Minimizing the Regret of an Influence Provider

Yipeng Zhang[1], Yuchen Li[2], Zhifeng Bao[1],* Baihua Zheng[2], H. V. Jagadish[3]
[1]RMIT University [2]Singapore Management University [3]University of Michigan
{yipeng.zhang, zhifeng.bao}@rmit.edu.au {yuchenli, bhzheng}@smu.edu.sg jag@umich.edu

ABSTRACT

Influence maximization has been studied extensively from the perspective of the influencer. However, the influencer typically purchases influence from a provider, for example in the form of purchased advertising. In this paper, we study the problem from the perspective of the influence provider. Specifically, we focus on influence providers who sell Out-of-Home (OOH) advertising on billboards. Given a set of requests from influencers, how should an influence provider allocate resources to minimize regret, whether due to forgone revenue from influencers whose needs were not met or due to over-provisioning of resources to meet the needs of influencers? We formalize this as the Minimizing Regret for the OOH Advertising Market problem (MROAM). We show that MROAM is both NP-hard and NP-hard to approximate within any constant factor. The regret function is neither monotone nor submodular, which renders any straightforward greedy approach ineffective. Therefore, we propose a randomized local search framework with two neighborhood search strategies, and prove that one of them ensures an approximation factor to a dual problem of MROAM. Experiments on real-world user movement and billboard datasets in New York City and Singapore show that on average our methods outperform the baselines in effectiveness by five times.

CCS CONCEPTS

• **Theory of computation** → **Optimization with randomized search heuristics**.

KEYWORDS

Outdoor Advertising, Regret Minimization, Influence Provider

ACM Reference Format:
Yipeng Zhang[1], Yuchen Li[2], Zhifeng Bao[1],* Baihua Zheng[2], H. V. Jagadish[3]. 2021. Minimizing the Regret of an Influence Provider. In *Proceedings of the 2021 International Conference on Management of Data (SIGMOD '21), June 20–25, 2021, Virtual Event, China.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3448016.3457257

1 INTRODUCTION

Influence Maximization (IM) has been studied extensively [17]. A typical problem setting is to maximize the influence subject to limits on the expenditure or the number of seeded nodes. In a marketplace, if customers request influence, they purchase it from some influence provider; the provider performs the services to provide influence and earn profits. Most, if not all, existing literatures related to IM study the problem from the perspective of a customer who requests certain influence. Different from existing work, we investigate the problem from the perspective of influence provider and aim to maximize the influence provider's profit. In the rest of this paper, we use the following terms interchangeably: host and influence provider; advertiser and customer.

Since the host-based optimization is much harder than that of the advertiser, as we will demonstrate shortly, we strategically choose a straightforward influence model, without follow-on network effects. For concreteness, we describe our problem under the Out-of-Home (OOH) advertising scenario, which has been studied extensively in recent literatures [19, 24, 26, 27, 29]. The general applicability of our problem will be illustrated later in this section.

Our Observation. Existing studies in the OOH scenario all share a common objective: to help the single advertiser achieve the largest influence under her budget constraint. However, a more challenging yet unexplored problem, as confirmed from real-world OOH hosts [3, 13, 16], has the following setting: the host needs to deal with multiple advertisers coming every day. It is a standard practice for each advertiser to submit a campaign proposal to the host by specifying a demanded influence and a corresponding committed payment that will be fully paid only if her demand is achieved.

Our Problem. Motivated by this observation, we propose and study the ad allocation problem from the host's perspective, who is responsible for assigning billboards to all advertisers. The host owns large number of billboards, each with known influence, and each advertiser seeks a subset of the billboards with aggregate influence reaching her demand. The host gains the maximum payment if all advertisers are satisfied. By following this rationale, we propose a novel "regret" model to guide the host in assigning billboards to advertisers. In particular, there are two types of regret that affect profit, namely *revenue regret* and *excessive influence regret*. The former case arises when the host cannot meet the demand of an advertiser, while the latter case arises when the host overly satisfies an advertiser. Note that both undesirable cases are independent and they could occur together, as we present in Example 1.

Example 1. *A host owns six billboards $\mathcal{U} = \{o_1, \cdots, o_6\}$, and $I(o_i)$ indicates the influence of the billboard o_i, as reported in Table 1. Three advertisers, $\mathcal{A} = \{a_1, a_2, a_3\}$, approach the host for advertisement services, with each requesting her demanded influence $\mathcal{I}$ and the payment L she is willing to pay if the demand is satisfied, as listed in Table 2. To serve these advertisers, the host needs to deploy a set of billboards S_i to each advertiser to satisfy her demand. Let us consider two different deployment strategies, as shown in Table 3 and Table 4, respectively. Strategy 1 fails to satisfy the advertiser a_3 and hence the host cannot collect the full payment from a_3. Besides, the host wastes*

SIGMOD '21, June 20–25, 2021, Virtual Event, China

ACM ISBN 978-1-4503-8343-1/21/06…$15.00
https://doi.org/10.1145/3448016.3457257

*Zhifeng Bao is the corresponding author.

Table 1: Billboard Influence

$\mathcal{U}$	o_1	o_2	o_3	o_4	o_5	o_6
$I(o_i)$	2	6	3	7	1	1

Table 2: Advertiser Contract

$\mathcal{A}$	a_1	a_2	a_3
I_i	5	7	8
L_i	\$10	\$11	\$20

Table 3: Strategy 1

$\mathcal{A}$	a_1	a_2	a_3
S_i	o_2	o_4	o_1, o_3, o_5, o_6
Satisfy	Y	Y	N
$I(S_i)-I_i$	1	0	-1

Table 4: Strategy 2

$\mathcal{A}$	a_1	a_2	a_3
S_i	o_1, o_3	o_4	o_2, o_5, o_6
Satisfy	Y	Y	Y
$I(S_i)-I_i$	0	0	0

certain resources as S_1 assigned to a_1 has its influence exceeding that demanded by a_1. Ideally, the assigned billboards are expected to just satisfy, but not exceed, the advertisers' demanded influences, as the exceeded influences could be assigned to other advertisers to bring in additional revenues. Therefore, Strategy 2 is better.

