic Analysis as the ground truth [2], which is one of the most adopted regional economic indicators. We evaluate our model's performance with the task of predicting the regional economy in 2015 with the migration data from 2013 to 2015. We make the following observations: *First*, the proposed model AMCN significantly outperforms all baselines in terms of reducing the relative mean square error by 43.8% against the classic support vector regression model and by 12.7% against the best deep learning baselines ($p<0.001$, two tailed Student's t-test). *Second*, the important neighbor discovery algorithm shows prominent effect in addressing the data sparsity issue in terms of supplementing more neighbors to the less connected nodes on average. *Third*, through visualizing the attention weight of the higher-order aggregator function, we make insightful observations on the different importance of the heterogeneous migration data sources. For example, the tax-exemptor's migration in current year and the taxpayer's migration in two years before carry most influence in the prediction of current year's regional economy.

The main contributions of our study can be summarized from the following aspects:

- We formally define the research problem of predicting regional economy with migration data in the framework of predicting node's label given migration multi-graph, which allows the predictive models to jointly model migration graph structure and region attribute.
- We propose a novel attentional multi-graph convolutional network to effectively carry out the predictive task. It consists of three key components: multi-graph convoluional network, attentional higher-order aggregator function and important neighbor discovery algorithm.
- We conduct extensive evaluations on real-world migration data among U.S. counties. Experiments show our model achieves significant performance gain against the state-of-art baselines, and in-depth model analysis shows the proposed model can reveal insightful correlations between regional economy and migration data.

2 RELATED WORKS

2.1 Economy and Migration Data

Migration's impact on the economy of destination and origination regions has been a constant research topic for the past decades [1–3], which has important influence over migration policy making and numerous applications. Previous studies in this area mostly focused on specific scenario and population. For example, survey showed economists and public generally believe high-skilled immigrants can improve the economy of destination regions [6]. On the other hand, many public members held the view that low-skilled immigrants have adverse impact on destination regions, while studies showed very mixed results in real-world scenarios [8, 9]. As for the origination regions, researchers found evidence that migration benefited the economy in terms of increasing investment through remittances [7] and increasing average wages for those remain [14]. In addition, studies also showed migration could promote trades in goods and service as well as technology innovation [15]. These works studied the implications of migration in specific circumstances. However, the investigation on how all kinds of migrations as a whole impact on the regional economy is woefully inadequate.

In summary, previous works mainly focus on the empirical and theoretical correlation analysis between the regional economy and migration. Different from them, we propose a predictive model — AMCN, which provides a principal framework to jointly model the impact of heterogeneous migration data, e.g., in different years, consist of different population and with different directions. We show that the proposed model achieves significant performance boost over classic regression model and is able to offer interpretable prediction results through the attention mechanism.

2.2 Mining Human Mobility Data

With the rapid proliferation of portable smart devices, human mobility data is now ubiquitously sensed and computed at population scale. To be specific, the newly emergent ubiquitous sensing technology facilitates the fine-grained mobility trace extraction from call detail record [16], mobile traffic consumption [17], GPS modules [18], crowd sensing [19] and IOT devices [20]. Such valuable data sources increasingly gain popularity in wide range of applications. At individual level, mobility data has been successfully exploited in monitoring mood [21], predicting social relation [22, 23] and improving application recommendation [24]. As for the population level mobility data, previous researches show promising results in urban activity detection [25], transportation scheduling [26] and improving intelligent urban sensing [19]. In addition, previous studies on migration data analytic mostly focus on understanding the patterns of migration behavior, such as predicting the churn rate of migrants [27] and understanding how they integrate into the native community [28].

In this paper, we study a novel problem of predicting regional economy with the migration data. Our study sheds light on how human mobility interplay with the economy growth, which has important implications for policy making and wide range of applications.

2.3 Graph Convolutional Network (GCN)

The key idea of GCNs is a end-to-end deep learning model that jointly captures the graph structure and node feature [12]. It is achieved by learning a spatial in-variant aggregator function to aggregate important feature from each node's neighborhood to predict its label. The recently emergent GCNs has rapidly set a series of new records in graph learning tasks, such as link prediction [29],

[1] https://www.irs.gov/statistics/soi-tax-stats-migration-data
[2] https://www.bea.gov/data/gdp/gdp-county

node classification [30] and community detection [11]. In addition, it has also been successfully applied in numerous applications, such as urban traffic forecast [31], social influence modelling [32] and recommender system [33]. Since our task relies on both the county's feature and the migration graph structure, the GCN framework naturally fits our research. However, the spatial in-variant aggregator of classic GCN models means applying same aggregator function to every node on every edge, which prevents it to capture the different semantics of different types of edges. Therefore, they cannot be readily applied in our problem. On the other hand, several recent works have generalized the GCN framework to heterogeneous graph by leveraging the Metapath technique [34], which essentially constructs a homogeneous graph by fitting the original graph to a template of edges, i.e. Metapath [35]. However, designing suitable Metapath is knowledge intensive process and the current models cannot adequately address the challenges in predicting regional economy with migration data, e.g., heterogeneous data sources and data sparsity. In this paper, we propose a novel attentional multi-graph convolutional network. It fundamentally extends the classic GCN framework to migration multi-graph, which achieves significant performance gain over state-of-art baselines.

3 PROBLEM DEFINITION AND CHALLENGES

3.1 Research Problem

To properly formulate the investigated research problem, we formally define the data structure and notations we use throughout this paper. Specifically, the migration record is defined as follow.

