How Business Agglomeration Affects Individual Points-of-Interest: A Causal Effect Estimation Perspective

HAORAN XIN, Thrust of Artificial Intelligence, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
XINJIANG LU, Baidu Frontier Research Department, Beijing, China
YING SUN, Thrust of Artificial Intelligence, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
NENGJUN ZHU, School of Computer Engineering and Science, Shanghai University, Shanghai, China
TONG XU, University of Science and Technology of China, Hefei, China
JINGBO ZHOU, Baidu Frontier Research Department, Beijing, China
HUI XIONG, Thrust of Artificial Intelligence, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China and Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China

In modern cities, there is an increasing trend for the development of business agglomeration, which can foster the prosperity of individual businesses by clustering stores and industries. Recently, the advent of **Point-of-Interest (POI)** data enables a new paradigm for studying the causal effect of business agglomeration in a data-driven way. To this end, we aim to quantify the contribution of the agglomeration effect to the check-in volume at POIs. This is a non-trivial causal effect estimation task due to the higher-order spatial interference typically exhibited by the agglomeration distribution. Moreover, the confounding bias can be exacerbated due to the complex spatial and functional properties inherent to confounders. Therefore, we propose a **Causal effect estimation framework for AgglomeRation Effect (CARE)** measurement, which includes a **Spatial Interference Diffusion Network (SIDN)** and a **Disentangled Propensity Estimator (DPE)**. SIDN captures spatial interference by spreading the treatment effect among POIs through a dedicated spatial agglomeration hypergraph. Then, DPE models a POI's propensity of receiving the treatment and

This work was supported in part by the National Key R&D Program of China (Grant No. 2023YFF0725001), in part by the National Natural Science Foundation of China (Grant No. 92370204, 62306255, 62202282), in part by the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2023B1515120057), in part by the Key-Area Special Project of Guangdong Provincial Ordinary Universities (Grant No. 2024ZDZX1007).
Authors' Contact Information: Haoran Xin (corresponding author), Thrust of Artificial Intelligence, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China; e-mail: hxin883@connect.hkust-gz.edu.cn; Xinjiang Lu, Baidu Frontier Research Department, Beijing, China; e-mail: luxinjiang@baidu.com; Ying Sun, Thrust of Artificial Intelligence, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China; e-mail: sunyinggilly@gmail.com; Nengjun Zhu, School of Computer Engineering and Science, Shanghai University, Shanghai, China; e-mail: zhu_nj@shu.edu.cn; Tong Xu, University of Science and Technology of China, Hefei, China; e-mail: tongxu@ustc.edu.cn; Jingbo Zhou, Baidu Frontier Research Department, Beijing, China; e-mail: zhoujingbo@baidu.com; Hui Xiong, Thrust of Artificial Intelligence, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China and Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China; e-mail: xionghui@ust.hk.
This work was done when the first author was an intern at Baidu under the supervision of the second author.
* For correspondence, please contact Xinjiang Lu and Hui Xiong.



This work is licensed under Creative Commons Attribution International 4.0.

© 2026 Copyright held by the owner/author(s).
ACM 1556-472X/2026/3-ART46
https://doi.org/10.1145/3793854

further unravels the spatial and inherent aspects of propensity by disentangled learning objectives. In addition, we incorporate SIDN and DPE into a unified causal effect estimation architecture using neural Robinson decomposition. Finally, extensive experiments on three real-world datasets validate the effectiveness and universality of CARE for measuring the agglomeration effect.

CCS Concepts: • **Mathematics of computing → Causal networks**; • **Information systems → Business intelligence**;

Additional Key Words and Phrases: Business agglomeration, causal effect estimation, hypergraph neural networks

ACM Reference format:
Haoran Xin, Xinjiang Lu, Ying Sun, Nengjun Zhu, Tong Xu, Jingbo Zhou, and Hui Xiong. 2026. How Business Agglomeration Affects Individual Points-of-Interest: A Causal Effect Estimation Perspective. *ACM Trans. Knowl. Discov. Data.* 20, 3, Article 46 (March 2026), 28 pages.
https://doi.org/10.1145/3793854

1 Introduction

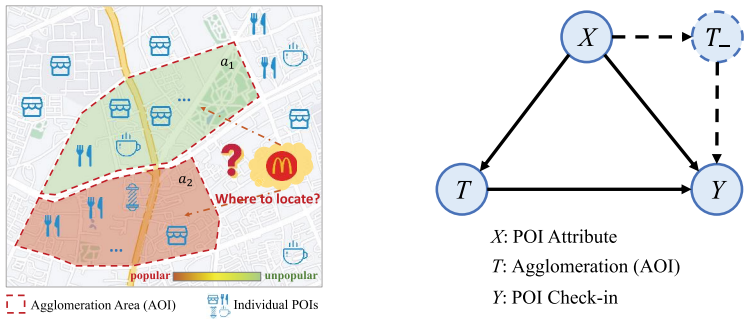
With urbanization advances, agglomeration development has become indispensable in various urban constructions. In particular, as shown in Figure 1(a), there exist many commercial agglomeration areas within the city, such as high streets and shopping centers, consisting of closely located individual businesses. The prosperity of such agglomerations can significantly benefit individual businesses by attracting more customers and creating additional commercial opportunities. Analyzing how agglomeration influences the commercial performance of individual businesses offers valuable insights for urban management and commercial strategy. However, existing research concerning urban agglomeration [20, 32, 44, 64, 65] has rarely studied the specific agglomeration effects on individual businesses.

In this work, we propose to systematically investigate the effect of agglomeration on individual businesses through a quantitative lens. Given that check-in volume serves as a reliable indicator of popularity and represents commercial opportunities, we use it as a valuable surrogate to study the commercial effect. Specifically, as **Point-of-Interest (POI)** and **Area-of-Interest (AOI)** data play important roles in **Location-Based Services (LBS)**, for example, Google Maps and Yelp, we aim to *quantify the contribution of the agglomeration effect to the check-in volume at POIs* in a data-driven way. Here, we present an illustrative example of business site selection to showcase our motivation.

Motivating Example: *Consider the scenario depicted in Figure 1(a) where a new fast-food chain, specifically a McDonald's, is to be established within one of the agglomeration areas (i.e., AOI a_1 or a_2). Suppose our observations reveal that the POIs within a_1 generally experience a low check-in volume, whereas those within a_2 are more popular and attract a larger customer base. Initially, placing the store in a_2 might seem wise due to its generally higher popularity. Yet, it's essential to delve deeper. The lower foot traffic in a_1 could stem from a mismatch between existing businesses and the area's focus, not a lack of customer interest. If the new store's concept resonates more closely with a_1's characteristics, it might benefit from a stronger agglomeration effect, potentially attracting more customers and making a_1 a more strategic choice.*

Therefore, it is important to thoroughly investigate the agglomeration effect on POIs, as it can offer fresh insights into leveraging agglomeration for increased business visibility and customer attraction. This knowledge can significantly aid in commercial decision-making and ultimately support the growth of the urban business economy.

In the literature, as understanding popularity can reveal key commercial insights, many research efforts [19, 31, 60, 63] focus on using spatial–temporal clues and auxiliary POI information to predict



(a) (b)

Fig. 1. (a) An illustration of business agglomeration and the motivating example. (b) The causal graph illustrates the business agglomeration effect on POI check-ins. The check-in outcome (Y) is jointly influenced by the agglomeration level (treatment, T)—which includes interference effects, depicted by dashed lines—and the POI attributes (covariates, X). Additionally, the agglomeration effect on Y is confounded by the POI's attributes. POI, Point-of-Interest.

the POI's popularity and demand. However, these works often cannot predict the specific portion of popularity resulting from agglomeration, due to a lack of detailed exploration into how POIs interact with their agglomeration areas. Meanwhile, several qualitative studies conceptually examined the impact of agglomeration in urban areas, encompassing aspects such as cooperation and competition within agglomerations [48], as well as its impact on consumer choices concerning shopping purposes and destinations [2]. Nevertheless, these studies fall short of providing comprehensive decision-making support due to the inability to uncover the quantitative agglomeration impact on POIs.

Owing to the capability of causal inference theory in exploring the underlying mechanism of the outcome generation process, causal effect estimation emerges as a promising approach for agglomeration effect quantification. Therefore, we employ a causal graph to tackle the counterfactual question: *What would have been the check-in volume of the POI if it had not been influenced by its associated agglomeration area?* Specifically, as shown in Figure 1(b), the check-in behavior of a POI (outcome, Y) is jointly influenced by the agglomeration (treatment, T) and the POI attribute (covariate, X). Moreover, the agglomeration effect on the check-in outcome is confounded by the POI's attributes. However, disentangling the pure causality of the agglomeration effect on the check-in outcome presents genuine challenges. First, agglomerations may have spillover effects, influencing not only the POIs within their territory but also neighboring POIs (the dashed lines in Figure 1(b)). Such interference often exhibits higher-order impacts due to the spatial agglomeration distribution. As a result, solely modeling the direct treatment effect from an agglomeration area to POIs is inadequate. Second, the role of the confounders, that is, the POI attribute, becomes more complicated in geographical settings. This complexity arises because the spatial and functional characteristics of POIs are intertwined to confound the check-in outcome, making it hard to eliminate the blended confounding bias.

To overcome the aforementioned challenges, we propose a **Causal inference framework for AgglomeRation Effect measurement (CARE)**. We majorly equip CARE with two dedicated modules, namely **Spatial Interference Diffusion Network (SIDN)** and **Disentangled Propensity Estimator (DPE)**. Specifically, SIDN aims to propagate spatial interference based on the spatial agglomeration distribution. It starts by aggregating the POIs within each hyperedge to create a hyperedge representation on top of a proposed agglomeration hypergraph. Subsequently,

the treatment representation is integrated into the aggregated hyperedge representation, facilitating spatial-aware propagation. An environment representation, depicting other POIs and their received treatments, is then derived by distributing the shared hyperedge information back to each POI according to the hypergraph connectivity. On the other hand, DPE is designed to model the propensity of POIs to receive treatment while disentangling spatial and inherent factors. By constructing semantic corrupted treatment samples and employing a triple-view negative sampling strategy, propensity representations are learned in both the geographical and inherent subspaces. To achieve a unified causal effect estimation framework, we adopt neural Robinson decomposition, effectively integrating SIDN and DPE. This integration enables the removal of confounding bias, allowing us to obtain the pure contribution of the agglomeration to individual POIs.

The main contributions of this work are as follows:

- To the best of our knowledge, we are among the first to quantitatively study the causal effect of agglomeration areas on individual POIs. We propose CARE, a causal effect estimation framework designed to capture spatial interference and provide disentangled propensity estimation, enabling us to identify the causal effect of agglomerations on POIs.
- We incorporate the spatial spillover effect into the causal graph by modeling higher-order spatial interference with the help of agglomeration hypergraph, multiset aggregation, and treatment-aware spatial propagation.
- To obtain optimal propensity estimation, we unravel the tendency of spatial objects receiving treatments by the feature disentanglement and a triple-view negative sampling with respective corrupted treatments.
- Extensive experiments on three real-world datasets validate the effectiveness and universality of CARE in inferring the agglomeration effect on POIs.

2 Related Work

POI Mining. As LBS rapidly develops, POIs have become the focus of many urban applications. One of the most representative works is the POI recommendation, which aims to provide a list that fits a user's interest and current location. Existing works [28, 29, 57] mainly follow the collaborative filtering paradigm with spatiotemporal awareness and auxiliary information of POIs [52, 54]. In addition, some works have investigated various spatial knowledge to fulfill the POI-related inference. For example, [9, 27] proposed spatial-aware **Graph Neural Networks (GNNs)** to infer the relationships among POIs. multiple social clues on LBS. [19] devised a dynamic graph-based approach to predict the popularity of stores to support site selection based on historical data. [31, 63] modeled the POI demand with human mobility patterns. Nevertheless, while the agglomeration of POIs acts as a universal phenomenon in cities, few studies have discussed the agglomeration effect on individual POIs.

Causal Effect Estimation. It is helpful for decision-making to identify the causal effect. With Random Controlled Trial being impractical, learning the ITE from observational data has attracted lots of research attention, regression model BART to estimate the causal effect. Most existing works assume the **Stable Unit Treatment Value Assumption (SUTVA)**, which ignores potential interference among units. To address this limitation, recent studies explore causal effect estimation in settings where units influence each other through network connections. Two main approaches have emerged: one focuses on designing assignment strategies that allow interference by relaxing SUTVA and explicitly modeling spillover effects [3, 6, 21]; the other investigates observational data where outcomes depend not only on an individual's own treatment and covariates but also on those of their neighbors [16, 17, 22, 34, 36, 37]. In addition, causal effect estimation has also been applied to practical applications such as bike-sharing station construction [53], team competition [61], and

online review assessment [10]. In this work, we resort to causal effect estimation to quantify the agglomeration effect on the performance of individual POIs.

Representation Learning on Hypergraphs. As a generalization of the graph in which an edge can join any number of vertices, the efforts on hyperGNNs mainly fall into two main categories, namely **Clique-Expansion (CE)**-based and tensor-based approaches. The key idea of CE-based methods is to convert hypergraphs to ordinary graphs by replacing hyperedges with cliques, higher-order complex data correlations, hypergraphs based on the spectral theory of hypergraphs, mechanism in the learning process on hypergraphs to nurture flexibility. In parallel, some works [7, 15] designed more sophisticated propagation rules on hypergraphs, which they formulate as the Z eigenproblems for hypergraphs without using CE, merits of both CE and tensor-based methods with the multiset function. In this work, we first carefully design a spatial agglomeration hypergraph based on real-world geographical distribution and capture the spatial interference with the aid of hypergraph learning techniques.