Note that the host may suffer from revenue regret without any excessive influence regret (for example, if o_2 had an influence of 5 instead of 6 in strategy 2), or suffer from excessive influence regret without revenue regret (for example, if o_5 had an influence value of 2 instead of 1 in both strategies 1 and 2). Furthermore, to simplify our illustration, here we calculate $I(S)$ by simply aggregating $I(o_i)$, while in the real case, how to calculate $I(S)$ depends on concrete applications.

To this end, we formulate it as the Minimizing Regret for the OOH Advertising Market (MROAM) problem.

Hardness. We prove that MROAM is not only NP-hard but also NP-hard to approximate within any constant factor by using a reduction from the numerical 3-dimensional matching problem [8].

Our Solutions. Since MROAM is not tractable in general, we propose four heuristic methods by providing end-users with different levels of a trade-off between the efficiency and the regret value to achieve. The first two are greedy methods as baselines, where we assign the billboards based on a descending order of either the budget-effectiveness of advertisers or the regret-effectiveness of billboards. Without surprise, the greedy methods may produce a poor local minimum because the objective function of MROAM is non-monotone and non-submodular. To avoid the local minimum, we propose a randomized local search framework that iteratively samples a deployment plan and then performs a local search from the sampled plan until no further improvement could be achieved. In particular, it offers two local search methods: (1) *advertiser-driven local search*, and (2) *billboard-driven local search*. The former tries to exchange the set of billboards of one advertiser with the set of billboards of another advertiser. The latter performs a fine-grained search by examining whether any swap of two billboards can lead to a better solution. We also prove that the billboard-driven method guarantees an approximation factor to a dual problem of MROAM.

Empirical Evaluation. We use two real datasets to study how our methods behave w.r.t. audiences of two different transport modes. One dataset consists of taxi trips and roadside billboards in New York City, while the other consists of bus trips and bus stop billboards in Singapore. We also design the concept of *demand-supply ratio* and *average-individual demand ratio* to set up the demand at the macro-level (e.g., from low to excessive global demand) and micro-level (e.g., big vs. small advertiser), respectively, which in turn well-capture a wide range of real scenarios. Last, we report the evaluation results as well as insights on how and why different deployment strategies would benefit the host in practice.

General Applicability. The regret formulation proposed in this paper is applicable to many other scenarios where a company has to provision resources to meet customer needs. If the company provisions insufficient resources, it is unable to meet all customer needs; if it provisions excessive resources, it wastes valuable resources. The resources could be trucks for a logistics company, store locations for a large retail chain, or workers or staff at a temp agency. The exact optimization function could differ slightly, based on specific application needs. However, we believe that the techniques developed in this paper are applicable for different needs, with appropriate minor modifications. For instance, in the telecommunication marketing [22], the host owns telecommunication towers and mobile operators renting towers play the role of advertisers, where the demand of an operator is the number of customers accessing its network. The regret occurs if the host provides excessive or insufficient networking capability to the operator.

Main Contributions. *First*, we define a regret minimization framework for the host. Specifically, we study billboard placement from the host's perspective when dealing with multiple advertisers, which we call Minimizing Regret for OOH Advertising Market (MROAM) (Section 3). *Second*, we prove that MROAM is not only NP-hard but also NP-hard to approximate within a constant factor (Section 4). *Third*, we design two greedy methods. The first one satisfies advertisers based on the effective budget, while the second one treats each advertiser equally and assigns billboards to all advertisers synchronously (Section 5). *Fourth*, as these greedy heuristics can fall into a poor local minimum, we propose a randomized local search framework to cater to end-users with different requirements on the trade-off between efficiency and quality of the deployment strategies. We prove that one local search method guarantees an approximation factor to a dual problem of MROAM (Section 6). *Finally*, we design a novel setup to exhibit various real-world demand-supply relationships, and conduct extensive experiments on two real-world billboard and trajectory datasets to verify the effectiveness and the efficiency of our methods over different marketing conditions and transport modes (Section 7).

2 RELATED WORK

2.1 Regret Minimization in SVM

In the Social Viral Marketing (SVM) scenario, the host, such as Twitter and Facebook, provides a service of promoting the influence of ads on the social network to get payments from advertisers. It adopts a *cost per engagement* (CPE) business model [7, 12, 23], where the advertiser will pay the host for each click received by its ad.

Several studies on SVM [1, 2, 4] have been proposed recently. The most relevant one [2] tries to minimize the regret of the host, where unsatisfied or over-satisfied advertisers cause regrets. Each advertiser specifies a demanded influence. For the unsatisfied case, the advertiser only pays for the achieved influence; if the achieved influence exceeds the demand, the advertiser does not pay for the excessive influence. Consequently, both the underachieved payment from an advertiser and the excessive influence offered to an advertiser contribute to the regret of the host. The other two studies [1, 4] aim to maximize the revenue under CPE model. The revenue is defined as the sum of payments from advertisers, where the payment of each advertiser is the sum of CPE of activated nodes.

There are two main differences between SVM and our MROAM. *(1) Business models*: Unlike the CPE model adopted in SVM, in MROAM the host could receive a substantially reduced payment, and possibly even no payment at all if the demanded influence is not achieved. *(2) Influence models*: In MROAM, the geographical properties of billboards and users are utilized to build the influence relationship–a billboard influences all users who can "meet" that billboard (i.e., users and billboards are geographically close enough to each other), and the influence does not diffuse among users [26, 27, 29]. In contrast, regret minimization in SVM [2] is based on diffusible and probabilistic models such as the Independent Cascade model and the Linear Threshold model [5, 6, 9, 11, 14], under which the ad will diffuse from one user to another following a probability. Consequently, literatures in SVM mainly focus on how to efficiently and accurately evaluate the node's influence in a virtual social network. Given the above, the respective optimization problems under these two settings are fundamentally different.