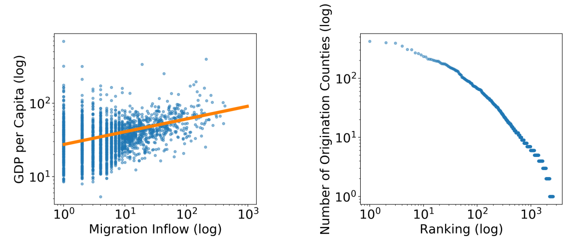
Definition 3.1. **Migration Record**. It is defined as a quadruple (c_i, c_j, t, n), which denotes the number of people n that migrate from county c_i to c_j at time t, with $n > 0$.

We organize the migration record as a migration graph among counties, which allows us to model the overall migration pattern as a whole.

Definition 3.2. **Migration Graph**. It is defined as a directed and weighted graph $\mathbf{G}_{t,s} = (\mathbf{V}, \mathbf{E}_{t,s})$, where $\mathbf{V}$ denotes the node set with node $v_i \in \mathbf{V}$ representing county c_i and $\mathbf{F}(v_i)$ denotes the attributes of node v_i. Each category s denotes a combination of population profile and directions. In addition, the weight of edge $\mathbf{E}_{t,s}(v_i, v_j)$ denotes the normalized number of people migrate from c_i to c_j (or from c_j to c_i w.r.t category s), which is computed by dividing the actual number of people with total number of people departing from (or arriving at w.r.t s) c_i.

In addition, the migration data often can be classified into different categories, i.e.. migration in different years, of different demographic (e.g., taxpayer and tax-exemptor) and with different directions (i.e., inflow and outflow). Therefore, we propose a multi-graph structure to simultaneously capture the heterogeneos migration behavior. That is there could be numerous parallel edges between two nodes that denote different migration data respectively. Specifically, it is formally defined as follows.

Definition 3.3. **Migration Multi-graph**. It is defined as a directed and weighted multi-graph $\mathbf{G} = \bigcup\{\mathbf{G}_{t,s}\} = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ denotes the node set that corresponds to counties and $\mathbf{F}(v_i)$ denotes the node attribute of $v_i \in \mathbf{V}$. In addition, $\mathbf{E} = \{\mathbf{E}_{t1,s1}, \mathbf{E}_{t2,s2}, \mathbf{E}_{t3,s3}, \ldots\}$ denotes the heterogeneous migration edge sets with $\mathbf{E}_{t1,s1}$ representing the migration data in $t1$ year and belong to $s1$ semantic categories.



(a) Correlation between regional economy and migration (b) Distribution of migration inflow

Figure 1: Visualizing the empirical patterns between migration data and regional economy across U.S. counties in 2015.

With the notations properly defined, we formulate the research problem as follow.

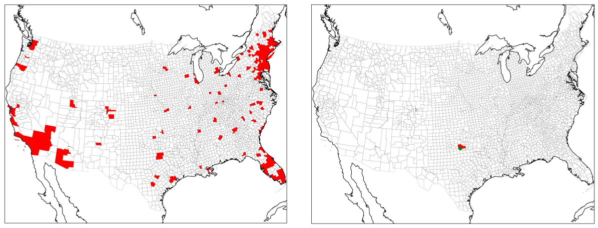
Problem 1. ***Regional Economy Prediction with Migration Data.*** *Given a migration multi-graph* $\mathbf{G} = (\mathbf{V}, \mathbf{E})$*, we aim to learn a prediction function* $\Phi(\mathbf{F}(v_i), \mathbf{E})$*. It takes both the node attribute* $\mathbf{F}(v_i)$ *and the graph structure in* $\mathbf{E}$ *as input, and output the prediction of regional economy* $\hat{y}_i$ *for each node* $v_i \in \mathbf{V}$.

3.2 Not Straightforward Path from Migration to Economy.

To showcase the empirical correlation between regional economy and migration, we present the GDP per capita and migration inflow of U.S. counties in 2015 in Figure 1(a). From the figure, we can observe that there is a general trend of positive correlation between them, which can be fitted by a linear function with 0.1734 slope coefficient in log-log plot. It indicates the regional economy generally is higher if it has higher migration inflow, which is consistent with the conventional wisdom. In addition, it suggests migration data does possess predictive signal on regional economy, which supports the feasibility of this task. However, we also notice that there is a large variation in the empirical data, where the maximum deviation in regional economy of counties with same migration inflow can reach to about 2 magnitude. Therefore, it cannot achieve accurate regional economy prediction by leveraging simple migration statistics, which suggests more sophisticated algorithms are needed to reveal the underlying correlation between them. Regional economy often correlated with each other through the flow of migration [1, 2]. Therefore, the predictive model should be able to capture the region's attribute as well as the migration graph structure, which poses a significant challenge to the classic feature based regression models.

3.3 Data sparsity for Less Populated Counties.

Finally, the unevenly distributed migration data is another challenge for achieving accurate prediction. That is the migration records mostly take place in the counties with large population, while



(a) Origination Counties to New York (b) Origination Counties to Haskell

Figure 2: An example of uneven data distribution in empirical migration data.

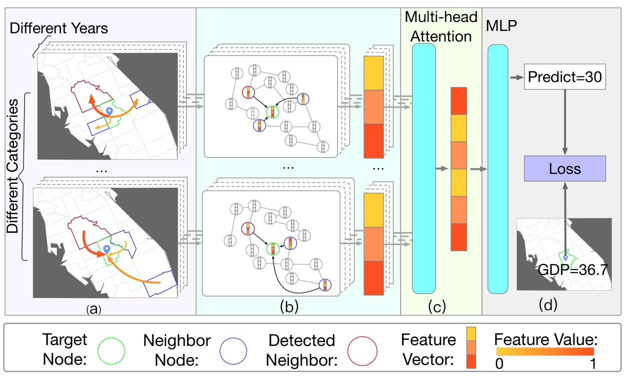


Figure 3: Overview of the attentional multi-graph convolutional network for regional economy prediction with migration data.