3 Preliminaries

This section will first introduce several basic causal definitions following the potential outcome framework [43]. Then, a formal problem definition is presented.

3.1 Background Causal Concepts

In the traditional potential outcome theory, a treatment refers to an intervention or condition whose causal effect on an outcome is of interest. Treatment t usually falls in binary values, that is, $t \in \{0, 1\}$ (either applied or not). Nevertheless, here we generalize to the multiple treatment setting for clearer elaboration. In other words, suppose we have a treatment set $\mathcal{T} = \{t_1, t_2, \ldots, t_{|\mathcal{T}|}\}$. Then, we define potential outcome as:

Definition 1 (*Potential Outcome*). Given a unit i and a treatment $t \in \mathcal{T}$, the potential outcome of unit i under treatment t is the outcome value that would have been if the treatment of the unit i had been set to t, which is denoted as $Y_i(t, -i, T_{-i}) \in \mathbb{R}$, where $-i$ denotes the units except for i, and T_{-i} refers to the treatments received by all units except for i.

Particularly, in the common SUTVA, the outcome of each unit is not affected by the treatments assigned to other units, that is, $\forall T_{-i}, T'_{-i}, Y_i(t, -i, T_{-i}) = Y_i(t, -i, T'_{-i})$. However, in this work, we will study a problem where interference is present, such that the outcome is also influenced by the treatment assignments of other units, that is, $\forall T_{-i} = T'_{-i}, Y_i(t, -i, T_{-i}) = Y_i(t, -i, T'_{-i})$. In addition, for a unit i, among all the potential outcomes under different treatments, only one outcome, where a specific treatment is applied, can be observed. We call such a potential outcome the *factual potential outcome*. Conversely, the remaining potential outcomes are known as the *counterfactual potential outcomes*.

Based on the definition of the potential outcome, ITE is defined as the difference between the potential outcomes at treatment value t and t' for a unit i:

$$\tau_i(t, t', -i, T_{-i}) = Y_i(t, -i, T_{-i}) - Y_i(t', -i, T_{-i}). \tag{1}$$

3.2 Problem Definition

We start by providing several important data structures related to our problem and then introduce the problem definition studied in this work. Key notations employed throughout this article are summarized in Table 1.

Initially, following many previous works [25, 30, 56] that divide a city into grids for various urban analyses, we also divide the city into grids, each spanning an area of $\Delta d \times \Delta d\ km^2$, to signify

Table 1. Key Notations in the Main Text

Symbol	Definition
$Y_i(t, -i, T_{-i})$	Potential outcome of unit i under treatment t with interference from other units $-i$
$\tau_i(t, t', -i, T_{-i})$	ITE for unit i
$\mathcal{V}, \mathcal{A}, \mathcal{R}$	POI, AOI, and region set
v, a, r	POI, AOI, and region
$\mathcal{T}$	Spatial treatment set
$\mathcal{V}^0, \mathcal{V}^1$	Independent and affiliated POI set
$\widetilde{t}(v_i)$	Factual spatial treatment of POI v_i
$Y_i(t, -v_i, T_{-v_i})$	PCV of POI v_i under treatment t
o_i	Environment summary of POI v_i
$\mathcal{O}$	Space of all the possible environment summaries
$\tau_i(a_j)$	ITE of AOI a_j on POI v_i
$\boldsymbol{p}_v, \boldsymbol{q}_a$	POI and AOI feature vector
$\boldsymbol{x}(v), \boldsymbol{a}, \boldsymbol{r}$	POI, AOI, and region embedding
$\mathcal{H}_s$	Spatial agglomeration hypergraph
$\boldsymbol{e}_i, \boldsymbol{e}(a\|v)$	Hyperedge representation indexing by ID and by the object that it is assigned to
$\mathcal{D}(e)$	Geographical closeness set of hyperedge e
$\psi(e)$	Look up function for coordinates of the object that e is assigned to
$\varsigma(\cdot, \cdot)$	Coordinates distance function
$\boldsymbol{\omega}, \boldsymbol{\omega}^g, \boldsymbol{\omega}^h$	Complete, geo, and inherent propensity representation
$\mathcal{S}^g, \mathcal{S}^h, \mathcal{S}^r$	Geo, inherent, and random treatment negative sampler
m	Conditional mean outcome

the region. We denote the set of regions as $\mathcal{R} = \{r_i\}_{i=1}^{|\mathcal{R}|}$, with the center coordinates of each r_i denoted as ℓ_i^r. Besides, the POI and AOI are defined as follows.

Definition 2 (*POI*). A POI is a spatial item allocated to a geographical location. Formally, we use v_i to represent a POI, $\widetilde{r}(v_i)$ to denote its located region, ℓ_i^v to represent its coordinates, $\boldsymbol{p}_i$ to denote its intrinsic attributes, and y_i to indicate its check-in volume during a time period Δt.[1] Thus, each POI v_i can be represented by a four-tuple $v_i = (\widetilde{r}(v_i), \ell_i^v, \boldsymbol{p}_i, y_i)$.

Definition 3 (*AOI*). An AOI refers to an agglomeration area in LBS denoted by a_i, which can be deemed a special POI with a fixed bounding area. Formally, for an AOI a_i, its center coordinates, area, intrinsic attributes, and geographical boundary are denoted by ℓ_i^a, s_i, $\boldsymbol{q}_i$, and Γ_i, respectively. Each AOI can be identified by a tuple $a_i = \left(\ell_i^a, s_i, \boldsymbol{q}_i, \Gamma_i\right)$.

Note that since AOIs are often clearly defined and easily searchable by users in modern LBS, we treat each AOI as unique, with its own fixed identification, location, and boundaries in this work. Based on the AOIs within a city, POIs can be classified into two categories: affiliated POIs that belong to one of the AOIs, and independent POIs that do not belong to any AOIs. Formally, given a set of POIs $\mathcal{V} = \{v_i\}_{i=1}^{|\mathcal{V}|}$ and a set of AOIs $\mathcal{A} = \{a_i\}_{i=1}^{|\mathcal{A}|}$, $\mathcal{V}$ can be divided into two disjoint subsets

[1]If a person has had direct interaction with the POI in the real world, we call such interaction a check-in of the POI. As our problem is studied within a fixed period, we omit Δt for brevity.

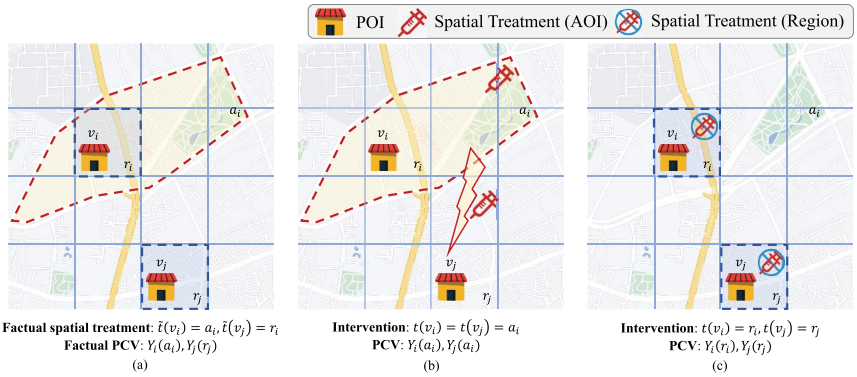


Fig. 2. The demonstration of the spatial treatment and PCV where the region assignments are obtained from dividing the city into grids, and the AOIs are well identified by modern LBS. (a) The factual spatial treatment of POIs v_i and v_j are given by a_i and r_j, respectively. This means that v_i is located within the AOI a_i. On the other hand, v_j is not located within any AOI, so we use a geographical indicator to represent its spatial treatment. (b) The spatial treatment a_i is implemented to both v_i and v_j. (c) The spatial treatment r_j is implemented to both v_i and v_j.

according to $\mathcal{A}$:

$$\mathcal{V}^1 = \left\{v_i | v_i \in \mathcal{V}, \left(\exists! a_j \in \mathcal{A}, \ell_i^v \in \Gamma_j\right)\right\},$$
$$\mathcal{V}^0 = \left\{v_i | v_i \in \mathcal{V}, \left(\forall a_j \in \mathcal{A}, \ell_i^v \notin \Gamma_j\right)\right\},$$

where $\mathcal{V}^0, \mathcal{V}^1$ are the independent and affiliated POI set, respectively, $\mathcal{V}^0 \cup \mathcal{V}^1 = \mathcal{V}$ and $\mathcal{V}^0 \cap \mathcal{V}^1 = \emptyset$. Then, we introduce the definitions of spatial treatment and **Potential Check-In Volume (PCV)** as follows.

Definition 4 (*Spatial Treatment*). The spatial treatment set $\mathcal{T}$ is composed of the region set $\mathcal{R}$ and the AOI set $\mathcal{A}$ (i.e., $\mathcal{T} = \mathcal{R} \cup \mathcal{A}$). Then, given a POI v, if it is assigned to the spatial treatment $a \in \mathcal{A}$, it means that v receives the direct agglomeration effect from the studied AOI a. Conversely, if v is given the spatial treatment $r \in \mathcal{R}$, it means that v locates in r and receives no direct agglomeration effect from any AOIs $\forall a \in \mathcal{A}$.

Definition 5 (*PCV*). Given a POI v, the potential check-in volume is the check-in volume that would have been with $v$receiving (not receiving) the direct agglomeration effect from a given AOI (any AOIs).

The illustration of the spatial treatment and PCV is shown in Figure 2. Note that the direct agglomeration effect here means the effect from the studied AOI. As stated before, there exists interference in our studied problem, which contributes to the indirect agglomeration effect and will be discussed later.

For $\forall v_i \in \mathcal{V}^1$, we use $\widetilde{a}(v_i)$ to denote the AOI that v_i belongs to. Then for $\forall v_i \in \mathcal{V}$, we determine its factual treatment (i.e., located AOI or region) $\widetilde{t}(v_i) \in \mathcal{T}$ based on whether v_i is an affiliated or independent POI:

$$\widetilde{t}(v_i) = \begin{cases} \widetilde{r}(v_i), & \text{if } v_i \in \mathcal{V}^0, \\ \widetilde{a}(v_i), & \text{if } v_i \in \mathcal{V}^1. \end{cases} \tag{2}$$

Afterward, we propose quantifying PCV with the potential outcome.

Proposition 1 (PCV Quantification). *Given a POI v_i, its located region $\widetilde{r}(v_i)$, and an AOI a_j, when v_i receives the direct effect from a_j, PCV equals $Y_i(a_j, -v_i, T_{-v_i})$. While v_i does not receive the direct effect from any AOIs, PCV equals $Y_i(\widetilde{r}(v_i), -v_i, T_{-v_i})$.*

First, we can naturally derive $Y_i(a_j, -v_i, T_{-v_i})$ based on the definition of PCV and spatial treatment $a \in \mathcal{A}$. Second, since $\widetilde{r}(v_i)$ is seen as a spatial treatment that describes the factual geographical context when no direct agglomeration effect is present, we can utilize it to derive $Y_i(\widetilde{r}(v_i), -v_i, T_{-v_i})$. Proposition 1 provides the groundwork for studying the problem within the potential outcome framework.

Due to the existence of spatial interference, inspired from [34], we further suppose that there exists a summary function Ψ, which can cover all the environmental knowledge and yield $o_i \in \mathcal{O}$ for each POI $v_i \in \mathcal{V}$ (i.e., other POIs and their received treatment), where $\mathcal{O}$ denotes the space of all the possible environment summaries:

$$o_i = \Psi(-v_i, T_{-v_i}). \tag{3}$$

In this article, we would like to study the effect of the AOI on the POI's check-in volume, which can be cast as the ITE estimation. Specifically, given a POI $v_i \in \mathcal{V}$ and an AOI $a_j \in \mathcal{A}$, the ITE $\tau_i(a_j)$ is given by:

$$\tau_i(a_j) = \mathbb{E}\left[Y_i(a_j, o_i)|\boldsymbol{p}_i\right] - \mathbb{E}\left[Y_i(\widetilde{r}(v_i), o_i)|\boldsymbol{p}_i\right].$$

Problem 1 (ITE Estimation of AOIs on POIs). *Given a POI $v_i \in \mathcal{V}$ ignorant about its factual check-in volume and an AOI $a_j \in \mathcal{A}$, predict the ITE $\tau_i(a_j) \in \mathbb{R}$ with the learned ITE estimator.*

For a single POI, since we can observe only one of its PCVs (with or without receiving the explicit effect from AOIs), following [34], we make an unconfoundedness assumption under interference to ensure the ITE is identifiable.

Assumption 1 (Unconfoundedness). *The unconfoundedness assumes the PCVs of a POI v_i are independent with the treatment and summary of the environment given its attribute, i.e., $Y_i(t, o) \perp t', o'|\boldsymbol{p}_i, \forall t, t' \in \mathcal{T}, \forall o, o' \in \mathcal{O}$.*

Note that since the summary function in Equation (3) is assumed to sufficiently capture interference effects, there is no need to condition further on the covariates of other POIs. The interference variance is already encapsulated in the summary representation, which is reflected in o'.