2.2 OOH Influence Maximization

With an unprecedented increase in the availability and collection of trajectory data [25], recent studies [26, 27, 29] initiate the problem of maximizing the influence in OOH advertising. Although they adopt slightly different influence measurements to cover different business needs, they both stand in the advertiser's shoes and share a common goal of *maximizing the influence for a single advertiser*. In particular, a billboard is considered to influence an audience only if this billboard is close enough to the trajectory that this audience travels along. To be more specific, the studies [26, 27] adopt the traffic volume (with influence overlap considered). To avoid double-counting the same user that may meet multiple billboards posting the same ad; the study [29] considers the impression count (the times that a user meets the same ad) to trigger an influence flag.

A major difference between this category of work and our work is on the objectives to be served. In this work, we serve the host who deals with multiple advertisers to minimize the "loss" of the host and meanwhile provide "just-the-right" amount of influence to meet all advertisers' demands, while they simply focus on maximizing the influence for a single advertiser.

There are also some loosely related studies. In [10, 18, 28], visualization tools are designed to help an advertiser to select billboards. The authors in [15] try to extract meaningful data from social media and use them to improve the influence of targeted OOH advertising.

3 PROBLEM FORMULATION

3.1 Preliminaries

Billboard Influence. Given a billboards set S_i, their influence is denoted by $I(S_i)$. As reported in a recent work art [26], $I(S_i)$ can be measured in various ways. We adopt the same influence measurement as [26]. To avoid distracting the readers, we will present details on measuring $I(S_i)$ later in Section 7.1.2. Note that our approaches are orthogonal to the choices of measurements.

Advertiser. Given an advertiser set $\mathcal{A} = \{a_1, a_2, \cdots, a_{|\mathcal{A}|}\}$, each advertiser a_i submits a campaign proposal to the host for deploying her ads. The proposal contains a payment L_i and a minimum demanded influence $\mathcal{I}_i$ for a_i.

3.2 Problem Definition

Without loss of generality, we argue that a *regret* model should cater for the following two cases.

Case 1: Revenue Regret. The host can receive the full payment L_i from an advertiser a_i only if the host assigns a billboard set S_i that meets her demand $a_i.\mathcal{I}_i$. Otherwise (i.e., $a_i.\mathcal{I}_i > I(S_i)$), the host will only receive partial payment and hence suffer from a revenue loss.

Case 2: Excessive Influence Regret. When the achieved influence exceeds this advertiser's demand (i.e., $I(S_i) > a_i.\mathcal{I}_i$), the host does not receive any additional payment. In this case, the host would rather use such excessive influences to fulfill requests from other advertisers. Hence, it renders an opportunity cost (i.e., excessive influence regret) of the current plan, which is defined as $L_i \cdot \frac{I(S_i)-\mathcal{I}_i}{\mathcal{I}_i}$.

Regret Model. Following the above, we formulate the regret model for the host to assign a billboard set S_i to an advertiser a_i as

$$R(S_i) = \begin{cases} L_i(1 - \gamma\frac{I(S_i)}{\mathcal{I}_i}) & \text{if } a_i.\mathcal{I}_i > I(S_i) \\ L_i \cdot \frac{I(S_i)-\mathcal{I}_i}{\mathcal{I}_i} & \text{otherwise} \end{cases} \tag{1}$$

where $I(S_i)/\mathcal{I}_i$ is the fraction of satisfied influence by required influence, and γ is the penalty ratio due to the unsatisfied demand. When $\gamma = 1$, the host can receive the fraction of payment as the same fraction of influence that has been satisfied (i.e., $I(S_i)/\mathcal{I}_i$); when $\gamma = 0$, the host cannot receive any payment unless the required influence is fully satisfied. The choice of γ is orthogonal to our problem. For more details on the selection of γ, please refer to the experiments reported in Section 7.4.

Finally, we are ready to present the MROAM problem.

Definition 3.1. *(MROAM.) Given a billboard database $\mathcal{U}$, a trajectory database $\mathcal{T}$, and an advertiser set $\mathcal{A}$, the goal of MROAM is to find a billboard deployment strategy $S = \{S_1, ..., S_{|\mathcal{A}|}\}$ for all advertisers, which can minimize the total regret of the host, such that each billboard is only assigned to at most one advertiser. Formally:*

$$\arg\min_{S} \quad R(S) = \sum_{S_i \in S} R(S_i), \quad \textit{subject to: } S_i \cap S_j = \emptyset$$

Discussion. First, the problem of revenue maximization is actually a subclass of regret minimization. Generally, the revenue is defined as the income [1, 4]. In the unsatisfied case, the regret is the 'lost' revenue. In the over-satisfied case, the regret can essentially capture the 'free services' provided by the host. Unfortunately, free services cannot be captured by the revenue at all. Second, since the billboard cost is a fixed portion in both satisfied and unsatisfied cases, we do not incorporate it in the regret function, and it does not affect our solution no matter whether the cost is considered or not. Last, the billboard can be a digital one, where we treat each digital billboard as "multiple billboards", one for a certain time slot.

4 PROBLEM HARDNESS

In this section, we conduct a theoretical analysis on the hardness of the MROAM problem.

Theorem 1. *MROAM is NP-hard, and is NP-hard to approximate within any constant factor.*

We use a reduction from the numerical 3-dimensional matching (N3DM) problem [8] to prove the hardness of MROAM. Let b denote a bound and X, Y and Z denote three multisets of integers respectively, each containing n elements. The N3DM is a decision problem that asks whether there is a matching relation M of $X \times Y \times Z$ such that every integer in X, Y and Z occurs exactly once, and for every triple $(x_i, y_i, z_i) \in M$, we have $x_i + y_i + z_i = b$ hold. This problem is known to be NP-complete. It is noted that the matching exists only if $b = (\sum X + \sum Y + \sum Z)/n$.