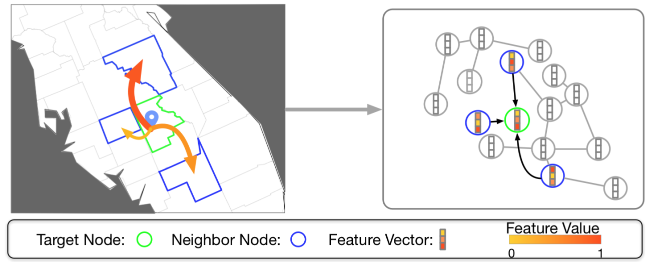


Figure 4: Illustration of graph convolutional network on migration subgraph (Best view in color).

Algorithm 1 : Aggregator $\Psi_{t,s}$ for migration graph of year t and category s.

Input: Node embedding set $\{\mathbf{H}_{t,s}(v)|\forall v \in \mathbf{V}\}$, migration graph $\mathbf{E}_{t,s}$, target node v;

Output: Updated embedding $\mathbf{H}'_{t,s}(v)$ of node v;

$\mathcal{N}_{t,s}(v) \leftarrow \{w| \; \forall \mathbf{E}_{t,s}(v,w) > 0\}$

$\mathbf{h}_c \leftarrow \sum_{\forall w \in \mathcal{N}_{t,s}(v)} \mathbf{E}_{t,s}(v,w)\mathbf{H}_{t,s}(w)$

$\mathbf{H}'_{t,s}(v) \leftarrow \mathrm{ReLU}\left(\mathbf{W}_{t,s}\mathrm{concat}(\mathbf{H}_{t,s}(v),\; \mathbf{h}_c) + \mathbf{b}_{t,s}\right)$

the relative smaller counties have few migration records, which might hinder the prediction accuracy. Figure 1(b) shows the number of origination counties distribution in 2015, which denotes the counties have immigrant from how many other counties. We can observe that the distribution follows a well defined power law, with 1972 (76.1%) counties have less than 10 origination counties. In addition, Figure 2 visualizes the origination counties of New York county and Haskell county. New York county has immigration from all over the country, while Haskell only has migration from a neighboring county. Moreover, Figure 1(a) shows the county with less migration inflow generally has higher variation in regional economy, which indicates the less connected counties are more difficult to predict. Therefore, the data sparsity for less populated counties presents a challenge to achieve accurate regional economy prediction.

4 METHOD

To achieve accurate prediction, we propose a novel Attentional Multi-graph Convolutional Network (AMCN), which is illustrated in Figure 3. Now, we elaborate on the design of each module as following.

4.1 Preliminary: Graph Convolutional Network on Migration Multi-graph

The key idea behind graph convolutional networks (GCN) is to learn a powerful aggregator function to capture the important feature for each node's prediction from its neighborhood on graph. That is each node's prediction is based on its own attribute and the context of the neighbor nodes' attribute, which allows the predictive model to jointly model the node attribute and graph structure. However, classic GCNs are designed to model homogeneous graph [11, 12], i.e., the graph with only one type of node and edge. When modeling the migration data from different years, of different population profile and with different directions, it cannot explicitly differentiate the semantic differences of heterogeneous edges in migration multi-graph. To remedy this problem, we extend the classic GCN framework to account for semantics of heterogeneous edges.

Specifically, we propose to first aggregate the feature on the subgraphs that induced by certain types of edges with customized aggregator functions respectively, and then fuse the features from subgraphs with a higher-order aggregator function. Algorithm 1 describes the customized aggregator function $\mathbf{H}_{t,s}(v)$ for each subgraph. The underlying intuition is that the counties have higher migration flows with the target county should have higher weight in predicting its economy. The computation process is also illustrated in Figure 4. The left figure shows the empirical migration graph with thicker arrows in deeper color representing more migrants, and the right figure shows the feature aggregate process.

Based on the customized aggregator function $\Psi_{t,s}$, we design the complete model, i.e., multi-graph convolutional network (MCN), showing in Algorithm 2. Specifically, we first iteratively sample out the l-hop neighborhood of target node on the subgraphs induced by certain type of edge set $\mathbf{E}_{t,s}$ (Line 3-6), which is storied in $\{R_{t,s}(i), for\ i = 0, 1, ..., l\}$. Then, we iteratively aggregate the node

Algorithm 2 : Multi-graph Convolutional Network

Input: Migration multi-graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, node attributes $\mathbf{F}$, aggregator functions $\{\Psi_{t,s}, \forall \mathbf{E}_{t,s} \in \mathbf{E}\}$, target node v;
Output: Predicted GDP per capita $\hat{\mathbf{y}}(v)$ for node v;
1: **for** $\forall \mathbf{E}_{t,s} \in \mathbf{E}$ **do**
2: /*Iteratively sample the neighbors in this migration graph */
3: $R_{t,s}(0) \leftarrow v$
4: **for** $i = 1$ to l **do**
5: $R_{t,s}(i) \leftarrow \bigcup \{\mathcal{N}_{t,s}(w) | \; \forall \; w \in R_{t,s}(i-1)\}$
6: **end for**
7: /*Aggregating features */
8: $\mathbf{H}_{t,s}(w) \leftarrow \mathbf{F}(w), \forall w \in R_{t,s}(l)$
9: **for** $i = l$ to 1 **do**
10: $\mathbf{H}_{t,s}(w) \leftarrow \Psi_{\mathrm{t,s}}(\mathbf{H}_{t,s}, \mathbf{E}_{t,s}, w), \forall w \in R_{t,s}(i-1)$
11: **end for**
12: **end for**
13: $\mathbf{h}(v) \leftarrow \text{Average}(\{\mathbf{H}_{t,s}(v), \forall \mathbf{E}_{t,s} \in \mathbf{E}\})$
14: $\hat{\mathbf{y}}(v) \leftarrow \text{Predictor}(\mathbf{h}(v))$