4 Methodology

Figure 3 presents the architecture of our proposed approach (CARE). It mainly consists of two modules, that is, SIDN and DPE, which are incorporated into a unified causal effect estimation framework with the help of neural Robinson decomposition. In particular, (1) SIDN diffuses interference in forms of the indirect treatment effect over the designed spatial agglomeration hypergraph; and (2) DPE unravels the original propensity estimation process into the geographical and inherent aspects through a triple-view negative sampling. Afterward, with the conditional mean outcome, the neural Robinson decomposition is employed to partial out confounding effects and further identify the distinct contributions of an AOI on a POI's check-in volume.

4.1 Spatial Treatment Representation Learning

We first embed the AOIs and regions into latent variables. Specifically, for an AOI $a_i \in \mathcal{A}$, we examine its attributes and employ a **Multilayer Perceptron (MLP)** to transform these attributes

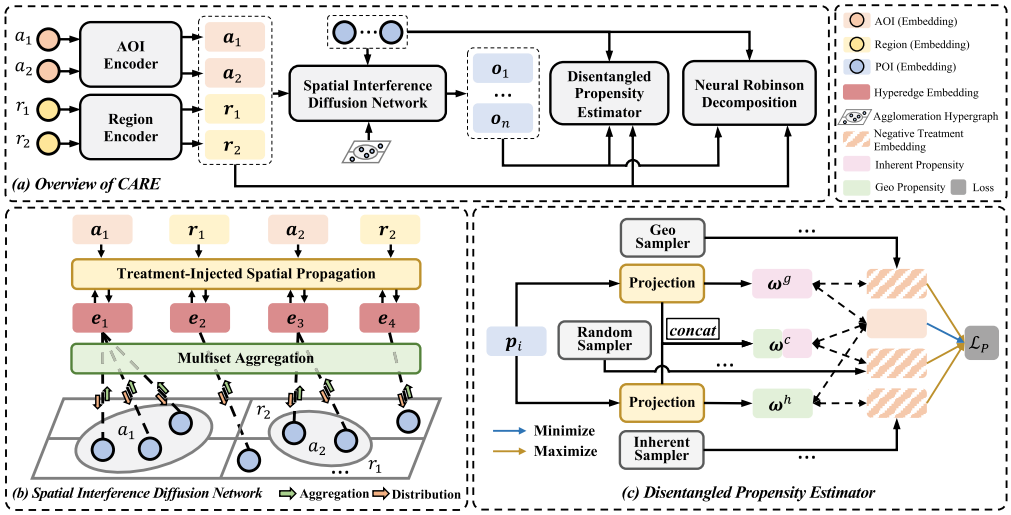


Fig. 3. The overview of CARE framework and its two major components, that is, SIDN and DPE.

into meaningful representations [63] $\boldsymbol{a}_i \in \mathbb{R}^d$:

$$\boldsymbol{a}_i = \text{MLP}_a(\boldsymbol{q}_i), \tag{4}$$

where $\boldsymbol{q}_i$ is the feature of a_i, and d denotes the dimensionality. Concerning the regions, as we expect them to act as geographical indicators that are indifferent to the functional semantics, we randomly assign an initial embedding $\boldsymbol{r}_i \in \mathbb{R}^d$ for each $r_i \in \mathcal{R}$. Afterward, based on the definition of the spatial treatment (Definition 4), we derive the spatial treatment embedding $\boldsymbol{t}$ by incorporating the AOI embedding and the region embedding as:

$$\boldsymbol{t}_i = \begin{cases} \boldsymbol{a}_i, & \text{if } t_i \in \mathcal{A}, \\ \boldsymbol{r}_i, & \text{if } t_i \in \mathcal{R}, \end{cases}$$

4.2 SIDN

In real-world scenarios, except for the direct effect from the studied AOI, there commonly exist indirect effect from other nearby potential AOIs, and even POIs, that is, the interference. For example, when a person pays a visit to an AOI, say, a shopping center, she or he may also visit surrounding POIs that are not located within the shopping center, which means that an AOI could affect surrounding POIs implicitly even if the AOI does not have a direct impact on these POIs. The diffusion of such interference among the spatial objects heavily depends on geographical distribution. Therefore, to deal with the spatial interference, we first define a spatial agglomeration hypergraph according to the agglomeration and independence of POIs in the real world.

Definition 6 (*Spatial Agglomeration Hypergraph*). A spatial agglomeration hypergraph $\mathcal{H}_s$ characterizes the agglomeration and independence of POIs. $\mathcal{H}_s$ can be denoted by $\mathcal{H}_s = \{\mathcal{V}_s, \mathcal{E}_s\}$, where $\mathcal{V}_s = \mathcal{V}$ is the node set and $\mathcal{E}_s$ is the hyperedge set that suggests hyper-connectivity among nodes. Each hyperedge $e \in \mathcal{E}_s$ can be seen as a subset of $\mathcal{V}_s$, i.e., $e \subseteq \mathcal{V}_s$. Each AOI and independent POI is assigned a hyperedge, denoted by $e(a_i)$ and $e(v_i)$, respectively, where $a_i \in \mathcal{A}, v_i \in \mathcal{V}^0$. The connectivity in $\mathcal{H}_s$ can be depicted by an incidence matrix $\boldsymbol{H} \in \mathbb{R}^{|\mathcal{V}_s| \times |\mathcal{E}_s|}$, and the entries are

defined as:

$$h(v, e(a_i)) = \begin{cases} 1, & \text{if } v \in \mathcal{V}_s^1 \wedge \ell_v \in \Gamma_i, \\ 0, & \text{otherwise}, \end{cases} \quad a_i \in \mathcal{A},$$

$$h(v, e(v_i)) = \begin{cases} 1, & \text{if } v \in \mathcal{V}_s^0 \wedge v = v_i, \\ 0, & \text{otherwise}, \end{cases} \quad v_i \in \mathcal{V}^0,$$

where $\mathcal{V}_s^0 = \mathcal{V}^0$, $\mathcal{V}_s^1 = \mathcal{V}^1$.

The presence of aforementioned spatial interference breaks the SUTVA that the outcome of each unit is not affected by the treatments assigned to other units [16, 22]. Several works have put efforts into relaxing the SUTVA [3, 34, 47, 66], and [34] is the most related work to ours which utilized the standard HyperGNN to model the interference among treatments and units. Nevertheless, due to the limitation of the binary treatment setting and rigid message-passing mechanism in standard HyperGNN, [34] fails to capture the relationships between the POIs and the agglomeration effectively. Bearing these in mind, we propose a novel SIDN, which can diffuse the indirect treatment effect based on the designed spatial agglomeration hypergraph. There are three key components in SIDN, termed *Multiset Aggregation*, *Treatment-Injected Spatial Propagation*, and *Interference Distribution*.

4.2.1 Multiset Aggregation ($\boldsymbol{x} \rightarrow \boldsymbol{e}$). First, for $\forall v_i \in \mathcal{V}$, similarly with Equation (4), the POI embedding $\boldsymbol{x}_i \in \mathbb{R}^d$ is initialized with an MLP encoder taking the POI attribute $\boldsymbol{p}_i$ as input:

$$\boldsymbol{x}_i = \text{MLP}_p(\boldsymbol{p}_i). \tag{5}$$

Inspired by [11, 24], as the message-passing process on hypergraphs can be formulated with multiset function,[2] and further enhanced by powerful function kernels, we utilize a multiset function f_m to exploit the interaction among the nodes within the same hyperedge and aggregate the information to obtain the hyperedge representation. The kernel of f_m is implemented with Transformer [50].

More specifically, given a hyperedge $e_i \in \mathcal{E}_s$, the nodes connected by e_i are aggregated with multi-head attention:

$$\begin{aligned} \boldsymbol{e}_i &= f_m\left(\boldsymbol{X}_{e_i}\right) = \sigma\left(\text{LayerNorm}\left(\boldsymbol{h}_i + \text{MLP}_e\left(\boldsymbol{h}_i\right)\right)\right), \\ \boldsymbol{h}_i &= \text{LayerNorm}\left(\|_{k=1}^{K} \boldsymbol{z}_M^{(k)} + \|_{k=1}^{K} \boldsymbol{o}_i^{(k)}\right), \\ \boldsymbol{o}_i^{(k)} &= \boldsymbol{N}_i^{(k)} \cdot \text{Softmax}\left(\left(\boldsymbol{M}_i^{(k)}\right)^{\mathrm{T}} \boldsymbol{z}_M^{(k)}\right), \\ \boldsymbol{M}_i^{(k)} &= \boldsymbol{W}_M^{(k)^{\mathrm{T}}} \boldsymbol{X}_{e_i} + \boldsymbol{b}_M^{(k)}, \quad \boldsymbol{N}_i^{(k)} = \boldsymbol{W}_N^{(k)^{\mathrm{T}}} \boldsymbol{X}_{e_i} + \boldsymbol{b}_N^{(k)}, \end{aligned} \tag{6}$$

where $\boldsymbol{X}_{e_i} = \underset{v_j \in e_i}{\|} \boldsymbol{x}_j \in \mathbb{R}^{d \times |e_i|}$, and $\|$ denotes a column-wise concatenation operation. $\boldsymbol{W}_M^{(k)}, \boldsymbol{W}_N^{(k)} \in \mathbb{R}^{d \times \frac{d}{K}}$ and $\boldsymbol{b}_M^{(k)}, \boldsymbol{b}_N^{(k)}, \boldsymbol{z}_M^{(k)} \in \mathbb{R}^{\frac{d}{K}}$ are learnable parameters. The aggregated representation $\boldsymbol{e}_i$ can be regarded as the hyperedge representation of e_i.

4.2.2 Treatment-Injected Spatial Propagation ($e \rightarrow e'$). The diffusion of spatial interference depends on geographical closeness. For example, when people check in an AOI/POI, they are unlikely to visit other AOIs/POIs located far away from it in a short period [49]. However, typical MPNN cannot fully handle the diffusion as mentioned earlier on the spatial agglomeration hypergraph

[2] A function is permutation invariant if the output is regardless of the order of inputs. Then, if a function is permutation invariant, we call it a multiset function.

due to the isolation among hyperedges. To remedy this, we inject the treatments into the acquired hyperedge representations and adaptively propagate them with spatial awareness:

$$\begin{aligned} \boldsymbol{u}_{e(a_i)} &= \mathrm{MLP}_u\left(\left[\boldsymbol{e}(a_i) \parallel \boldsymbol{a}_i\right]\right), a_i \in \mathcal{A} \\ \boldsymbol{u}_{e(v_i)} &= \mathrm{MLP}_u\left(\left[\boldsymbol{e}(v_i) \parallel \widetilde{\boldsymbol{r}}(v_i)\right]\right), v_i \in \mathcal{V}^0, \end{aligned} \tag{7}$$

where $\widetilde{\boldsymbol{r}}(v_i)$ is the region embedding of $\widetilde{r}(v_i)$. Afterward, we propagate the merged representation with the help of attention mechanism [51]:

$$\begin{aligned} \boldsymbol{e}_i' &= \sigma\left(\sum_{e_j \in \mathcal{D}(e_i)} \pi_{ij} \boldsymbol{W}_P \boldsymbol{u}_{e_j}\right), \\ \pi_{ij} = \frac{\exp(\mu_{ij})}{\sum_{e_k \in \mathcal{D}(e_i)} \exp\left(\mu_{ik}\right)}, &\quad \mu_{ij} = \sigma\left(\boldsymbol{z}_P^{\mathrm{T}} \boldsymbol{W}_P \left[\boldsymbol{u}_{e_i} \parallel \boldsymbol{u}_{e_j}\right]\right), \end{aligned} \tag{8}$$

where $\boldsymbol{z}_P \in \mathbb{R}^d$ and $\boldsymbol{W}_P \in \mathbb{R}^{d \times 2d}$ are trainable parameters. $\mathcal{D}\left(e_i\right)$ is the geographical closeness set w.r.t e_i, which is defined as:

$$\begin{aligned} \mathcal{D}(e_i) &= \left\{e_j | \varsigma\left(\psi(e_i), \psi(e_j)\right) < \rho^s, \forall e_i, e_j \in \mathcal{E}_s\right\}, \\ \psi\left(e(a_i)\right) &= \ell_i^a, a_i \in \mathcal{A}, \quad \psi\left(e(v_i)\right) = \ell_i^v, v_i \in \mathcal{V}^0, \end{aligned} \tag{9}$$

where $\varsigma(\cdot, \cdot)$ calculates the distance between two coordinates and ρ^s is a distance threshold for the propagation. Once the treatment effect and the hyperedge summary have been diffused through the spatial propagation pipe, the updated representation $\boldsymbol{e}_i'$ of e_i is ready to be distributed back to each node.

4.2.3 Interference Distribution ($e' \rightarrow o$). After the aggregation and propagation of node v_i, we can acquire its environment representation $\boldsymbol{o}_i$ by absorbing the representation from its hyperedge:

$$\boldsymbol{o}_i = \boldsymbol{W}_D \boldsymbol{e}_j', v_i \in e_j, \tag{10}$$

where $\boldsymbol{W}_D \in \mathbb{R}^{d \times d}$ is a learnable parameter. The environment representation $\boldsymbol{o}$ can selectively convey the environmental information of other POIs along with the treatments they have received.

Note that multiple SIDN layers can be stacked to capture higher-order spatial interference while only injecting the treatment representation in the first layer.