We reduce the N3DM decision problem to MROAM with the following process:

(1) Set the number of billboards in $\mathcal{U}$ to be $3n$. Set the number of advertisers in $\mathcal{A}$ to be n. Set $\gamma = 0$.
(2) We divide billboards into three disjoint sets D_1, D_2 and D_3 equally, and map each element in X, Y and Z in the N3DM problem to a billboard of the three sets. Then, $|X| = |Y| = |Z| = n$, $|D_1| = |D_2| = |D_3| = n$, and $b = (\sum X + \sum Y + \sum Z)/n$.
(3) For each billboard $o_i \in \mathcal{U}$, let o_i influence a disjoint set of trajectories. The influence of o_i is set as the integer value of the corresponding element in the N3DM. To facilitate proofs, we use o_i to denote its influence in this section only.
(4) Let c be a large number. We revise the influence value of all billboards as $\forall o_i \in D_1, o_i \leftarrow c + o_i$, $\forall o_j \in D_2, o_j \leftarrow 3 \cdot c + o_j$, and $\forall o_k \in D_3, o_k \leftarrow 9 \cdot c + o_k$. After the revision, we set the demanded influence of all advertisers to be $\mathcal{I}_i = b + 13 \cdot c$. Note that, when $c \to \infty$, the minimum regret value 0 is achieved for the above setting only if for all $S_i \in S$, $S_i = \{(o_i, o_j, o_k) | o_i \in D_1, o_j \in D_2, o_k \in D_3\}$.

Clearly, the reduction can be done in polynomial time. Next, we are ready to prove the hardness of MROAM.

Proof. We show that the answer to the *MROAM* decision problem is YES (if the minimum regret value is zero) *if* and *only if* the answer to the N3DM decision problem is YES.

The If Direction. When the answer to the *MROAM* decision problem is YES, there must exist S such that $R(S) = 0$. This implies that for every billboard set $S_i \in S$, $R(S_i) = 0$, i.e., $\mathcal{I}_i = I(S_i)$. It is because (1) $I(D_1 \cup D_2 \cup D_3) = \sum_{S_i \in S} I(S_i)$, and (2) for $1 \leq i < j \leq n$, $\mathcal{I}_i = \mathcal{I}_j$. If for any billboard set $S_i \in S$, $I(S_i) > \mathcal{I}_i$, then there must exist at least one billboard set S_j where $I(S_j) < \mathcal{I}_j$, which implies $R(S_j) > 0$. Since each element $x_i \in X$, $y_i \in Y$ and $z_i \in Z$ is mapped to the influence value of the corresponding billboard, for each triple $(x_i, y_i, z_i) \in M$, $(x_i + y_i + z_i) = I(S_i) - 13 \cdot c = \mathcal{I}_i - 13 \cdot c = b$. As a result, the answer to the N3DM decision problem is YES.

The Only-if Direction. When the answer to the N3DM decision problem is YES, there must exist $M \subset X \times Y \times Z$ such that, for all $(x_i, y_i, z_i) \in M$, $(x_i + y_i + z_i) = b$. Since the value of each x_i, y_i, z_i is mapped to the influence value of the corresponding billboard, for the corresponding billboard set $S_i \in S$, $I(S_i) = b + 13 \cdot c = \mathcal{I}_i$. We have $R(S) = 0$ as $R(S_i) = 0$ for all $S_i \in S$. Hence the answer to the decision problem of *MROAM* is YES.

Based on the above arguments, the N3DM decision problem is equivalent to deciding whether there is a billboard deployment strategy to achieve zero regret. Since the N3DM decision problem is NP-complete, the decision problem of MROAM is NP-complete, and the optimization problem is NP-hard, even if $|\mathcal{T}|$ is restricted to be the number of polynomial value $|\mathcal{U}|$.

Approximation Hardness. We next show: if MROAM can be approximated with any factor in polynomial time, then the N3DM decision problem can be solved in polynomial time. Let OPT_n denote the number of triples whose summations are not b in N3DM. Let OPT_m denote the instance of MROAM to which the N3DM problem matching is reduced. We can conclude $OPT_n = 0$ if and only if $OPT_m = 0$. Suppose *ALG* is an algorithm that approximates MROAM within a factor of τ. Then, the minimum $R(S)$ achieved by *ALG* on any instance of MROAM is smaller than $\tau \cdot OPT_m$, i.e., $R(S) \leq \tau \cdot OPT_m$. Hence, when $OPT_n = OPT_m = 0$, we have $R(S) = 0$; when $OPT_n \neq 0$, we have $R(S) \geq OPT_m > 0$. Based on the above, we can solve N3DM within polynomial time by checking whether $R(S) = 0$, which is impossible unless NP=P. Hence, it is NP-hard to approximate MROAM within any constant factor. □

Algorithm 1: Budget-Effective Greedy

Input: $\mathcal{U}$, $\mathcal{T}$, $\mathcal{A}$
Output: S

1.1 Order each advertiser $a_i \in \mathcal{A}$ based on descending order of $L_i/\mathcal{I}_i$
1.2 Initialize $S \leftarrow \{S_1, ..., S_{|\mathcal{A}|}\}$
1.3 **foreach** $a_i \in \mathcal{A}$ **do**
1.4   **while** $\mathcal{U} \neq \emptyset \wedge \mathcal{I}_i > I(S_i)$ **do**
1.5   Select $o \in \mathcal{U}$ that maximizes $\frac{R(S_i) - R(S_i \cup \{o\})}{I(\{o\})}$
1.6 $S_i \leftarrow S_i \cup \{o\}$
1.7 $\mathcal{U} \leftarrow \mathcal{U} \setminus \{o\}$
1.8 return S

5 TWO GREEDY HEURISTICS

The hardness of MROAM implies that efficient algorithms with theoretical guarantee w.r.t. the optimal regret do not exist unless NP=P. Given its hardness, we first propose an efficient greedy heuristic that orders the advertisers based on their *budget-effectiveness*, and then prioritizes more budget-effective ones in the deployment (Section 5.1). However, as all ideal billboards will likely be assigned to a few budget-effective advertisers, advertisers with lower budget-effectiveness might be unsatisfied due to lack of ideal billboards. Hence, we propose an improved greedy heuristic by deploying ideal billboards to all advertisers synchronously (Section 5.2).