feature from the sampled l-hop neighborhood back to the target node v with the corresponding customized aggregator function $\Psi_{t,s}$ in each hop (Line 8-11). The feature captured through each types of edges $\mathbf{H}t, s(v)$ is fused with a higher-order aggregator function (Line 13). Without loss of generality, we choose the average function here for simplicity. The fused embedding vector is fed into a 2-layers fully connected multi-layer perceptron (MLP) [36] to predict the economic status of that county. Since the prediction task is a regression problem, we adopt the relative Mean Square Error (MSE) to minimize the error of the model's prediction, which is computed as follows,

$$\mathbb{O} = \frac{1}{|V|} \sum_{\forall v \in \mathbf{V}} \left(\frac{\hat{\mathbf{y}}(v) - \mathbf{y}(v)}{\mathbf{y}(v)}\right)^2, \quad (1)$$

where $\hat{\mathbf{y}}(v)$ is the predicted result on node v and $\mathbf{y}(v)$ is the ground truth.

4.2 Differentiating the Features from Heterogeneous Migration Data

Although choosing average function as the higher-order aggregator function benefits the model from simplicity, it has limitations in differentiating the importance of features from heterogeneous edges for different counties. For example, the migration patterns of taxpayers and tax-exemptors might carry different importance for two different counties due to the migration policy and region population profile. However, simple average function models both features equivalently and statically, which prevents the predictive model to capture the complex correlation between migration data and regional economy.

To address this problem, we design a higher-order aggregator function that draws inspiration from the multi-head attention mechanism [13], which is illustrated in Figure 5. The key idea is to learn an additional deep learning model to predict the importance of the features from different subgraphs dynamically. Since attention mechanism can effectively prioritize the importance of different neighboring nodes [30], we design a multi-head attention module

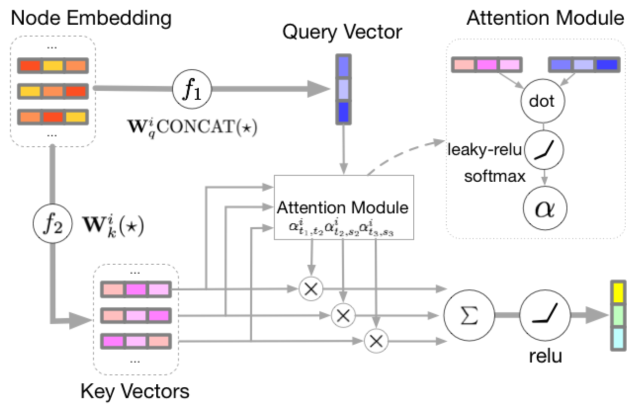


Figure 5: Illustration of the attentional higher-order aggregator function.

to serve as the higher-order aggregator function. It consists of multiple parallel attention modules which are referred to as "attention head", and each attention head outputs separate set of weight for the feature vectors independently. The underlying intuition is that the parallel attention heads can project the features into different semantic space and capture the important features from different perspective [13]. Specifically, given a m dimension feature $\mathbf{H}_{t,s}(v)$ captured from $\mathbf{E}_{t,s}$ edges, the attention weight $\alpha^i_{t,s}(v)$ of the i-th attention head on this feature is computed as follow,

$$\begin{aligned} Q^i_{t,s}(v) &= \mathbf{W}^i_q \text{CONCAT}(\{\mathbf{H}_{t,s}(v), \forall \mathbf{E}_{t,s} \in \mathbf{E}\}), \\ K^i_{t,s}(v) &= \mathbf{W}^i_k \mathbf{H}_{t,s}(v), \\ \alpha^i_{t,s}(v) &= \text{SOFTMAX}\big(\delta\big(Q^i_{t,s}(v) \cdot K^i_{t,s}(v)\big)\big), \end{aligned} \quad (2)$$

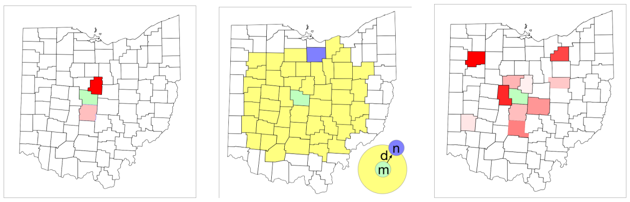
where $\mathbf{W}^i_q$ is a learnable weight matrix with $m/n \times m * e$ dimensions that projects the concatenated feature vectors from all types of edges to a m/n dimension query vector $Q^i_{t,s}(v)$, with n denoting the number of attention heads and e denoting the number of edge types. In addition, $\mathbf{W}^i_k$ is a $m/n \times m$ learnable weight matrix that projects the feature vector $\mathbf{H}_{t,s}(v)$ to a m/n dimension key vector $K^i_{t,s}(v)$. The attention weight is computed as the inner product of the query vector and key vector, which is activated with the widely adopted *Leaky_ReLU* function $\delta(\star)$ [37] and normalized with $SoftMax$ function [38]. The computation of the attention weight is illustrated in Figure 5. Finally, the fused feature $\mathbf{H}'(v)$ for node v is computed as follow,

$$\mathbf{H}'(v) = \text{CONCAT}_i\big(\{\text{ReLU}\big(\sum_{\forall \mathbf{E}_{t,s} \in \mathbf{E}} \alpha^i_{t,s}(v) \mathbf{K}^i_{t,s}(v)\big)\}\big), \quad (3)$$

where each attention head outputs the weighted sum the key vector $\mathbf{K}^i_{t,s}(v)$ activated with *ReLU* function. The output of each attention head is concatenated together as the final fused feature.