4.3 DPE

In causal studies, propensity score [41] generally refers to the conditional probability of a unit being exposed to a particular treatment given a set of observed covariates, which is widely used to alleviate the selection bias when estimating the causal effect from observational data [1, 22, 35, 41, 42]. However, when estimating the propensity score of spatial objects (e.g., POIs and AOIs), the situation gets more complicated. First, the original propensity score is inadequate in fully capturing the exposure due to the presence of spatial interference. Second, the propensity of a POI exposed to a specific treatment depends on geographical proximity. For example, a POI is more likely to be exposed to an AOI treatment if the POI is located within the boundary of the AOI. Moreover, different business characteristics of the AOI may have different impacts on the propensity of exposing the POI. Unfortunately, if we directly model the conditional probability $P(t|\boldsymbol{p})$ $(t \in \mathcal{T})$, not only the spatial interference but the entangled spatial and functional effects may lead to sub-optimal estimation of the propensity score and thus hinder the final ITE inference objective. To this end, we devise a DPE to unravel the estimation process into the *geographical aspect* and *inherent aspect*.

4.3.1 Feature Disentanglement. Specifically, to incorporate the spatial interference into the propensity estimation, we first introduce the propensity score under interference $P(t, o|\boldsymbol{p})$ [14]. Following [22], given a POI $v_i \in \mathcal{V}$, a treatment $\forall t \in \mathcal{T}$, and an environment summary $\forall o \in \mathcal{O}$, we then extend the propensity score into the propensity representation $\boldsymbol{\omega}(v_i)$:

$$\boldsymbol{\omega}(v_i) \triangleq \mathbb{E}\left[h(\boldsymbol{t}, \boldsymbol{o})|\boldsymbol{p}_i\right], \tag{11}$$

where $h(\cdot)$ is an arbitrary function responsible for blending the two types of representations, that is, the treatment and the environment representation. We implement $h(\cdot)$ with a linear transformation of the concatenation of the representations because of its simple yet effective implementation as:

$$h(\boldsymbol{t}, \boldsymbol{o}) = \boldsymbol{W}_T^{\mathrm{T}}\left[\boldsymbol{t} \parallel \boldsymbol{o}\right] + \boldsymbol{b}_T, \tag{12}$$

where $\boldsymbol{W}_T \in \mathbb{R}^{2d\times d}, \boldsymbol{b}_T \in \mathbb{R}^{2d}$ are parameters. As aforementioned, the direct approximation of propensity representation is entangled by geographical and inherent factors. To activate disentanglement, we use two separate MLPs to project the initial POI attribute $\boldsymbol{p}_i$ of v_i into geo and inherent spaces and obtain the propensity representations in these two spaces, respectively:

$$\boldsymbol{\omega}^g(v_i) = \mathrm{MLP}_g(\boldsymbol{p}_i), \quad \boldsymbol{\omega}^h(v_i) = \mathrm{MLP}_h(\boldsymbol{p}_i), \tag{13}$$

where $\boldsymbol{\omega}^g, \boldsymbol{\omega}^h \in \mathbb{R}^d$ are geo and inherent propensity representations, respectively. We also combine representations of these two subspaces to generate a complete propensity representation:

$$\boldsymbol{\omega}(v_i) = \boldsymbol{W}_C^{\mathrm{T}}\left[\boldsymbol{\omega}^g(v_i) \parallel \boldsymbol{\omega}^h(v_i)\right] + \boldsymbol{b}_C, \tag{14}$$

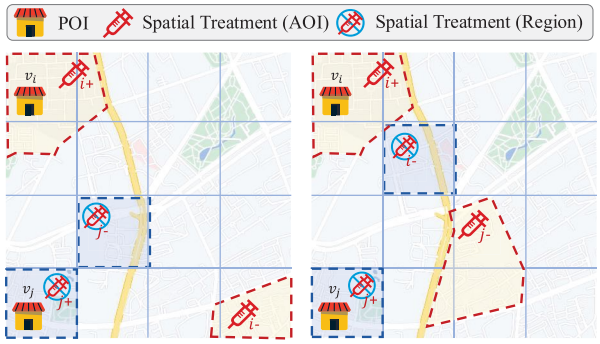
where $\boldsymbol{\omega} \in \mathbb{R}^d$ and $\boldsymbol{W}_C \in \mathbb{R}^{2d\times d}, \boldsymbol{b}_C \in \mathbb{R}^d$ are learnable parameters.

4.3.2 Triple-View Negative Sampling. At the POI level, the propensity to receive agglomeration treatment is confounded along two structurally distinct dimensions: (1) geography, which drives exposure to spatial interference, and (2) intrinsic POI attributes, which govern baseline attractiveness. These axes are often entangled in observational data, making it difficult to disentangle whether a POIs treatment likelihood arises from its location or inherent appeal. Negative sampling is a widely used training strategy and has been applied in various fields, such as recommendation [38, 39, 67], knowledge graph completion [55], and self-supervised learning [8, 26]. As negative sampling can efficiently address the complex geographical and inherent characteristics of the propensity by making informative contrasts, we propose a triple-view negative sampling strategy to achieve representation disentanglement, where three sets of representative negative treatment samples are constructed.

In particular, we build three samplers: geo sampler, inherent sampler, and random sampler, and adopt **Bayesian personalized ranking (BPR)** criterion [39] to learn the respective propensity representations. Given a POI v_i, since $\widetilde{t}(v_i)$ (Equation (2)) represents its factual treatment, we take $\widetilde{t}(v_i)$ together with o_i as the positive sample of v_i.

For negative sampling, we first build the geo sampler based on the following two empirical observations as illustrated in Figure 4(a):

- For an affiliated POI, that is, $\forall v \in \mathcal{V}^1$, it is less likely to receive direct treatment effect from those AOIs far away.
- For an independent POI, that is, $\forall v \in \mathcal{V}^0$, it is less likely to receive direct treatments of the regions $r \in \mathcal{R} \setminus \{\widetilde{r}(v)\}$ as v does not locate in these regions.



(a) Geo sampler (b) Inherent sampler

Fig. 4. Illustration of negative spatial treatments generated by the two proposed samplers: the geo sampler and the inherent sampler. In the notation of spatial treatments, subscripts i/j denote the target POI, and superscripts $+/-$ indicate positive and negative treatments, respectively.

Besides, the negative environment summary is also sampled corresponding to each sampled treatment. Therefore, the geo sampler is given by:

$$\begin{aligned}\mathcal{S}^g = &\left\{(v_i, t_j, o_k) | v_i \in \mathcal{V}^1, t_j \in \mathcal{A}, \varsigma\left(\ell_i^v, \ell_j^a\right) > \rho^a, \widetilde{t}(v_k) = t_j\right\} \\ &\cup \left\{(v_i, t_j, o_k) | v_i \in \mathcal{V}^0, t_j \in \mathcal{R} \setminus \{\widetilde{r}(v_i)\}, \widetilde{t}(v_k) = t_j\right\},\end{aligned} \tag{15}$$

where ρ^a is a distance threshold. For the inherent sampler, we instruct its sampling process inspired by the following observational insights as illustrated in Figure 4(b):

- —The affiliated POI $v \in \mathcal{V}^1$ is most likely to receive direct treatment effect from the AOI it belongs to instead of its nearby regions because it lies in a functional area with semantics, that is, the AOI.
- —The independent POI $v \in \mathcal{V}^0$ is less likely to be affected by direct treatment effect of nearby AOIs because there are no functional areas covering it.

In other words, we use counterpart types of treatments to reinforce the functional properties of AOIs. Consequently, with the environment summary similarly sampled, the negative samples generated by the inherent sampler are:

$$\begin{aligned}\mathcal{S}^h = &\left\{(v_i, t_j, o_k) | v_i \in \mathcal{V}^1, t_j \in \mathcal{R}, \varsigma\left(\ell_i^v, \ell_j^r\right) < \rho^a, \widetilde{t}(v_k) = t_j\right\} \\ &\cup \left\{(v_i, t_j, o_k) | v_i \in \mathcal{V}^0, t_j \in \mathcal{A}, \varsigma\left(\ell_i^v, \ell_j^a\right) < \rho^r, \widetilde{t}(v_k) = t_j\right\},\end{aligned} \tag{16}$$

where ρ^r is another distance threshold. Besides, to ensure the learning process is not overly biased toward the geo and inherent negative samples, we utilize an extra random sampler to discover the distribution of comprehensive negative facts:

$$\mathcal{S}^r = \left\{(v_i, t_j, o_k) | v_i \in \mathcal{V}, t_j \in \mathcal{T} \setminus \left\{\widetilde{t}(v_i)\right\}, \widetilde{t}(v_k) = t_j\right\}. \tag{17}$$

4.3.3 Learning Objectives. Afterward, given a POI v, a treatment $t \in \mathcal{T}$, and an environment summary $o \in \mathcal{O}$, we define the approximation score based on the representations as follows:

$$\eta(\boldsymbol{\omega}, v, t, o) \triangleq \|\boldsymbol{\omega}(v) - h(\boldsymbol{t}, \boldsymbol{o})\|_2^2. \tag{18}$$

Consequently, with the pairwise comparisons between negative and positive samples under the BPR criterion, we derive the following loss functions:

$$\begin{aligned}
\mathcal{L}_P^g &= \sum_{(v_i,t_j,o_k)\in\mathcal{S}^g} \left[\delta + \eta(\boldsymbol{\omega}^g, v_i, \widetilde{t}(v_i), o_i) - \eta(\boldsymbol{\omega}^g, v_i, t_j, o_k)\right]_+, \\
\mathcal{L}_P^h &= \sum_{(v_i,t_j,o_k)\in\mathcal{S}^h} \left[\delta + \eta(\boldsymbol{\omega}^h, v_i, \widetilde{t}(v_i), o_i) - \eta(\boldsymbol{\omega}^g, v_i, t_j, o_k)\right]_+, \\
\mathcal{L}_P^r &= \sum_{(v_i,t_j,o_k)\in\mathcal{S}^r} \left[\delta + \eta(\boldsymbol{\omega}, v_i, \widetilde{t}(v_i), o_i) - \eta(\boldsymbol{\omega}, v_i, t_j, o_k)\right]_+,
\end{aligned} \tag{19}$$

where $[\cdot]_+ = \max(0, \cdot)$ denotes the standard hinge loss, and δ is the margin size. Then, we derive the propensity loss by integrating the above three loss functions:

$$\mathcal{L}_P = \mathcal{L}_P^r + \mathcal{L}_P^g + \mathcal{L}_P^h. \tag{20}$$

4.4 Neural Robinson Decomposition.

Robinson decomposition was initially proposed for parametric components estimation in partially linear models [40]. Recently, the Robinson decomposition has been applied in causal effect estimation to partial out the confounding effect [4, 22, 35] such that the separate contributions of treatments and covariates to the outcome can be identified. The core idea is to use propensity to decouple the outcome into two parts: the conditional mean outcome, which is solely a function of the covariates, and the treatment effect, which captures the additional impact introduced by the treatment. These two components are modeled separately in a two-stage training process, allowing us to estimate causal effects effectively. In this section, we aim to incorporate the proposed SIDN and DPE into a unified causal effect estimation framework with the aid of neural Robinson decomposition [22].

We first define the conditional mean outcome of a POI $v_i \in \mathcal{V}$, which captures the expected outcome associated with the POI attribute:

$$m(v_i) \triangleq \mathbb{E}\left[y|\boldsymbol{p}_i\right]. \tag{21}$$

Besides, we employ another MLP to encode the POI attribute into low-dimensional representation $\boldsymbol{v}_i$ for Robinson Decomposition:

$$\boldsymbol{v}_i = \mathrm{MLP}_v(\boldsymbol{p}_i). \tag{22}$$

According to the product effect assumption [22], the effect of a treatment on a unit can be modeled by the inner product of their representations. We extend this assumption to deal with the causal effect under interference as:

Assumption 2 (Product Effect under Interference). *The potential outcome $Y_i(t_j, o_i)$ of POI v_i with treatment t_j under spatial interference can be inferred as:*

$$Y_i(t_j, o_i) = \boldsymbol{v}_i^{\mathrm{T}} h(\boldsymbol{t}_j, \boldsymbol{o}_i) + \varepsilon, \tag{23}$$

where $\boldsymbol{o}_i$ is the environment representation of v_i obtained from SIDN. Then, $\forall(v, t, o) \in \mathcal{V} \times \mathcal{T} \times \boldsymbol{O}, \mathbb{E}\left[\varepsilon|\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{o}\right] = \mathbb{E}\left[\varepsilon|\boldsymbol{x}\right] = 0$.

With Assumption 2, given a POI v_i, the conditional mean outcome $m(v_i)$ in Equation (21) can be further inferred by:

$$m(v_i) = \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{\omega}(v_i), \tag{24}$$

where $\boldsymbol{\omega}$ is the complete propensity representation (refer to Equation (14)). Afterward, the neural Robinson decomposition of Equation (23) is derived as:

$$Y_i(t_j, o_i) - m(v_i) = \boldsymbol{v}_i^{\mathrm{T}} \left(h(\boldsymbol{t}_j, \boldsymbol{o}_i) - \boldsymbol{\omega}(v_i) \right) + \varepsilon. \tag{25}$$

4.5 Optimization and Inference

We discuss the learning of model parameters on top of neural Robinson decomposition. With Equation (25), given the conditional mean out m and complete propensity representation $\boldsymbol{\omega}$, we have the following overall loss function:

$$\mathcal{L} = \sum_{v_i \in \mathcal{V}} \left\{ y_i - \left[m(v_i) + \boldsymbol{v}_i^{\mathrm{T}} \left(h\left(\widetilde{\boldsymbol{t}}(v_i), \boldsymbol{o}_i\right) - \boldsymbol{\omega}(v_i) \right) \right] \right\}^2, \tag{26}$$

where y_i is the factual check-in volume of POI v_i (i.e., $y_i = Y_i(\widetilde{t}(v_i), o_i)$) and $\widetilde{\boldsymbol{t}}(v_i)$ is the treatment representation of $\widetilde{t}(v_i)$. Since in Equation (26), the overall loss is coupled to the conditional mean outcome m and propensity $\boldsymbol{\omega}$, we learn the parameters in a two-stage manner.