5.1 Budget-effective Greedy

Algorithm 1 is an efficient greedy heuristic. In Line 1.1, we order all advertisers by their budget-effectiveness $L_i/\mathcal{I}_i$, (i.e., budget over demand). Then, we initialize an empty set of billboard sets for each advertiser (Line 1.2). Next, a greedy heuristic is employed to keep assigning the billboard that can best reduce the regret (i.e., maximizing $(R(S_i) - R(S_i \cup \{o\}))/I(\{o\})$) to fulfill the next budget-effective advertiser a_i (Lines 1.3-1.7). After all advertisers are satisfied or the host runs out of billboards, the billboard sets S will be returned (Line 1.8).

5.2 Synchronous Greedy

The above greedy method may fall into a trap where most of the ideal billboards are exhausted first. Subsequently, the rest of the

Algorithm 2: Synchronous Greedy

Input: $\mathcal{U}, \mathcal{T}, \mathcal{A}, S^{in}$ ($S^{in} = \{S_1^{in}, ..., S_{|\mathcal{A}|}^{in}\}$)
Output: S

2.1 $S \leftarrow S^{in}$
2.2 **while** *TRUE* **do**
2.3   **foreach** $a_i \in \mathcal{A}$ **do**
2.4   **if** $\mathcal{I}_i > I(S_i)$ **then**
2.5   **if** $\mathcal{U} \neq \emptyset$ **then**
2.6   Select $o \in \mathcal{U}$ that maximizes $\frac{R(S_i)-R(S_i \cup \{o\})}{I(\{o\})}$
2.7 $S_i \leftarrow S_i \cup \{o\}$
2.8 $\mathcal{U} \leftarrow \mathcal{U} \setminus \{o\}$
2.9   **if** *more than two* $a_i \in \mathcal{A}$ *are not satisfied* **then**
2.10   Release $S_j \in S$ such that $\mathcal{I}_j > I(S_j)$ and has minimum $L_j/\mathcal{I}_j$
2.11 $\mathcal{A} \leftarrow \mathcal{A} \setminus \{a_j\}$
2.12   **else**
2.13   return S

advertisers whose demands have not yet been satisfied may not have ideal billboard deployment strategies. Therefore, we extend Algorithm 1 by assigning ideal billboards to all advertisers synchronously. In Lines 2.3-2.8, we assign one billboard that can maximize $(R(S_i) - R(S_i \cup \{o\}))/I(\{o\})$ to each advertiser whose demand has not been satisfied. During the iterations, if there is no more billboard, we iteratively release billboards from the least budget-effective advertiser a_i (Line 2.10) and remove a_i from $\mathcal{A}$ (Line 2.11), until the billboards are enough for the rest of advertisers. Eventually, with the decrease of $|\mathcal{A}|$, the while loop breaks as fewer than two advertisers are unsatisfied (Line 2.13). In Algorithm 2, the input $S^{in} = \{S_1^{in}, ..., S_{|\mathcal{A}|}^{in}\}$; S_i^{in} denotes the billboards assigned to advertiser a_i. In this algorithm, S^{in} is an empty set, but it is non-empty when this algorithm is invoked as a routine by the local search methods to be presented in Sections 6.

6 A LOCAL SEARCH FRAMEWORK

The greedy heuristic can generate results with a constant approximation ratio to the optimal solution only when the objective function is monotone and submodular. Unfortunately, the objective $R(S)$ of MROAM is neither monotone nor submodular, as shown in the following counterexample.

Example 2. *Given two billboard sets S_1 and S_2, where $I(S_1) = 8$ and $I(S_2) = 9$, and $S_1 \subset S_2$, and a billboard o_1 such that $o_1 \notin S_1 \cup S_2$, we assume $I(S_1 \cup \{o_1\}) = 9$ and $I(S_2 \cup \{o_1\}) = 10$. Now we have an advertiser such that $\mathcal{I} = 10$ and $L = 10$. Obviously, $R(S_1) = 10 - 8\gamma$, $R(S_1 \cup \{o_1\}) = 10 - 9\gamma$, $R(S_2) = 10 - 9\gamma$, and $R(S_2 \cup \{o\}) = 0$. Hence, $R(S_1) - R(S_1 \cup \{o_1\}) < R(S_2) - R(S_2 \cup \{o_1\})$. Let $S_2' = S_2 \cup \{o_1\}$, for any $o_2 \notin S_2'$, $I(S_2' \cup \{o_2\}) > 10$. Hence, $R(S_2' \cup \{o_2\}) > 0 = R(S_2')$. Therefore, the objective $R(S)$ is neither monotone nor submodular.*

As a result, Algorithm 2 can easily produce a poor local minimum. To enhance the result quality, we introduce a local search framework where two local search methods are designed to provide different levels of a trade-off between the quality of the result and the efficiency of the search. In particular, we first propose a randomized greedy heuristic with a local search strategy that swaps deployment plans between *advertisers* (Section 6.1). We then introduce a fine-grained local search method that swaps the assignments of *billboards* to enhance the solution quality (Section 6.2). We also prove that our fine-grained local search method can provide theoretical guarantees for a dual problem of MROAM (Section 6.3).