4.3 Important Neighbors Discovery

One important obstacle for achieving accurate regional economy prediction is the data sparsity issue for the less connected counties. That is a large portion of counties only have migration records with



(a) Empirical Migration (b) Candidate Counties (c) Detected Neighbors

Figure 6: Illustration of the important neighbor discovery algorithm.

small amount of other counties. For example, Figure 6(a) shows the target county (green color) only has migration flow with two counties, which is marked red and deeper color indicating higher migration. Therefore, automatically discovering and supplementing the important neighbor counties (i.e. counties that have migration flow with target node) that are missed out in the empirical migration data is of significant importance to improve prediction accuracy.

We find inspiration from the prevalent theoretical model of human migration, i.e., gravity model [39–41]. Originated from the physical gravity law, it predicts the migration flow between two regions to be positively correlated with the product of two region's population and negatively correlated with the square of the distance between them, which is illustrated in Figure 6(b). Specifically, the migration flow between the blue county and green county is predicted to be $m * n/d^2$, where m, n denote the population of these two counties and d denotes the distance between them. The basic assumption is that nearby regions should be more connected and highly populated regions are more influential [40]. However, gravity model only generates coarse-grained predictions. For example, two county pairs with identical populations and distance can have different migration flows due to different migration policy. On the contrary, utilizing deep learning technique to predict the important neighbors is also difficult, since there are more than 2,500 counties as potential candidates. Therefore, we aim to combine the strength of gravity model and deep learning technique to achieve accurate neighbors discovery.

Specifically, we first leverage the gravity model to identify the most plausible candidates for each county to form candidate pool $\mathbf{P}(v)$, which can be derived as follow,

$$\mathbf{P}(v) = \{w | \frac{\rho(w)}{\text{Distance}(v, w)^2} > \theta, \forall w \in \mathbf{V}\}, \tag{4}$$

where $\rho(w)$ denotes the population in county w and θ is a predefined threshold. By leveraging the gravity model, we dramatically narrow down the candidate neighbors of each county. For example, the yellow color counties in Figure 6(b) denote the selected candidates. Moreover, we design a deep learning model to further select the neighbor counties and fine tune the migration flows between them. It is built on top of attention mechanism [13], where the learned attention weight between two counties served as the normalized predicted migration flow between them. Specifically, given target county v and a candidate neighbor county w, the attention weight between them can be computed as following,

$$\gamma_{t,s}(v, w) = \text{SoftMax}\big(\delta\big(\mathbf{F}(w)\mathbf{W}_n\mathbf{F}(v)\big)\big), \tag{5}$$

where $\mathbf{F}(\star)$ denotes the node feature and $\mathbf{W}_n$ is a learnable weight matrix that linearly transformed the node feature vector to attention space. $\delta(\star)$ denotes the activation function, which is chosen as the widely adopted *Leaky_ReLU* function [37]. The attention coefficient $\gamma_{t,s}(v, w)$ is computed as the activated value normalized with $SoftMax$ function [38]. The learned attention coefficient capture the importance of the candidate neighbor nodes in predicting the target node's label. Therefore, we use it as the weight of the newly added edges between target node and candidate nodes, with 0 indicating the candidate nodes are not selected as supplemented neighbors. The discovered neighbors are illustrated in Figure 6(c), where red color denotes the supplemented neighbors and deeper color indicates higher predicted migration flow. Specifically, the updated edge set $\mathbf{E}'_{t,s}(v, w)$ is computed as follows,

$$\mathbf{E}'_{t,s}(v, w) = \mathbf{E}_{t,s}(v, w) + \gamma_{t,s}(v, w), \text{ for } \forall w \in \mathbf{P}(v). \tag{6}$$

It is worth pointing out that the important neighbor discovery model also adopts the multi-head attention structure [13]. That is there can be multiple parallel attention heads with each head attending to a specific feature, i.e., detecting the important neighbors with a certain characteristics.

5 EXPERIMENT

5.1 Dataset

We evaluate the proposed model on the migration data extracted from U.S. Internal Revenue Service (IRS) public records. It is based on year-to-year address changes reported on individual tax returns and tax exemptions filed with the IRS and we use the county-to-county migration data from 2012 to 2015. Based on the combination of **inflow** and **outflow** with **taxpayer** and **tax-exemptor**, we construct 4 directed migration graphs for each year and finally we obtain 16 migration graphs in total. To ensure the connectivity of the graphs, we iteratively filter out the counties have no migration records. As a result, we obtain 2593 counties to evaluate different methods, which is 82.5% of the total number of counties. Besides, we employ GDP per capita statistics in 2015 published by Bureau of Economic Analysis as the groundtruth, which is one of the most adopted regional economic indicators. The total number of migrants is 61,058,028, while the average number of migrants and GDP per capita is 23,547.25 and 37.73 thousand dollars, respectively.

5.2 Experiment Settings

Baselines We compare our model with two categories of baselines: classic regression models and graph representation learning models. The classic regression models directly map the node attribute to the predicted GDP per capita. On the other hand, since the main contribution of our model is to extract powerful representation for each node, we also compare our model with the state-of-art graph representation learning models. The node embedding extracted by these baselines are also fed into a 2-layers MLP for prediction, which is consistent with our model's predictor module. In addition, we also report the performance of two degraded variants of our model

Table 1: Performance comparison with baseline models, where (∗∗) indicates p<0.001 significance over best baseline.