Stage 1. As the conditional mean outcome defined in Equation (21) only depends on the POI attribute (covariates), a sole MLP is employed to learn a basic regression function:

$$m(v_i) = \mathrm{MLP}_o(\boldsymbol{p}_i), \tag{27}$$

where the parameters are optimized with the following objective:

$$\mathcal{L}_O = \sum_{v_i \in \mathcal{V}} \| m(v_i) - y_i \|_2^2 . \tag{28}$$

Stage 2. Although with Equation (11), the propensity representation is only conditioned on the POI attribute. It is also bundled with the treatment representation, environment representation, and their interactions in the representation learning process (see Equation (20)). Therefore, we alternatively optimize the propensity loss (Equation (20) together with the overall loss (Equation (26)). Specifically, the alternative optimization in stage 2 until convergence is performed as follows:

- —Optimize $\mathcal{L}$ in Equation (26) for N steps while the parameters in conditional mean outcome m and propensity $\boldsymbol{\omega}$ are fixed.
- —Optimize $\mathcal{L}_P$ in Equation (20) for one step by only updating the parameters in propensity $\boldsymbol{\omega}$.

The summary of training CARE can be found in Algorithm 1. Afterward, we can use the trained CARE to estimate the ITE w.r.t. an AOI impacting a POI's check-in volume. Specifically, given a POI v_i and an AOI a_j, the ITE is inferred as:

$$\hat{\tau}_i(a_j) = \boldsymbol{v}_i^{\mathrm{T}} \left(h(\boldsymbol{a}_j, \boldsymbol{o}_i) - h(\widetilde{\boldsymbol{r}}(v_i), \boldsymbol{o}_i) \right), \tag{29}$$

where $\widetilde{\boldsymbol{r}}(v_i)$ is the region embedding of $\widetilde{r}(v_i)$.

Revisiting the motivating example in Section 1, consider a siting decision such as the McDonald's case. The ITE estimated by CARE can support more informed choices. Specifically, given two AOIs and a POI, one can compute the ITE using Equation (29) and then select the site with the higher ITE—subject to practical constraints.

5 Experiments

In this section, we conduct experiments on two real-world datasets to evaluate the effectiveness of the proposed CARE framework in estimating the agglomeration effect on POIs.

Algorithm 1: Training Procedure of CARE

Input : POI set $\mathcal{V}$, AOI set $\mathcal{A}$, region set $\mathcal{R}$, spatial agglomeration hypergraph $\mathcal{H}_s$
Output: The agglomeration ITE estimator $\mathcal{F}$

Initialize parameters: $\Theta_T = \left\{\text{MLP}_a, \{\boldsymbol{r}_i\}_{i=1}^{|\mathcal{R}|}, \boldsymbol{W}_T, \boldsymbol{b}_T\right\}$,
$\Theta_{SIDN} = \left\{\text{MLP}_{\{p,e,u\}}, \boldsymbol{W}_{\{M,N,P,D\}}, \boldsymbol{b}_{\{M,N\}}, \boldsymbol{z}_{\{M,P\}}\right\}$, $\Theta_{DPE} = \left\{\text{MLP}_{\{g,h\}}, \boldsymbol{W}_C, \boldsymbol{b}_C\right\}$,
$\Theta_{CMO} = \{\text{MLP}_o\}$, $\Theta_X = \{\text{MLP}_v\}$;
while *not converged* **do**
  Get conditional mean outcome $m(v)$ by Equation (27);
  Get loss $\mathcal{L}_O$ by Equation (28);
  Update parameters Θ_{CMO} according to the gradient of $\mathcal{L}_O$;
end
while *not converged* **do**
  /* *Spatial Interference Diffusion Network* */
  Get AOI representations $\boldsymbol{a}_i$ by Equation (4);
  Get POI representations $\boldsymbol{x}_i$ by Equation (5);
  Get hyperedge representation $\boldsymbol{e}_i$ by Equation (6);
  Get updated hyperedge representation $\boldsymbol{e}'_i$ by Equations (7) and (8);
  Get environment representation $\boldsymbol{o}_i$ by Equation (10);
  /* *Disentangled Propensity Estimator* */
  Get geo, inherent and complete propensity representations $\boldsymbol{\omega}^g(v), \boldsymbol{\omega}^h(v), \boldsymbol{\omega}(v)$ by Equations (13) and (14);
  Get negative sample sets $\mathcal{S}^g, \mathcal{S}^h, \mathcal{S}^r$ by Equations (15) to (17);
  Get propensity loss $\mathcal{L}_P$ by Equations (18) to (20);
  Get loss $\mathcal{L}$ by Equation (26);
  /* *Optimization Stage 1* */
  for $n = 1$ *to* N **do**
    Update parameters $\Theta_{\{T,SIDN,X\}}$ according to the gradient of $\mathcal{L}$;
  end
  /* *Optimization Stage 2* */
  Update parameters Θ_{DPE} according to the gradient of $\mathcal{L}_P$;
end
return $\mathcal{F}_{\Theta_{\{T,SIDN,X\}}}$

5.1 Experimental Setups

5.1.1 Dataset. We construct the datasets based on two real-world online map application datasets (*Shanghai* and *Chengdu*), where users' check-ins and queries on both POI and AOI data were recorded, and one public POI check-in dataset (*NYC* [59]). For the *Shanghai* and *Chengdu* datasets, we utilized the overall POI check-ins initiated by the users from 04 January 2019 to 04 July 2019 as the factual check-in volume of each POI. To assure the data quality, we filtered out the POIs where the check-in volume is less than ten or greater than 50,000. We also eliminated the AOIs with less than three affiliated POIs. Then, we incorporate the category, the coordinates, the located administrative region, and the existence duration to construct POI and AOI features. For the *NYC* dataset, we adopted the POI check-ins from 12 April 2012 and 16 February 2013. Similarly, POIs with check-ins fewer than 5 or more than 1,000 times were filtered out. Given the unavailability of

Table 2. Statistics of Our Datasets

Statistics		Values		
		Shanghai	Chengdu	NYC
# POIs	Affiliated	93,207	70,179	6,927
	Independent	72,808	57,938	3,061
# AOIs		8,996	5,980	200
Average check-in volume		80.61	82.58	17.85
Average # POIs in an AOI		10.36	11.74	34.64

AOI information in this dataset, a density peak clustering algorithm was employed as an adaptive measure to form synthetic agglomeration areas. Within these clusters, core and border points were identified as affiliated POIs, while noise points outside of the clusters were considered as independent POIs. To maintain diversity in areas of high density, a random selection of border and core points were reclassified as noise points. POI features were constructed using the category, coordinates, and ID, while, due to the absence of AOI information, only the ID was utilized for constructing AOI features. Moreover, for these three datasets, we randomly split the POIs into a training set, a validation set, and a test set following the proportions: 60%/20%/20%. And we evaluate the ITE of POIs in the test set. The statistics of datasets are given in Table 2.

Because of the lack of ground truth of counterfactual outcomes, it is hard to obtain the true causal effect of agglomeration from observational data. Therefore, we design a simulation strategy to generate the counterfactual check-in volume based on observational data. Although this approach follows standard practice [22, 34, 46] in causal inference, we acknowledge its limitations: performance is measured against simulated outcomes shaped by our modeling assumptions, not against real-world causal effects. To reduce overfitting to one design, we vary key parameters (introduced later) and run experiments to test robustness across different simulation environments. These checks strengthen confidence in our findings within the semi-synthetic setting. Note that the generated ITE is solely utilized for evaluation purposes. Conversely, our model is trained on real-world observations, guaranteeing its practical feasibility. For the affiliated POI $\forall v_i \in \mathcal{V}^1$ where $Y_i(\widetilde{a}(v_i), o_i) = y_i$ is the factual check-in volume, we construct its counterfactual check-in volume $Y_i(\widetilde{r}(v_i))$ as follows:

$$Y_i(\widetilde{r}(v_i)) = y_i \left(1 - \kappa \cdot g_f(v_i, \widetilde{a}(v_i)) \cdot g_d(v_i, \widetilde{a}(v_i))\right) + \xi, \tag{30}$$

where κ is a hyperparameter, ξ is a random noise from Gaussian distribution $\mathcal{N}(0, 1)$, and g_f and g_d are the function- and distance-aware decay factors, respectively, which will be introduced later. For the independent POI $\forall v_i \in \mathcal{V}^0$, since $Y_i(\widetilde{r}(v_i))$ serves as its factual check-in volume, we sample several AOIs next to v_i to evaluate how the nearby AOIs could affect v_i if it receives the explicit effect from these AOIs. Given a sampled AOI $a \in \mathcal{A}$ that is near the independent POI $v_j \in \mathcal{V}^0$, we calculate the counterfactual volume $Y_j(a)$ similarly with Equation (30):

$$Y_j(a) = y_j \left(1 + \kappa \cdot g_f(v_j, a) \cdot g_d(v_j, a)\right) + \xi. \tag{31}$$

Generally, the design intuition behind the two decay factors, g_f and g_d, stems from real-world observations. Initially, a distance-aware decay factor g_d is instantiated using a piecewise radial basis function. This function simulates the diminishing effects of increasing geographical distance as:

$$g_d(v_i, a_j) = \begin{cases} 1, & \text{if } \ell_i^v \in \Gamma_j^a, \\ e^{-\lambda_S(\ell_i^v, \ell_j^a)} & \text{otherwise}, \end{cases} \tag{32}$$

where λ is a hyperparameter. Afterward, since the impact of an AOI on a POI primarily hinges on their level of match, we employ the cosine similarity of the features in function-aware decay factor g_f to measure the functional affinity of a POI–AOI pair. This is scaled by the query volume of the AOI, reflecting its influence as follows:

$$g_f(v_i, a_j) = \text{Norm}\left(\log Q(a_i)\right)\left(\frac{1}{2} + \frac{\boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{q}_j}{2\|\boldsymbol{p}_i\|\|\boldsymbol{q}_j\|}\right), \tag{33}$$

where $Q(a)$ is the query volume of AOI a within the dataset time span, and Norm denotes a normalization operation. Specifically, for the dataset that lacks AOI information (*i.e.,* the NYC dataset), following [34], we adopt a linear Gaussian model to determine the factor g_f:

$$g_f(v_i, a_j) = \boldsymbol{\chi}_j^{\mathrm{T}}\boldsymbol{p}_i, \tag{34}$$

where $\boldsymbol{\chi}_j$ is the Gaussian vector for a_j, with each element adhering to the standard normal distribution.

5.1.2 Baselines. We compare the proposed CARE approach with the following baselines.

—Two vanilla methods:
- –*Mean* uses the average check-in volume of the units under a specific treatment as the outcome prediction of a POI given this treatment.
- –**Ordinary least squares (OLS)** linear regression takes the concatenation of the POI feature and the treatment feature (AOI feature or region one-hot vector) as input to fit a linear regression model.

—Two POI prediction methods:
- –*DeepR* [27] is originally proposed for POI competitiveness inference. It splits the neighbors of a POI into several sectors according to their location and aggregates information from each sector.
- –*PRIM* [9] is another POI relationship inference model that integrates the POI category taxonomy and utilizes a self-attentive extractor to distill spatial context.

We equip these two POI prediction methods with **Hilbert-Schmidt Independence Criterion (HSIC)** regularization [17] to enable counterfactual ability.

—Six causal effect estimation methods:
- –**Counterfactual Regression (*CFR)*** [46] balances the representations between treatment and control units to achieve counterfactual inference. We use Wasserstein distance as the **Integral Probability Metric (IPM)** to penalize imbalance. We extend CFR to the multi-treatment setting as [62] suggested.
- –*HyperSCI* [34] models higher-order interference powered by the hyperGNNs. The IPM choice and the multi-treatment extension are similar to CFR.
- –*SIN* [22] employs GNNs to encode the structured treatments and proposes the generalized Robinson Decomposition to isolate the causal components. As there are no structured treatments in our work, we replace the GNNs with MLPs.
- –*BART* [18] predicts the outcomes with an ensemble of multiple regression trees.
- –*GraphITE* [17] also utilizes GNNs to vectorize structured treatments into embeddings and uses HSIC regularization to enhance the independence of the targets and the treatments. Similarly, we replaced the GNNs with MLPs due to the absence of structured treatment data.
- –*HyperSCI(HSIC)* is an upgraded version of HyperSCI, where we replace Wasserstein distance with HSIC regularization and embed the treatments to support the multi-treatment setting.