Algorithm 3: Randomized Local Search

Input: $\mathcal{U}, \mathcal{T}, \mathcal{A}$
Output: S^{best}

3.1 $S^{best} \leftarrow$SynchronousGreedy($\mathcal{U}, \mathcal{T}, \mathcal{A}, \emptyset$)
3.2 **while** *the number of iterations < a preset count* **do**
3.3 $\mathcal{U}^* \leftarrow \mathcal{U}$
3.4   **for** $a_i \in \mathcal{A}$ **do**
3.5 $S_i \leftarrow$ {a random billboard $o \in \mathcal{U}^*$}
3.6 $\mathcal{U}^* \leftarrow \mathcal{U}^* \setminus \{o\}$
3.7 $S \leftarrow \{S_1, ..., S_{|\mathcal{A}|}\}$
3.8 $S^* \leftarrow$ SynchronousGreedy($\mathcal{U}^*, \mathcal{T}, \mathcal{A}, S$)
3.9 $S^{can} \leftarrow$ Advertiser-drivenLocalSearch($\mathcal{U}^*, \mathcal{T}, S^*$)
3.10   **if** $R(S^{can}) < R(S^{best})$ **then**
3.11 $S^{best} \leftarrow S^{can}$
3.12 return S^{best}

6.1 Advertiser-driven Local Search (ALS)

Our randomized local search strategy is presented in Algorithm 3. In Line 3.1, we initialize the current best plan using the synchronous greedy (i.e., Algorithm 2). Subsequently, we generate a number of baseline plans (S^* in Line 3.8 denotes a baseline plan) and execute the local search starting from S^* (Lines 3.2-3.9). In the following, we explain how to generate a baseline plan S^* and how to perform a local search based on S^*.

The generation of a baseline plan consists of two steps. First, we randomly assign a billboard o to an advertiser (Lines 3.4-3.7). In other words, we form a non-empty initial plan S_i^{in} by assigning a random billboard to each advertiser. Second, we execute the greedy search developed previously (i.e., Algorithm 2) to assign the remaining billboards to the advertisers. This two-step process is expected to generate a baseline plan with a reasonable regret value, and it incorporates probabilistic assignments to escape from a local minimum. Note that, the input S_i^{in} to Algorithm 2 is not empty.

For each generated baseline plan, we perform a local search to explore its "neighborhood" search space by exchanging the set of billboards assigned to one advertiser with the set of billboards assigned to another advertiser. We name this local search strategy as the *advertiser-driven local search*, and present it in Algorithm 4. It iteratively selects a pair of advertisers and checks the two sets of billboards assigned. If the exchange of the two sets of billboards can lead to a smaller regret of S^{can}, the exchange is executed and S^{best} is replaced by S^{can} (Lines 4.4-4.8). The local search terminates once no further improvement can be achieved from any candidate neighborhood plan.

6.2 Billboard-driven Local Search (BLS)

The advertiser-driven local search can escape a local minimum when an advertiser occupies a large number of billboards. However, releasing all the billboards allocated to an advertiser is a coarse-grained optimization and could miss a potentially better solution, as shown in the following example.

Algorithm 4: Advertiser-driven Local Search (ALS)

Input: $\mathcal{U}, \mathcal{T}, S^{best}$
Output: S^{best}
4.1 **while** *TRUE* **do**
4.2 $\quad S^{can} \leftarrow S^{best}$
4.3 $\quad$ **foreach** $a_i \in \mathcal{A}$ **do**
4.4 $\quad\quad$ **foreach** $a_j \in \mathcal{A} \backslash \{a_i\}$ **do**
4.5 $\quad\quad\quad$ **if** *Exchange S_i with S_j will reduce $R(S^{can})$* **then**
4.6 $\quad\quad\quad\quad$ Exchange S_i with S_j
4.7 $\quad$ **if** *$R(S^{can}) < R(S^{best})$* **then**
4.8 $\quad\quad S^{best} \leftarrow S^{can}$
4.9 $\quad$ **else**
4.10 $\quad\quad$ return S^{best}

Example 3. *Assume we have two advertisers a_1 and a_2, such that $\mathcal{I}_1 = x$, $L_1 = x$, $\mathcal{I}_2 = x - 1$, and $L_2 = x - 1$, where $x > 3$, and three billboards o_1, o_2 and o_3, where o_1 influences $\{t_1, \cdots, t_{x-1}\}$, o_2 influences $\{t_1, \cdots, t_{x-2}, t_x\}$, and o_3 influences $\{t_x, t_{x+1}\}$. Let $S_1 = \{o_1, o_2\}$ and $S_2 = \{o_3\}$ be two plans to serve two advertisers a_1 and a_2. Accordingly, we have $I(S_1) = x$, and $I(S_2) = 2$. The total regret $R(S) = R(S_1) + R(S_2) = x - 1 - 2\gamma$. Exchanging S_1 with S_2 will lead to a larger regret $R(S) = x + 1 - 2\gamma$. However, if we only exchange o_1 with o_3, then $S_1 = \{o_2, o_3\}$ and $S_2 = \{o_1\}$. Subsequently, $I(S_1) = x$, and $I(S_2) = x - 1$. Consequently, we are able to achieve a smaller regret as $R(S) = R(S_1) + R(S_2) = 0$.*

Motivated by this drawback, we propose a fine-grained local search approach by only exchanging two billboards, instead of two sets of billboards, at a time, as shown in Algorithm 5. In particular, given a current best billboard assignment S^{best}, we examine the neighborhood search space around S^{best} by performing the following four moves as long as they could help reduce the regret.

(1) Exchange a billboard assigned to an advertiser with another billboard assigned to a different advertiser (Lines 5.4-5.6).
(2) Replace a billboard assigned to an advertiser with an unassigned billboard (Lines 5.7-5.8).
(3) Release a billboard assigned to an advertiser (Lines 5.9-5.10).
(4) Allocate unassigned billboards (Lines 5.11-5.13).

The local search terminates once no further improvement can be achieved from the moves.