Method	MSE	Standard Deviation	Relative Improvement
SVR	0.2793	0.3709	–
NODE2VEC	0.1800	0.2522	35.56%
DEEPWALK	0.2219	0.2860	20.55%
LINE	0.1880	0.2418	32.69%
MCN	**0.1644**∗∗	0.2371	41.14%
AMCN(w/o ND)	**0.1591**∗∗	0.2373	43.04%
AMCN	**0.1571**∗∗	0.2374	43.75%

to show the performance gain of each key components. Specifically, the baseline models are introduced as follow.

- SVR [42]: It is the classic support vector regression model. It is a powerful regression model to capture complex correlation without deep learning.
- DeepWalk [43]: It is a state-of-art representation learning model to capture graph structure by simulating truncated random walks in the graph. This approach only supports the graph with binary edges.
- Node2Vec [44]: It is an extended version of DeepWalk, which can model weighted edges with biased random walks.
- LINE [45]: It is a state-of-art representation learning model that can efficiently learn node representation to preserve both proximity and structural role on graph.

There are two variations of our model:

- MCN: The vanilla version of our model, which not includes the attentional higher-order aggregator function and neighbor discovery mechanism. It also serves as a baseline of GCN.
- AMCN(w/o ND): Our complete model without neighbor discovery mechanism.

To evaluate the performance of different methods, we adopt relative Mean Square Error(MSE) as the metric, which is computed by relative error to avoid the impact of uneven distributed ground truth. To improve stability, we randomly split the counties into 5 groups and perform cross-validation by training on four groups and test on the other group iteratively.

5.3 Overall Performance

The experiment results are reported in Table 1. From the results, we make the following observations.

1) The proposed AMCN model outperforms all the baselines on both effectiveness and robustness. Specifically, it provides relative performance gain of 43.8%(p<0.001) over SVR and 12.7% (p<0.001) over best deep learning baseline. In addition, its standard deviation is much smaller than SVR and deep learning baselines. This result demonstrate that the AMCN model is able to successfully aggregate information from the multi-graph to uncover the complex correlation between regional economy and migration.

2) AMCN and its variants outperform the deep learning baselines, while the deep learning baselines perform better than the classic SVR. It suggests that migration graph is indeed conducive to regional economy prediction, therefore it is important to explicitly capture the graph structure. In addition, our model significantly outperforms previous deep learning methods in capturing the feature of migration graph.

3) We find AMCN performs best while MCN is poorest among these three variants. It indicates both neighbor discovery and higher-order aggregator are effective in terms of consistent performance boost.

To conclude, AMCN and its variants significantly outperform all the baselines, and have stronger robustness at the same time. In addition, ablation study shows the key components of neighbor discovery mechanism and higher-order aggregator are effective in improving the prediction accuracy.

Table 2: Higher-order aggregator function's attention weight on heterogeneous migration data.

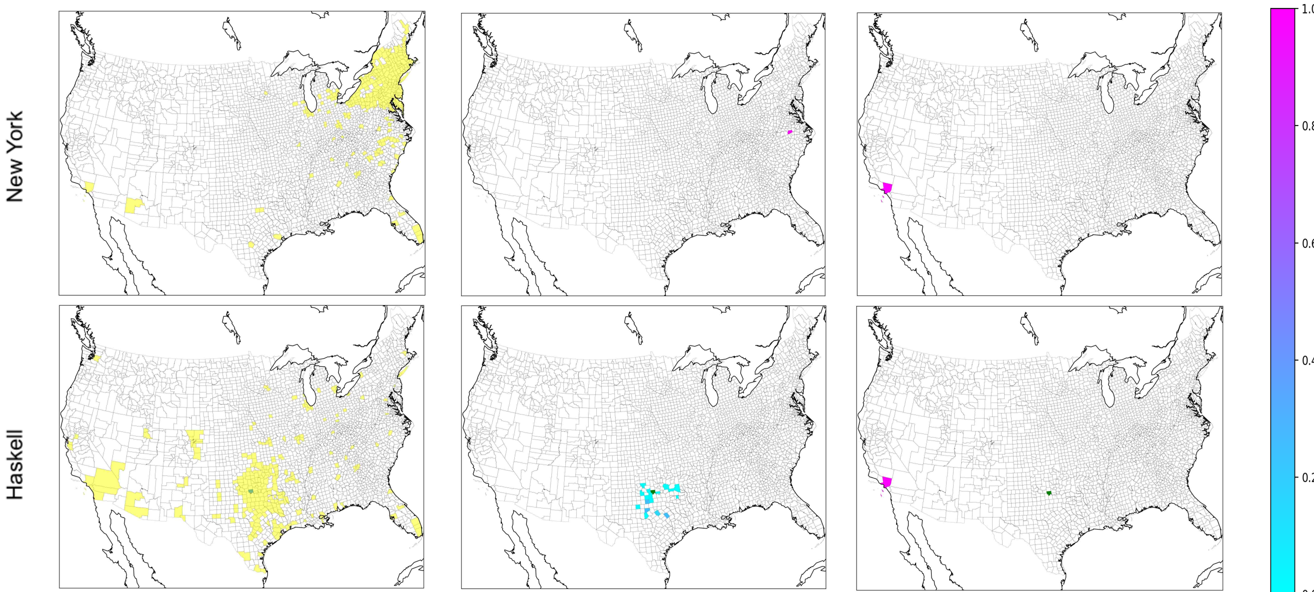
Year	Category				Attention Head
	Taxpayer		Tax-exemptor		
	Outflow	Inflow	Outflow	Inflow	
2012	0	0	0	0	Head 1
2013	0	0	0.001	0	
2014	0	0.004	0.008	0.008	
2015	0.078	0.156	0.353	0.392	
2012	0	0	0.001	0.002	Head 2
2013	0	0	0.001	0.003	
2014	0.001	0.008	0.009	0.011	
2015	0.080	0.160	0.339	0.384	
2012	0.149	0.084	0.018	0.027	Head 3
2013	0.347	0.235	0.040	0.096	
2014	0.002	0.002	0	0.001	
2015	0	0	0	0	

5.4 In-depth Model Analysis

To better understand AMCN, we take a series of in-depth analysis based on higher-order aggregator and neighbors discovery as flows.