5.1.3 Implementation Details. We implemented CARE using PyTorch 1.12.1 and PyTorch Geometric [13] 2.2.0. We trained the model on a CentOS 6.3 Linux machine with 24 cores, 251 GB of RAM, and NVIDIA P40 GPU with 22 GB of GPU memory. Then, we performed a random parameter search on the validation set regarding the latent dimensionality in $\{50, 100, 200\}$, # MLP and SIDN layers in $\{1, 2, 3\}$, # attention heads in $\{1, 2, 4\}$, # steps of optimizing $\mathcal{L}$ in $\{3, 5, 10, 15\}$, dropout in $\{0, 0.1, 0.3, 0.5, 0.7\}$, learning rate in $\{0.001, 0.005, 0.01\}$, and # negative samples in $\{1, 3, 5\}$. Besides, we set the distance thresholds ρ^s, ρ^a as 3 km and ρ^r as 8 km. We choose Adam as the optimizer to train CARE using an L2 regularization with weight 10^{-5}. In the experiments, the decay coefficient λ in g_d (Equation (32)) is set to 0.4. Given an independent POI $v \in \mathcal{V}^0$, we pick 3 closest AOIs to v as its counterfactual treatments. Regarding the hyper-parameter configurations of the baselines, for PRIM, the number of GNN layers is set to 1, and the coefficient of the radial basis function kernel is fixed at 2. For DeepR, the spatial context is divided into four sectors, and the grid size is set to 0.1 km. In the case of CFR, the coefficient of the IPM is set to 1. For HyperSCI, the IPM coefficient is set to 0.1 on the Shanghai and NYC datasets, and 10 on the Chengdu dataset to account for its denser spatial structure. For SIN, the hyper-parameters are selected via random search following the same strategy as CARE. For BART, the number of regression trees is set to 100. Finally, for GraphITE and HyperSCI(HSIC), the coefficient of the HSIC regularization term is set to 10.

5.1.4 Evaluation Metrics. In the literature, there are two widely used evaluation metrics for treatment effect estimation, namely *Rooted Precision in Estimation of Heterogeneous Effect* ($\sqrt{\epsilon_{PEHE}}$) [18] and *Mean Absolute Error on average treatment effect* (ϵ_{ATE}) [46]. As our problem setting can be seen as a multi-treatment one, following [33, 45], we extend the two original metrics into the multi-treatment variants $\overline{\sqrt{\epsilon_{PEHE}}}$ and $\overline{\epsilon_{ATE}}$ as:

$$\overline{\sqrt{\epsilon_{PEHE}}} = \frac{1}{|\mathcal{A}|}\sum_{j=1}^{|\mathcal{A}|}\sqrt{\frac{1}{n_j}\sum_{i=1}^{n_j}\left(\hat{\tau}_{v_i}(a_j) - \tau_{v_i}(a_j)\right)^2},$$

$$\overline{\epsilon_{ATE}} = \frac{1}{|\mathcal{A}|}\sum_{j=1}^{|\mathcal{A}|}\left|\frac{1}{n_j}\sum_{i=1}^{n_j}\hat{\tau}_{v_i}(a_j) - \frac{1}{n_j}\sum_{i=1}^{n_j}\tau_{v_i}(a_j)\right|,$$

where n_j denotes the number of POIs to be evaluated that receive the explicit effect from AOI a_j. Such POIs include affiliated POIs that belong to a_j and independent POIs that have the sampled nearby AOI as a_j (see Equation (31)). The lower the metrics, the better the performance.

5.2 Overall Performance Comparison

We conduct the experiments on Shanghai and Chengdu datasets by varying the hyperparameter κ, which controls the significance of the effect from AOIs in the counterfactual outcome simulation, to validate the universality of CARE in various simulation environments. The overall performance results of CARE compared to the baselines in three different simulation settings on Shanghai and Chengdu datasets are presented in Table 3, where the mean and standard deviation of the two metrics are obtained through five random runs using different initial parameters. As can be seen, CARE consistently achieves the best performance w.r.t. all metrics and all κ settings on both datasets. Particularly, CARE outperforms the most competitive baseline in terms of $\overline{\sqrt{\epsilon_{PEHE}}}$, achieving an average improvement of 5.87% and 11.45% on Shanghai and Chengdu datasets, respectively, across different κ values. In addition, the average improvements of $\overline{\epsilon_{ATE}}$ across different κ are 15.78% and 14.68% on Shanghai and Chengdu datasets, respectively, surpassing the best baselines. Moreover, we have the following observations.

Table 3. Performance Comparison between CARE and All the Baselines under Different κ Settings

Datasets	Methods	$\kappa = 0.3$		$\kappa = 0.5$		$\kappa = 0.7$	
		$\overline{\sqrt{\epsilon_{PEHE}}}$	$\overline{\epsilon_{ATE}}$	$\overline{\sqrt{\epsilon_{PEHE}}}$	$\overline{\epsilon_{ATE}}$	$\overline{\sqrt{\epsilon_{PEHE}}}$	$\overline{\epsilon_{ATE}}$
Shanghai	Mean	79.81	75.77	82.16	77.26	84.70	78.82
	OLS	87.31	68.65	89.13	69.14	91.14	69.77
	PRIM	28.39±12.2	24.55±12.0	33.30±12.0	27.34±12.0	38.57±11.8	30.43±12.0
	DeepR	13.40±3.87	9.58±4.02	18.12±3.32	11.76±3.46	23.37±2.84	14.56±2.93
	CFR	50.27±0.15	43.85±0.15	52.76±0.18	45.43±0.17	55.44±0.18	47.08±0.18
	HyperSCI	36.53±0.11	30.84±0.10	38.44±0.11	31.62±0.09	40.64±0.10	32.58±0.08
	SIN	18.03±6.05	13.36±5.35	23.31±5.87	16.55±5.32	28.96±5.69	20.06±5.23
	BART	14.56±1.11	9.30±0.52	17.77±1.10	11.40±0.51	21.08±1.09	13.51±0.50
	GraphITE	12.71±5.08	9.30±5.53	18.67±5.10	13.09±5.69	24.80±5.11	17.05±5.80
	HyperSCI(HSIC)	10.10±0.42	6.39±0.53	15.99±0.55	9.98±0.74	22.11±0.66	13.87±0.89
	CARE	**9.31±0.49**	**5.23±0.48**	**14.76±0.22**	**8.27±0.22**	**20.64±0.19**	**11.88±0.38**
Chengdu	Mean	68.84	64.47	71.28	65.97	73.92	67.55
	OLS	63.93	50.24	65.83	50.88	68.01	51.64
	PRIM	20.39±7.20	16.37±7.18	25.11±7.43	18.73±7.77	30.36±7.50	21.64±8.07
	DeepR	12.24±2.91	8.23±3.11	16.90±2.31	10.20±2.45	22.20±1.82	12.96±1.87
	CFR	41.76±0.12	36.08±0.18	44.66±0.13	37.98±0.18	47.72±0.13	39.92±0.18
	HyperSCI	46.41±0.13	40.54±0.16	48.39±0.13	41.44±0.16	50.62±0.12	42.43±0.14
	SIN	21.31±7.29	13.82±5.26	26.18±7.11	16.45±5.37	31.56±6.94	19.59±5.40
	BART	15.01±0.57	9.41±0.20	18.37±0.52	11.63±0.21	21.87±0.51	13.86±0.21
	GraphITE	11.96±1.00	8.18±0.90	17.41±0.65	11.20±0.78	23.24±0.73	14.68±1.14
	HyperSCI(HSIC)	11.18±1.57	6.49±0.90	16.39±1.63	9.19±1.08	22.05±1.59	12.59±1.21
	CARE	**9.10±0.33**	**4.85±0.29**	**14.67±0.09**	**8.02±0.18**	**20.72±0.19**	**11.83±0.41**
NYC	Mean	10.06	8.56	10.38	8.69	12.65	10.06
	OLS	7.06	6.11	7.31	6.09	9.06	6.36
	PRIM	3.59±0.13	2.55±0.12	4.21±0.11	2.96±0.09	7.92±0.05	5.82±0.08
	DeepR	3.55±0.11	2.58±0.12	4.20±0.11	3.03±0.13	8.01±0.11	5.97±0.13
	CFR	6.02±0.29	4.66±0.28	6.30±0.22	4.69±0.21	9.07±0.23	6.29±0.19
	HyperSCI	8.06±0.03	7.12±0.07	8.19±0.10	7.01±0.09	9.27±0.10	6.56±0.17
	SIN	3.99±0.39	2.86±0.39	4.51±0.35	3.15±0.35	7.97±0.28	5.72±0.27
	BART	3.58±0.13	2.67±0.09	4.15±0.08	3.07±0.12	8.03±0.16	6.07±0.10
	GraphITE	3.81±0.21	2.70±0.24	4.39±0.21	3.05±0.25	7.98±0.26	5.78±0.30
	HyperSCI(HSIC)	3.69±0.09	2.52±0.07	4.31±0.09	2.97±0.08	7.99±0.06	5.82±0.03
	CARE	**3.33±0.08**	**2.31±0.11**	**3.96±0.09**	**2.75±0.11**	**7.74±0.10**	**5.68±0.13**

The best results are highlighted in bold and the second best are underlined.

Among all the baselines, the naive baselines (Mean and OLS) perform the worst in all conditions, demonstrating that simply using the average volume level or a polynomial regression model is not enough to depict the ITE distribution. The POI prediction methods (PRIM and DeepR) achieve relatively better performance, and even attain some second-best performances on NYC dataset, with the spatial awareness of POIs captured by the devised GNNs. However, as each hyperedge is expanded into a fully connected graph, the expression power of the geographical agglomeration is weakened. Besides, the implicit interaction between the treatments and the POIs is poorly modeled. Within the causal effect estimation methods, CFR and the original version of HyperSCI show unsatisfying performance despite the counterfactual inference ability. The main reason is that they

are originally designed for the binary treatment setting, and could become overwhelmed when encountering so many treatments, as demonstrated in our problem. By modeling the treatments as representations, baselines like SIN, BART, and GraphITE can deal with the multi-treatment setting and thus perform relatively more promising. Nevertheless, without incorporating geographical information, which is beneficial to spatial prediction, the performance of these methods is not as good as CARE. As an upgraded version of HyperSCI, the power of HyperSCI(HSIC) in handling multiple treatments is unleashed with HSIC regularization. Owing to the ability to learn the spatial agglomeration and spillover effects through standard HyperGNN, it achieves competitive performance among the baselines. However, it is not qualified to cope with the isolated hyperedges on the spatial agglomeration hypergraph and the entangled confounding bias. In CARE, we design SIDN to smoothly diffuse the spatial interference based on the agglomeration hypergraph and DPE to accomplish the disentanglement of geographical and inherent propensities. Therefore, the performance of CARE is greatly improved on all metrics and all datasets with satisfactory margins.

5.3 Ablation Study

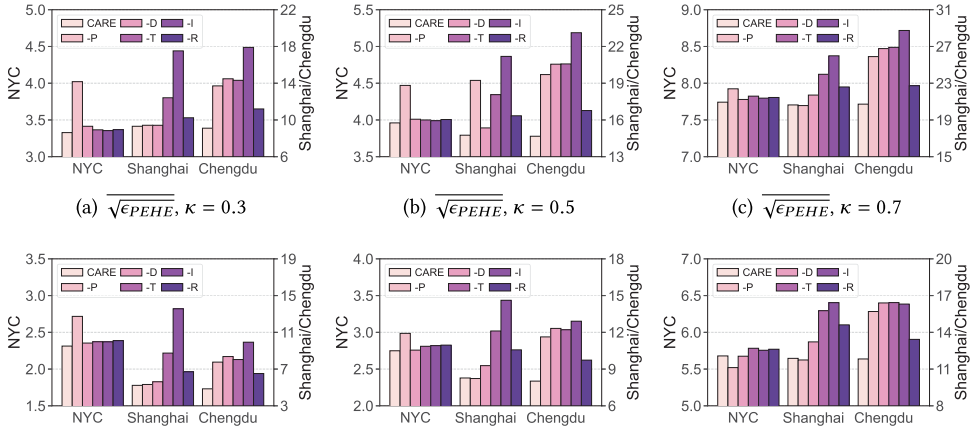
To evaluate the effectiveness of different modules, we compare CARE with its different variants: (1) *CARE-T* removes the treatment injection and only spatially propagates the hyperedge representations, so the spatial diffusion of interference is eliminated. (2) *CARE-P* removes the spatial propagation process while keeping the treatment injection. As a result, there will be no information sharing among hyperedges. (3) *CARE-I* removes the environment representation learned by SIDN, such that the spatial interference based on the agglomeration is neglected. (4) *CARE-D* removes the disentanglement in the propensity estimator, which only uses a random negative sampling strategy to learn the propensity representation. (5) *CARE-R* removes the random negative sampler from the propensity estimator, only retaining two disentangled samplers.

As can be seen from Figure 5, CARE shows enhanced performance over a majority of the variants, proving the critical roles of our designs in estimating the ITE of AOIs on POIs. Specifically, the performance degrades if we remove the treatment injection (CARE-T), which indicates that the indirect treatment effect does deserve careful management. Besides, CARE outperforms CARE-P in most cases, showing that explicitly modeling geographical distance among AOIs and POIs can boost the diffusion of spatial interference. Also, the performance experiences a decrease when the environment representation, which captures spatial interference, is not included (CARE-I), providing further evidence of the advantages of incorporating the agglomeration spillover effects for predicting the ITE. In addition, removing the disentanglement learning of propensity estimator (CARE-D) also results in performance decline, which verifies that the propensity entanglement of spatiality and functionality could restrict the ITE estimation of spatial objects. Finally, eliminating the random negative sampler in DPE (CARE-R) leads to the deterioration in performance, underscoring the essential role of such a design in regularizing disentangled propensity learning.

5.4 Parameter Sensitivity Analysis

Figure 6 presents the parameter sensitivity analysis of CARE on two datasets.