6.3 Theoretical Analysis

Since MROAM cannot be approximated by an efficient algorithm, we rewire the minimum regret problem to a maximum revenue problem to facilitate the proof on the approximation ratio. We show that under the rewired problem, the billboard-driven local search (BLS) method can achieve an approximation factor. We define the rewired objective $R' = \sum_{S_i \in S} R'(S_i)$ as follows:

$$R'(S_i) = \begin{cases} L_i \cdot \frac{I(S_i)}{\mathcal{I}_i} & \text{if } \; a_i.\mathcal{I}_i > I(S_i) \\ L_i - L_i \cdot \frac{I(S_i) - \mathcal{I}_i}{\mathcal{I}_i} & \text{otherwise} \end{cases} \tag{2}$$

R' mimics R as $R(S_i) = 0$ iff $R'(S_i) = L_i$. Furthermore, $R(S_i) + R'(S_i) = L_i$ for any $I(S_i)$. Thus, minimizing R and maximizing R' are *dual* problems when the demanded influence can be achieved. We note that the rewired objective R' remains to be *neither* monotone *nor* submodular. Hence, we do not oversimplify the problem to make the analysis easier.

Algorithm 5: Billboard-driven Local Search (BLS)

Input: $\mathcal{U}, \mathcal{T}, S^{best}$
Output: S^{best}
5.1 **while** *TRUE* **do**
5.2 $\quad S^{can} \leftarrow S^{best}$
5.3 $\quad$ **foreach** $S_i \in S^{can}$ **do**
5.4 $\quad\quad$ **foreach** $S_j \in S^{can} \setminus S_i$ **do**
5.5 $\quad\quad\quad$ **if** *$\exists o_m \in S_i \wedge o_n \in S_j$ such that Exchange(o_m, o_n) will decrease $R(S^{can})$* **then**
5.6 $\quad\quad\quad\quad$ Exchange(o_m, o_n)
5.7 $\quad\quad$ **if** *$\exists o_m \in S_i \wedge o_n \in \mathcal{U}$ such that Exchange(o_m, o_n) will decrease $R(S^{can})$* **then**
5.8 $\quad\quad\quad$ Exchange(o_m, o_n)
5.9 $\quad\quad$ **if** *$\exists o_m \in S_i$ such that releasing o_m will decrease $R(S^{can})$* **then**
5.10 $\quad\quad\quad$ Release $o_m \in S_i$
5.11 $\quad S \leftarrow$ SynchronousGreedy$(\mathcal{U}, \mathcal{T}, \mathcal{A}, S^{can})$
5.12 $\quad$ **if** *$R(S) < R(S^{can})$* **then**
5.13 $\quad\quad S^{can} \leftarrow S$
5.14 $\quad$ The same as Lines 4.7-4.10

Next, we analyze the approximation factor of BLS if the objective is to maximize R'. To simplify the presentation, we study the case for one advertiser but the analysis procedure can be easily extended to support the case of multiple advertisers.

We define the local maximum obtained by BLS as the following:

Definition 6.1. *A deployment plan S is called a $(1{+}r)$-approximate local maximum, if $(1 + r)R'(S) \geq R'(S \backslash \{o\})$ for any billboard $o \in S$, and $(1 + r)R'(S) \geq R'(S \cup \{o\})$ for any $o \notin S$.*

Next, we prove the following lemma to demonstrate the properties of the local maximum S obtained by BLS.

Lemma 6.1. *If S is a $(1 + r)$-approximate local maximum, for any deployment plan V,*

$$R'(V) \leq \max\left((1 + r|\mathcal{U}|), (1 - \psi)^{-|\mathcal{U}|}\right) \cdot R'(S) \tag{3}$$

where $\psi = \max_{o \in \mathcal{U}} \frac{I(\{o\})}{\mathcal{I}}$ denotes the ratio of the maximum influence of one billboard to the demanded influence of the advertiser.

Proof. Let $V = V_1 \subseteq V_2 \subseteq V_3 \subseteq \ldots \subseteq V_k$ and $V_i \backslash V_{i-1} = \{o_i\}$. Let us consider two cases: (a) $I(S) < \mathcal{I}$; and (b) $I(S) \geq \mathcal{I}$.

In case (a), if $S \subseteq V$, then we can always find $V_a = S$ for some $a \leq k$. There also exists a V_b where $a \leq b \leq k$ such that $R'(V_b) \geq R'(V)$ and $R'(V_i) \leq R'(V_b)$ for all $i \leq b$. Further, we have $R'(V_i) - R'(V_{i-1}) \leq R'(S \cup \{o_i\}) - R'(S) \leq r \cdot R'(S)$ for $a \leq i \leq b$. The first inequality is due to the submodularity of the minimization part of R' whereas the second inequality is due to the local maximum of S. By summing up the inequalities between indices a and b, we get $R'(V) - R'(S) \leq r(b-a)R'(S)$. Hence, $R'(V) \leq [1 + r(b-a)]R'(S) \leq (1 + r|\mathcal{U}|)R'(S)$. Using a symmetric argument, we can also show that $R'(V) \leq (1 + r|\mathcal{U}|)R'(S)$ when $V \subseteq S$.

In case (b), if $S \subseteq V$, it is trivial to see that $R'(V) \leq R'(S)$. If $V \subseteq S$, let $V = V_k \subseteq V_{k+1} \subseteq \ldots \subseteq V_{k+p} = S$. We have $R'(V_{i-1}) - R'(V_i) \leq \psi \cdot R'(V_{i-1})$ and then lead to $R'(V_{i-1}) \leq (1 - \psi)^{-1} R'(V_i)$, for $k < i \leq k + p$. By recursively applying the inequalities between indices $k \leq i \leq k + p$ times, we have $R'(V) = R'(V_k) \leq (1 - \psi)^{-p} R'(V_{k+p}) \leq (1 - \psi)^{-|\mathcal{U}|} R'(S)$.