Attention Weight. Firstly, we visualize the attention heads in higher-order aggregator in Table 2. The Table is divided into 3 parts as there are 3 heads in the higher-order aggregator. For every head, the number in a cell represents how important the embedding generated by corresponding graph contributes to the result. Specially, different rows in the a head mean the corresponding graphs are constructed by the migration data from different years while different columns represent different categories of the graphs(inflow taxpayer, outflow taxpayer, inflow tax-exemptor and outflow tax-exemptor).

Intuitively, the model indeed learns the various representation with different significance of graphs in the three heads, as we can observe that the first head and second head mostly focus on the 2015 while the third head take more attention on 2012 and 2013. More interestingly, it seems that the first two columns in the third head are much darker in 2012 and 2013 while the other two heads are more attentive to the last two columns in 2014 and 2015, which may indicates that the tax-exemptor migration have a immediate effect to the regional economy when the influence of the taxpayer migration exists some delays, where the delays may be caused by the lag of tax effectiveness.



(a) Candidate Counties (b) First Neighbor Detector (c) Second Neighbor Detector

Figure 7: Neighbor discovery algorithm's performance on New York and Haskell.

Neighbor Discovery. Here we select two counties to visualize the neighbor discovery. The inflow migration data for New York and Haskell has been shown in the Figure 2 (a) and (b), it indicates that New York has many neighbors while Haskell has only one neighbor. So what would happen after neighbor discovery? The process for neighbor discovery in showed as Figure 7 where the upper row corresponds to New York and the the bottom row correspond to Haskell. Figure 7 (b) presents the recommended neighbors by the gravity model with yellow and the selected county (New York or Haskell) with green, Figure 7 (c) and (d) presents two neighbor detectors of ultimately neighbors discovered where various color symbolizes different weights or significance. Figure 7 (b) shows that gravity model would recommend nearly the same number of neighbors for different counties, but the following attention mechanism would learn differences for counties. More interestingly, Figure 7 (c) and (d) give us the feeling that there are few neighbors would be added to the county which already has many neighbors (New York) when the less connected county (Haskell) prefer to discover more neighbors.

To validate its effect in addressing data sparsity issue, we research on the correlation between the the number of migration neighbors and discovered neighbors as shown in Figure 8. We can observe that there is a general trend of negative correlation between them, which can be fitted by a linear function with -0.2056 slope coefficient in log-log scale. It indeed represents that neighbor discovery can help less connected counties to discover more neighbors to some extent.

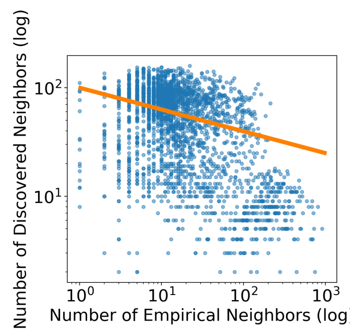


Figure 8: Correlation between the number of empirical neighbors and discovered neighbors.

6 CONCLUSION

In this paper, we investigate the long-standing research question of predicting regional economy with migration data. We propose a novel attentional multi-graph convolutional network to effectively aggregate feature from the heterogeneous migration behavior between regions. Extensive experiments on real-world migration data among U.S. counties show our model can significantly outperform all baseline methods in terms of reducing the relative mean square error by 43.8% against the classic SVR model and by 12.7% against the best deep learning baselines. Besides, in-depth model analysis reveals insightful correlations, e.g., the migration of tax-extemptors tends to impact on the regional economy immediately while the migration of taxpayers has a two year delay to reach peak impact.

ACKNOWLEDGMENTS

This work was supported in part by The National Key Research and Development Program of China under grant 2018YFB1800804, the National Nature Science Foundation of China under U1936217, 61971267, 61972223, 61941117, 61861136003, Beijing Natural Science Foundation under L182038, Beijing National Research Center for Information Science and Technology under 20031887521, and research fund of Tsinghua University - Tencent Joint Laboratory for Internet Innovation Technology.