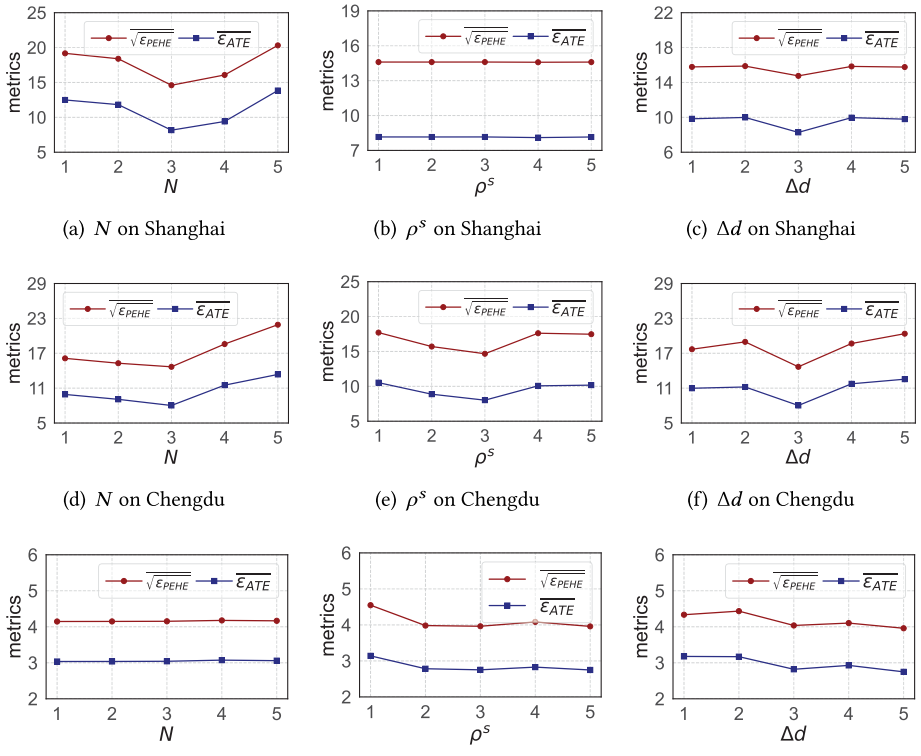
The Number of Steps N for Optimizing $\mathcal{L}$. We analyze how the number of steps for optimizing $\mathcal{L}$ affects the performance by varying N from 1 to 5. As shown in Figure 6(a), (d), and (g), utilizing a too small value for N leads to an increase in errors, while a relatively large value for N results in poorer performance. When N grows, the optimization frequency of the propensity estimator decreases during the alternative optimization process. Too frequent optimization of propensity



(a) $\overline{\sqrt{\epsilon_{PEHE}}}$, $\kappa = 0.3$ (b) $\overline{\sqrt{\epsilon_{PEHE}}}$, $\kappa = 0.5$ (c) $\overline{\sqrt{\epsilon_{PEHE}}}$, $\kappa = 0.7$

(d) $\overline{\epsilon_{ATE}}$, $\kappa = 0.3$ (e) $\overline{\epsilon_{ATE}}$, $\kappa = 0.5$ (f) $\overline{\epsilon_{ATE}}$, $\kappa = 0.7$

Fig. 5. Contributions of different modules in CARE.



(a) N on Shanghai (b) ρ^s on Shanghai (c) Δd on Shanghai

(d) N on Chengdu (e) ρ^s on Chengdu (f) Δd on Chengdu

(g) N on NYC (h) ρ^s on NYC (i) Δd on NYC

Fig. 6. Parameter sensitivity analysis on three datasets ($\kappa = 0.5$).

could disturb the primary optimization objective, while the opposite could lead to poor elimination of the confounding bias. Therefore, it is beneficial to find a tradeoff of the value for N.

The Distance Threshold ρ^s for the Spatial Propagation. To further probe the impact of the distance threshold ρ^s in the treatment-injected spatial propagation of SIDN, we vary ρ^s from $1km$ to $5km$ on

Table 4. Contribution of Two Training Stages ($\kappa = 0.5$)

Training stage	Shanghai		Chengdu		NYC	
	$\sqrt{\epsilon_{PEHE}}$	ϵ_{ATE}	$\sqrt{\epsilon_{PEHE}}$	ϵ_{ATE}	$\sqrt{\epsilon_{PEHE}}$	ϵ_{ATE}
Only Stage 1	16.14±0.03	10.60±0.04	16.51±0.04	10.94±0.05	4.16±0.03	3.06±0.02
Only Stage 2	16.98±0.17	10.16±0.20	17.84±0.65	9.75±0.34	4.07±0.15	2.90±0.17
Stage 1 and 2	**14.76±0.22**	**8.27±0.22**	**14.67±0.09**	**8.02±0.18**	**3.96±0.09**	**2.75±0.11**

The best results are highlighted in bold and the second best are underlined.

both datasets. As illustrated in Figure 6(b), (e), and (h), the performance remains stable on Shanghai and NYC dataset while fluctuating over a reasonable range on Chengdu dataset. This variability on Chengdu dataset may be attributed to the unique spatial distribution of its aggregation areas, where business circles are predominantly aligned along the city's urban axis. Besides, we can also observe that a large distance threshold (e.g., 5 km) generally produces slightly worse performance on the Chengdu dataset. This is because a large threshold could weaken the discriminative ability of the spatial propagation as more and more members are involved in the closeness set (Equation (9)). Although a larger threshold does not adversely affect performance on Shanghai and NYC datasets, it is still not advisable as it may lead to computational inefficiency by overextending the closeness set. Thus, a proper ρ^s could bring performance and efficiency improvements.

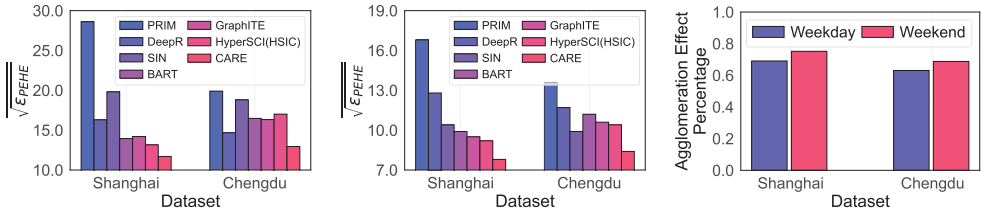
The Grid Size Δd. We also investigate the influence of the size of the grid that determines the division of urban regions by varying Δd from 1 km to 5 km. Figure 6(c), (f), and (i) show that the performance generally reaches its peak at a grid size of 3 km × 3 km. On the one hand, with the increase in grid size, the receptive field regarding geography within the control group expands, as neighboring independent POIs could receive the same treatment. On the other hand, larger granularity is not always good, as it may jeopardize the spatial discrimination of treatments concerning the control group.

5.5 Contribution of Two Training Stages

To further validate the effectiveness of the two-stage training strategy, we compare the performance of CARE against its single-stage variants, each omitting one of the two stages described in Section 4.5. The results are presented in Table 4. As shown, the two-stage training consistently outperforms its single-stage counterparts, confirming its advantage in disentangling confounding effects. Specifically, removing Stage 1 prevents the model from properly estimating the conditional mean outcome, leading to inferior performance. Conversely, omitting Stage 2 results in entanglement between treatment and outcome, which undermines the accuracy of ITE estimation.

5.6 Temporal Dynamics Analysis

We further investigate the fine-grained temporal dynamics captured by CARE on the Shanghai and Chengdu datasets. Specifically, we divide the data into weekdays and weekends, and compare the performance of CARE against several competitive baselines. The results are presented in Figure 7(a) and (b). As illustrated, CARE consistently outperforms the baselines across both datasets, demonstrating its ability to accurately estimate treatment effects under varying temporal conditions, thereby confirming its robustness and generalization capability. In addition, we compute the proportion of the agglomeration effect on POIs relative to the total check-in volume for weekdays



(a) Performance of weekday ($\kappa = 0.5$) (b) Performance of weekend ($\kappa = 0.5$) (c) Agglomeration effect proportion

Fig. 7. Detailed performance analysis on weekdays and weekends for the Shanghai and Chengdu datasets.

Table 5. Efficiency Comparison across the Three Datasets

Methods	Training time (s)			Inference time (s)		
	Shanghai	Chengdu	NYC	Shanghai	Chengdu	NYC
DeepR	96.284	65.526	6.343	1.261	1.064	0.095
GraphITE	99.449	77.607	22.301	0.336	0.339	0.018
HyperSCI(HSIC)	94.021	36.559	9.143	0.418	0.416	0.048
PRIM	66.032	47.359	7.381	0.923	0.925	0.058
SIN	1.094/0.515	0.845/0.442	0.068/0.197	0.080	0.077	0.021
CARE	0.339/5.999	0.275/0.660	0.030/1.052	0.434	0.148	0.052

Training times for different stages are reported separately if multiple stages are required.

and weekends, as shown in Figure 7(c). The results reveal that the agglomeration effect is more pronounced during weekends, which aligns with the intuitive observation that agglomeration areas tend to attract more visitors on weekends, when people are more inclined to engage in leisure activities.

5.7 Efficiency Comparison

To assess efficiency, we present the results of our proposed CARE framework and comparable baselines in terms of the training time and inference time in Table 5, standing for the offline training efficiency and deployed ITE estimation efficiency, respectively. The training time is measured as the average duration of one epoch, while the inference time refers to the processing time to obtain estimated ITE results with the inference batch size of 1.

It is evident that DeepR, GraphITE, HyperSCI(HSIC), and PRIM exhibit lower efficiency in terms of time consumption due to the computational requirements of the HSIC regularization term, which consumes significant time and cannot be accelerated by GPU. SIN, on the other hand, is the most efficient baseline, as it removes the computationally intensive HSIC regularization component and adopts a lighter two-stage training strategy to mitigate confounding factors. However, it comes at the expense of sub-optimal performance in estimating the ITE. As for our CARE framework, it takes slightly longer times than SIN due to the incorporation of spatial interference modeling and the negative sampling process. However, CARE achieves the best overall performance, making this modest increase in computational cost justifiable. This is especially notable when contrasted with baselines such as HyperSCI(HSIC), which exhibit significantly higher computational overhead yet deliver subpar performance.

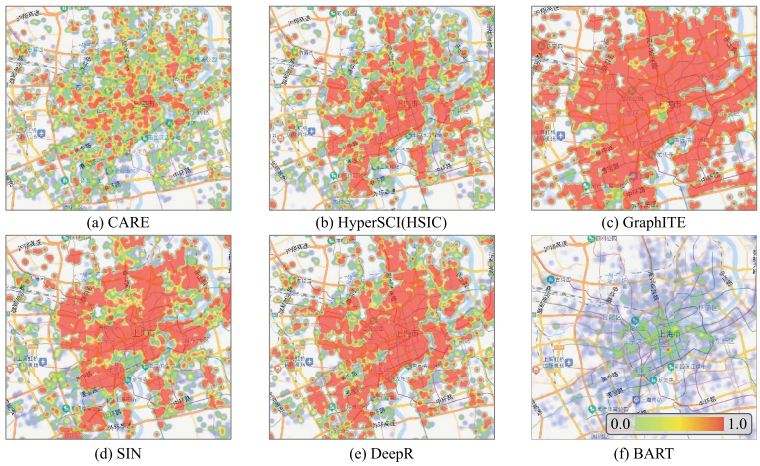


Fig. 8. The visualization of the ITE proportion inference by different methods.

5.8 Visualization Analysis

To intuitively demonstrate the superiority of CARE in inferring the agglomeration effect on POIs, we visualize the ITE proportion predicted by CARE and some competitive baselines for each affiliated POI in the Shanghai test set. The results are shown in Figure 8 and deeper color means the agglomeration area has a more significant effect on the POI's check-in volume. As can be observed, CARE has distinguished several agglomerations that have a potentially high effect, while the baseline seems to predict the ITEs all over the city evenly.

6 Conclusion and Future Work

In this article, we studied the business agglomeration effect by quantifying the contribution of the agglomeration effect to the check-in volume at POIs. We formulate the problem as a causal effect estimation task and introduce a causal framework, namely CARE. This framework is designed to partial out the entangled confounding variables and accurately identify the agglomeration effect while accounting for the interference that is spatially diffused among POIs. Along this line, we devised a SIDN to capture spatial interference by spreading the treatment effect on a dedicated agglomeration hypergraph. Then, the DPE was proposed to unravel the propensity into geographical and inherent aspects by a triple-view negative sampling strategy. After that, the agglomeration effect was estimated on top of the neural Robinson decomposition. Extensive experiments on three real-world datasets demonstrated the effectiveness and the universality of the proposed approach. In future work, we aim to incorporate more complex factors that may influence POI check-in volumes, such as weather conditions, traffic patterns, and population density, to gain a deeper understanding of check-in behavior.

References

[1] Ahmed M. Alaa, Michael Weisz, and Mihaela Van Der Schaar. 2017. Deep counterfactual networks with propensity-dropout. arXiv:1706.05966. Retrieved from https://arxiv.org/abs/1706.05966

[2] Theo A. Arentze, Harmen Oppewal, and Harry J. P. Timmermans. 2005. A multipurpose shopping trip model to assess retail agglomeration effects. *Journal of Marketing Research* 42, 1 (2005), 109–115.

[3] Peter M. Aronow and Cyrus Samii. 2017. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*11, 4 (2017), 1912–1947.