REFERENCES

[1] Brinley Thomas. *Migration and economic growth: a study of Great Britain and the Atlantic economy*, volume 12. CUP Archive, 1973.
[2] Oded Stark and David E Bloom. The new economics of labor migration. *The american Economic review*, 75(2):173–178, 1985.
[3] Adrian J Bailey. Worlds in motion: Understanding international migration at the end of the millennium, 2001.
[4] Hendrik Van den Berg and Örn B Bodvarsson. *The Economics of Immigration: Theory and Policy*. Springer, 2009.
[5] Doug Saunders. *Arrival city: How the largest migration in history is reshaping our world*. Vintage, 2011.
[6] David Card, Christian Dustmann, and Ian Preston. Immigration, wages, and compositional amenities. *Journal of the European Economic Association*, 10(1):78–119, 2012.
[7] Claire L Adida and Desha M Girod. Do migrants improve their hometowns? remittances and access to public services in mexico, 1995-2000. *Comparative Political Studies*, 44(1):3–27, 2011.
[8] Mette Foged and Giovanni Peri. Immigrants' effect on native workers: New analysis on longitudinal data. *American Economic Journal: Applied Economics*, 8(2):1–34, 2016.
[9] Sari Pekkala Kerr and William R Kerr. Economic impacts of immigration: A survey. Technical report, National Bureau of Economic Research, 2011.
[10] Joao Pedro Azevedo, Judy S Yang, and Osman Kaan Inan. *What are the impacts of Syrian refugees on host community welfare in Turkey? a subnational poverty analysis*. The World Bank, 2016.
[11] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.
[12] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.
[14] Benjamin Elsner. Emigration and wages: The eu enlargement experiment. *Journal of International Economics*, 91(1):154–163, 2013.
[15] Jennifer Hunt and Marjolaine Gauthier-Loiselle. How much does immigration boost innovation? *American Economic Journal: Macroeconomics*, 2(2):31–56, 2010.
[16] Zhihan Fang, Fan Zhang, Ling Yin, and Desheng Zhang. Multicell: Urban population modeling based on multiple cellphone networks. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):106, 2018.
[17] Fengli Xu, Pengyu Zhang, and Yong Li. Context-aware real-time population estimation for metropolis. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1064–1075. ACM, 2016.
[18] Farbod Faghihi and Petteri Nurmi. An empirical study on the regularity of route mobility. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 1418–1425. ACM, 2016.
[19] Shenggong Ji, Yu Zheng, and Tianrui Li. Urban sensing based on human mobility. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1040–1051. ACM, 2016.
[20] Aku Visuri, Kennedy Opoku Asare, Elina Kuosmanen, Yuuki Nishiyama, Denzil Ferreira, Zhanna Sarsenbayeva, Jorge Goncalves, Niels van Berkel, Greg Wadley, Vassilis Kostakos, et al. Ubiquitous mobile sensing: Behaviour, mood, and environment. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pages 1140–1143. ACM, 2018.
[21] Luca Canzian and Mirco Musolesi. Trajectories of depression: Unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 1293–1304. ACM, 2015.
[22] Fei Yi, Zhiwen Yu, Qin Lv, and Bin Guo. Toward estimating user-social event distance: mobility, content, and social relationship. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 233–236. ACM, 2016.
[23] Yanwei Yu, Hongjian Wang, and Zhenhui Li. Inferring mobility relationship via graph embedding. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):147, 2018.
[24] Donghan Yu, Yong Li, Fengli Xu, Pengyu Zhang, and Vassilis Kostakos. Smartphone app usage prediction using points of interest. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):174, 2018.
[25] Tong Xia, Yue Yu, Fengli Xu, Funing Sun, Diansheng Guo, Depeng Jin, and Yong Li. Understanding urban dynamics via state-sharing hidden markov model. In *The World Wide Web Conference*, pages 3363–3369. ACM, 2019.
[26] Stefan Foell, Gerd Kortuem, Reza Rawassizadeh, Santi Phithakkitnukoon, Marco Veloso, and Carlos Bento. Mining temporal patterns of transport behaviour for predicting future transport usage. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing*, pages 1239–1248. ACM, 2013.
[27] Yang Yang, Zongtao Liu, Chenhao Tan, Fei Wu, Yueting Zhuang, and Yafeng Li. To stay or to leave: Churn prediction for urban migrants in the initial period. In *Proceedings of the 2018 World Wide Web Conference*, pages 967–976. International World Wide Web Conferences Steering Committee, 2018.
[28] Yang Yang, Chenhao Tan, Zongtao Liu, Fei Wu, and Yueting Zhuang. Urban dreams of migrants: A case study of migrant integration in shanghai. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
[29] Jinyin Chen, Xuanheng Xu, Yangyang Wu, and Haibin Zheng. Gc-lstm: Graph convolution embedded lstm for dynamic link prediction. *arXiv preprint arXiv:1812.04206*, 2018.
[30] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.
[31] Binbing Liao, Jingqing Zhang, Chao Wu, Douglas McIlwraith, Tong Chen, Shengwen Yang, Yike Guo, and Fei Wu. Deep sequence learning with auxiliary information for traffic prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 537–546. ACM, 2018.
[32] Jiezhong Qiu, Jian Tang, Hao Ma, Yuxiao Dong, Kuansan Wang, and Jie Tang. Deepinf: Social influence prediction with deep learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2110–2119. ACM, 2018.
[33] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 974–983. ACM, 2018.
[34] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *The World Wide Web Conference*, WWW '19, pages 2022–2032, New York, NY, USA, 2019. ACM.
[35] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *Proceedings of the VLDB Endowment*, 4(11):992–1003, 2011.
[36] Matt W Gardner and SR Dorling. Artificial neural networks (the multilayer perceptron)âĂŤa review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14-15):2627–2636, 1998.
[37] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015.
[38] Roland Memisevic, Christopher Zach, Marc Pollefeys, and Geoffrey E Hinton. Gated softmax classification. In *Advances in neural information processing systems*, pages 1603–1611, 2010.
[39] George Kingsley Zipf. The p 1 p 2/d hypothesis: on the intercity movement of persons. *American sociological review*, 11(6):677–686, 1946.
[40] Marc Barthelemy. Spatial networks. *Phys. Rep.*, 499:1–101, 2010.
[41] Woo-Sung Jung, Fengzhong Wang, and H Eugene Stanley. Gravity model in the korean highway. *EPL (Europhysics Letters)*, 81(4):48005, 2008.
[42] Harris Drucker, Christopher JC Burges, Linda Kaufman, Alex J Smola, and Vladimir Vapnik. Support vector regression machines. In *Advances in neural information processing systems*, pages 155–161, 1997.
[43] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710. ACM, 2014.
[44] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864. ACM, 2016.
[45] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on world wide web*, pages 1067–1077. International World Wide Web Conferences Steering Committee, 2015.