[4] Susan Athey, Julie Tibshirani, and Stefan Wager. 2019. Generalized random forests. *The Annals of Statistics* 47, 2 (2019), 1148–1178.
[5] Song Bai, Feihu Zhang, and Philip H. S. Torr. 2021. Hypergraph convolution and hypergraph attention. *Pattern Recognition* 110 (2021), 107637.
[6] Guillaume Basse and Avi Feller. 2018. Analyzing two-stage experiments in the presence of interference. *Journal of the American Statistical Association* 113, 521 (2018), 41–55.
[7] Austin R. Benson, David F. Gleich, and Lek-Heng Lim. 2017. The spacey random walk: A stochastic process for higher-order data. *SIAM Review* 59, 2 (2017), 321–345.
[8] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *Proceedings of the International Conference on Machine Learning*. PMLR, 1597–1607.
[9] Yile Chen, Xiucheng Li, Gao Cong, Cheng Long, Zhifeng Bao, Shang Liu, Wanli Gu, and Fuzheng Zhang. 2022. Points-of-interest relationship inference with spatial-enriched graph neural networks. arXiv:2202.13686. Retrieved from https://arxiv.org/abs/2202.13686
[10] Lu Cheng, Ruocheng Guo, and Huan Liu. 2022. Estimating causal effects of multi-aspect online reviews with multi-modal proxies. In *Proceedings of 15th ACM International Conference on Web Search and Data Mining*, 103–112.
[11] Eli Chien, Chao Pan, Jianhao Peng, and Olgica Milenkovic. 2021. You are AllSet: A multiset function framework for hypergraph neural networks. arXiv:2106.13264. Retrieved from https://arxiv.org/abs/2106.13264
[12] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. 2019. Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 3558–3565.
[13] Matthias Fey and Jan E. Lenssen. 2019. Fast graph representation learning with PyTorch geometric. In *Proceedings of the ICLR Workshop on Representation Learning on Graphs and Manifolds*.
[14] Laura Forastiere, Edoardo M. Airoldi, and Fabrizia Mealli. 2021. Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association* 116, 534 (2021), 901–918.
[15] David F. Gleich, Lek-Heng Lim, and Yongyang Yu. 2015. Multilinear PageRank. *SIAM Journal on Matrix Analysis and Applications* 36, 4 (2015), 1507–1541.
[16] Ruocheng Guo, Jundong Li, and Huan Liu. 2020. Learning individual causal effects from networked observational data. In *Proceedings of 13th International Conference on Web Search and Data Mining*, 232–240.
[17] Shonosuke Harada and Hisashi Kashima. 2021. Graphite: Estimating individual effects of graph-structured treatments. In *Proceedings of 30th ACM International Conference on Information & Knowledge Management*, 659–668.
[18] Jennifer L. Hill. 2011. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics* 20, 1 (2011), 217–240.
[19] Hsun-Ping Hsieh, Fandel Lin, Cheng-Te Li, Ian En-Hsu Yen, and Hsin-Yu Chen. 2019. Temporal popularity prediction of locations for geographical placement of retail stores. *Knowledge and Information Systems* 60, 1 (2019), 247–273.
[20] Changhui Hu, Weidong Liu, Yuqiu Jia, and Yaya Jin. 2019. Characterization of territorial spatial agglomeration based on POI data: A case study of Ningbo City, China. *Sustainability* 11, 18 (2019), 5083.
[21] Kosuke Imai, Zhichao Jiang, and Anup Malani. 2021. Causal inference with interference and noncompliance in two-stage randomized experiments. *Journal of the American Statistical Association* 116, 534 (2021), 632–644.
[22] Jean Kaddour, Yuchen Zhu, Qi Liu, Matt J. Kusner, and Ricardo Silva. 2021. Causal effect inference for structured treatments. In *Advances in Neural Information Processing Systems* 34, 24841–24854.
[23] Eun-Sol Kim, Woo Young Kang, Kyoung-Woon On, Yu-Jung Heo, and Byoung-Tak Zhang. 2020. Hypergraph attention networks for multimodal learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14581–14590.
[24] Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. 2019. Set transformer: A framework for attention-based permutation-invariant neural networks. In *Proceedings of the International Conference on Machine Learning*. PMLR, 3744–3753.
[25] Shuangli Li, Jingbo Zhou, Ji Liu, Tong Xu, Enhong Chen, and Hui Xiong. 2023. Multi-temporal relationship inference in urban areas. In *Proceedings of 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1316–1327.
[26] Shuangli Li, Jingbo Zhou, Tong Xu, Dejing Dou, and Hui Xiong. 2022. GeomGCL: Geometric graph contrastive learning for molecular property prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence* 36, 4541–4549.
[27] Shuangli Li, Jingbo Zhou, Tong Xu, Hao Liu, Xinjiang Lu, and Hui Xiong. 2020. Competitive analysis for points of interest. In *Proceedings of 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1265–1274.
[28] Defu Lian, Yongji Wu, Yong Ge, Xing Xie, and Enhong Chen. 2020. Geography-aware sequential location recommendation. In *Proceedings of 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2009–2019.

[29] Defu Lian, Cong Zhao, Xing Xie, Guangzhong Sun, Enhong Chen, and Yong Rui. 2014. GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Proceedings of 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 831–840.
[30] Yuxuan Liang, Kun Ouyang, Junkai Sun, Yiwei Wang, Junbo Zhang, Yu Zheng, David Rosenblum, and Roger Zimmermann. 2021. Fine-grained urban flow prediction. In *Proceedings of the Web Conference 2021*, 1833–1845.
[31] Yanchi Liu, Chuanren Liu, Xinjiang Lu, Mingfei Teng, Hengshu Zhu, and Hui Xiong. 2017. Point-of-interest demand modeling with human mobility patterns. In *Proceedings of 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 947–955.
[32] Chenyu Lu, Min Pang, Yang Zhang, Hengji Li, Chengpeng Lu, Xianglong Tang, and Wei Cheng. 2020. Mapping urban spatial structure based on poi (point of interest) data: A case study of the Central city of Lanzhou, China. *ISPRS International Journal of Geo-Information* 9, 2 (2020), 92.
[33] Jing Ma, Ruocheng Guo, Aidong Zhang, and Jundong Li. 2021. Multi-cause effect estimation with disentangled confounder representation. In *Proceedings of 30th International Joint Conference on Artificial Intelligence*, 2790–2796.
[34] Jing Ma, Mengting Wan, Longqi Yang, Jundong Li, Brent Hecht, and Jaime Teevan. 2022. Learning causal effects on hypergraphs. In *Proceedings of 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1202–1212.
[35] Xinkun Nie and Stefan Wager. 2021. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* 108, 2 (2021), 299–319.
[36] Elizabeth L. Ogburn, Ilya Shpitser, and Youjin Lee. 2020. Causal inference, social networks and chain graphs. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)* 183, 4 (2020), 1659–1676.
[37] Elizabeth L. Ogburn, Oleg Sofrygin, Ivan Diaz, and Mark J. Van der Laan. 2024. Causal inference for social network data. *Journal of the American Statistical Association* 119, 545 (2024), 597–611.
[38] Yingrong Qin, Chen Gao, Yue Wang, Shuangqing Wei, Depeng Jin, Jian Yuan, and Lin Zhang. 2022. Disentangling geographical effect for point-of-interest recommendation. *IEEE Transactions on Knowledge and Data Engineering* 35, 8 (2022).
[39] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2012. BPR: Bayesian personalized ranking from implicit feedback. arXiv:1205.2618. Retrieved from https://arxiv.org/abs/1205.2618
[40] Peter M. Robinson. 1988. Root-N-consistent semiparametric regression. *Econometrica* 56, 4 (1988), 931–954.
[41] Paul R. Rosenbaum and Donald B. Rubin. 1983. The Central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 1 (1983), 41–55.
[42] Paul R. Rosenbaum and Donald B. Rubin. 1984. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* 79, 387 (1984), 516–524.
[43] Donald B. Rubin. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association* 100, 469 (2005), 322–331.
[44] Mehebub Sahana, Haoyuan Hong, and Haroon Sajjad. 2018. Analyzing urban spatial patterns and trend of urban growth using urban sprawl matrix: A study on Kolkata urban agglomeration, India. *The Science of the Total Environment* 628 (2018), 1557–1566.
[45] Shiv Kumar Saini, Sunny Dhamnani, Akil Arif Ibrahim, and Prithviraj Chavan. 2019. Multiple treatment effect estimation using deep generative model with task embedding. In *Proceedings of the World Wide Web Conference*, 1601–1611.
[46] Uri Shalit, Fredrik D. Johansson, and David Sontag. 2017. Estimating individual treatment effect: Generalization bounds and algorithms. In *Proceedings of the International Conference on Machine Learning*. PMLR, 3076–3085.
[47] Eric J. Tchetgen Tchetgen and Tyler J. VanderWeele. 2012. On causal inference in the presence of interference. *Statistical Methods in Medical Research* 21, 1 (2012), 55–75.
[48] Christoph Teller, Andrew Alexander, and Arne Floh. 2016. The impact of competition and cooperation on the performance of a retail agglomeration and its stores. *Industrial Marketing Management* 52 (2016), 6–17.
[49] Waldo R. Tobler. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46, Supp. 1 (1970), 234–240.
[50] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems* 30.
[51] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. arXiv:1710.10903. Retrieved from https://arxiv.org/abs/1710.10903
[52] Hao Wang, Yanmei Fu, Qinyong Wang, Hongzhi Yin, Changying Du, and Hui Xiong. 2017. A location-sentiment-aware recommender system for both home-town and out-of-town users. In *Proceedings of 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1143.
[53] Qianru Wang, Bin Guo, Lu Cheng, Zhiwen Yu, and Huan Liu. 2022. CausalSE: Understanding varied spatial effects with missing data towards adding new bike-sharing stations. *ACM Transactions on Knowledge Discovery from Data* 17, 2 (2022), 1–24.

[54] Suhang Wang, Yilin Wang, Jiliang Tang, Kai Shu, Suhas Ranganath, and Huan Liu. 2017. What your images reveal: Exploiting visual contents for point-of-interest recommendation. In *Proceedings of 26th International Conference on World Wide Web*, 391–400.

[55] Shen Wang, Xiaokai Wei, Cicero Nogueira Nogueira dos Santos, Zhiguo Wang, Ramesh Nallapati, Andrew Arnold, Bing Xiang, Philip S. Yu, and Isabel F. Cruz. 2021. Mixed-curvature multi-relational graph neural network for knowledge graph completion. In *Proceedings of the Web Conference 2021*, 1761–1771.

[56] Congxi Xiao, Jingbo Zhou, Jizhou Huang, Hengshu Zhu, Tong Xu, Dejing Dou, and Hui Xiong. 2023. A contextual master-slave framework on urban region graph for urban village detection. In *Proceedings of the 2023 IEEE 39th International Conference on Data Engineering (ICDE)*. IEEE, 736–748.

[57] Haoran Xin, Xinjiang Lu, Nengjun Zhu, Tong Xu, Dejing Dou, and Hui Xiong. 2022. CAPTOR: A Crowd-aware pre-travel recommender system for out-of-town users. In *Proceedings of 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1174–1184.

[58] Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. 2019. HyperGCN: A new method for training graph convolutional networks on hypergraphs. In *Advances in Neural Information Processing Systems* 32.

[59] Dingqi Yang, Daqing Zhang, Vincent W. Zheng, and Zhiyong Yu. 2014. Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 45, 1 (2014), 129–142.

[60] Yang, Yang Yaqian Duan, Xinze Wang, Zi Huang, Ning Xie, and Heng Tao Shen. 2018. Hierarchical multi-clue modelling for poi popularity prediction with heterogeneous tourist information. *IEEE Transactions on Knowledge and Data Engineering* 31, 4 (2018), 757–768.

[61] Teng Ye, Wei Ai, Lingyu Zhang, Ning Luo, Lulu Zhang, Jieping Ye, and Qiaozhu Mei. 2020. Predicting individual treatment effects of large-scale team competitions in a ride-sharing economy. In *Proceedings of 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2368–2377.

[62] Jinsung Yoon, James Jordon, and Mihaela Van Der Schaar. 2018. GANITE: Estimation of individualized treatment effects using generative adversarial nets. In *Proceedings of the International Conference on Learning Representations*.

[63] Ruiyun Yu, Dezhi Ye, and Jie Li. 2020. RePiDem: A refined poi demand modeling based on multi-source data. In *Proceedings of the IEEE INFOCOM 2020—IEEE Conference on Computer Communications*. IEEE, 964–973.

[64] Xiang Yu, Zhanyun Wu, Heren Zheng, Manqi Li, and Tianle Tan. 2020. How urban agglomeration improve the emission efficiency? A spatial econometric analysis of the Yangtze river Delta urban agglomeration in China. *Journal of Environmental Management* 260 (2020), 110061.

[65] Wenhua Yuan, Jianchun Li, Li Meng, Xiaonan Qin, and Xiaoxing Qi. 2019. Measuring the area green efficiency and the influencing factors in urban agglomeration. *Journal of Cleaner Production* 241 (2019), 118092.

[66] Yuan Yuan, Kristen Altenburger, and Farshad Kooti. 2021. Causal network motifs: Identifying heterogeneous spillover effects in a/b tests. In *Proceedings of the Web Conference 2021*, 3359–3370.

[67] Yu Zheng, Chen Gao, Xiang Li, Xiangnan He, Yong Li, and Depeng Jin. 2021. Disentangling user interest and conformity for recommendation with causal embedding. In *Proceedings of the Web Conference 2021*, 2980–2991.

Received 27 March 2024; revised 11 September 2025; accepted 6 January 2026