By combining cases (a) and (b), we prove the lemma. □

Table 5: Statistics of Datasets

	$\lvert\mathcal{T}\rvert$	$\lvert\mathcal{U}\rvert$	AvgDistance	AvgTravelTime
NYC	1.7×10^6	1462	2.9km	569s
SG	2.2×10^6	4092	4.2km	1342s

Given Lemma 6.1, we are now ready to prove the approximation guarantee for BLS.

Theorem 2. *The BLS method returns an approximation factor of* $\rho = \max[(1 + r|\mathcal{U}|), (1 - \psi)^{-|\mathcal{U}|}]$ *for maximizing* R'.

Proof. Consider an optimal plan OPT and let BLS swap billboards only if the improvement ratio is r. If the algorithm terminates, the set S obtained is a $(1 + r)$-approximate local maximum. We have the following inequalities: $\rho \cdot R'(S) \geq \rho \cdot R'(S \cup OPT) \geq R'(OPT)$. The first inequality is due to the fact that S is a local minimum and adding more billboards only results in a lower R' value. The second inequality holds since OPT is also a local minimum. □

7 EXPERIMENT

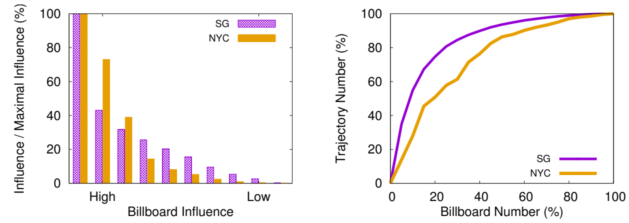
We conducted experiments to determine how well the various algorithms did in minimizing regret, and also how much computational time they required. Additionally, we highlight two key questions:

- Q1: What will happen if the **global demand** of all the advertisers is (far) below, close to, or over the maximum supply of the host?
- Q2: Which type of advertisers is better for the host in terms of minimizing the regret? A small number of big advertisers with high **individual demand** or a large number of small advertisers with low individual demand?

Since this is the *first* work on regret minimization for OOH advertising, there is no existing method to evaluate against. Instead, we provide extensive investigation of these two questions, thereby providing guidance on how to minimize the regret under different real-world scenarios. Please note that, in the rest of this section, we will use 'supply' as supplied influence for short, and 'demand' as demanded influence for short.

A straightforward setup is to select different numbers of advertisers $|\mathcal{A}|$, and then randomly set a demand $\mathcal{I}_i$ for each advertiser $a_i \in \mathcal{A}$ from an estimated range, say $[1000, 100000]$. However, we want to gain a deeper understanding of how the global demand affects regret. Consequently, we plan to change the ratio between the global demand and the supply to simulate different workloads and to study how different algorithms perform. An adjustable parameter ***Demand-Supply Ratio*** α is introduced to facilitate the study of Q1. Using this ratio saves us from finding and justifying proper absolute values for demand or supply, which are highly dependent on multiple factors (e.g., the market share of the host) and vary from case to case. In addition, we introduce the ***Average-Individual Demand Ratio*** $p(\overline{I^{\mathcal{A}}})$ in order to answer Q2. Given a global demand, we can control the number of advertisers to adjust the individual demand of advertisers represented by $p(\overline{I^{\mathcal{A}}})$.

In the following, we first introduce the datasets, settings of the above two parameters, running environment and performance metrics; we then present the major experimental results and our insights gained under different real scenarios.



(a) Influence Distribution of Billboards **(b) Impression Count Distribution of Billboards**

Figure 1: Influence Distribution on NYC and SG

7.1 Experiment Setup

7.1.1 *Datasets.*

We carefully choose two real datasets, New York City (NYC) and Singapore (SG), as reported in Table 5. They enable us to study our problem when facing audiences from two representative transport modes in reality (i.e., taxis vs. buses, respectively). For NYC, same as [26, 27, 29], we crawled the billboard dataset from LAMAR [16], one of the largest OOH advertising companies worldwide. The trajectory dataset contains two million taxi trajectories from the publicly available TLC trip records [21]. For SG [20], we use the EZ-link (smart cards used in SG for cashless public transport) data to obtain trajectories and billboards. The trajectories are from bus stops to bus stops, and each bus stop is also the location of a billboard operated by JCDecaux [13].

We plot several unique features of these two datasets in Figure 1. Figure 1a lists the influence distribution of billboards: the x-axis presents different influences of billboards in descending order, and the y-axis reports the proportion of an influence over the maximal influence. Figure 1b reports the impression counts achieved by the set of selected billboards, where the impression count of a billboard set is the number of trajectories influenced by the billboards. We sort all the billboards in descending order of their influences and report the impression counts when a certain percentage of billboards are selected. Instead of reporting the exact values of the impression counts, we use the percentage (i.e., impression count/total trajectory count) for ease of illustration.

Observation. NYC has more high-influence billboards than SG, while trajectories that are influenced by these high-influence billboards are highly overlapping in NYC. That explains why the yellow curve in Figure 1b increases slower than the purple one.

7.1.2 *Billboard.*

Billboard Influence. Each billboard o has a location loc. The billboard influence can be measured in various ways, and we follow the same setting of the existing work [26, 27] that is briefly explained below. Each trajectory $t = \{p_1, \cdots, p_{|t|}\}$ is a set of points recording an audience's movement. A Bernoulli random variable $p(o, t)$ is used to denote the state whether a trajectory t meets a billboard o: $p(o, t) = 1$ iff $\exists p_i \in t$ such that $dist(t.p_i, o.loc) \leq \lambda$. Here, $dist(\cdot)$ is the Euclidean distance between p_i and $o.loc$, and λ is a given distance threshold. We denote the influence of a billboard o_i to a trajectory t_j as $I(o_i, t_j) = p(o_i, t_j)$. Accordingly, we denote the influence of a billboard set S_i to a trajectory t_j as $I(S_i, t_j)$: $I(S_i, t_j) = 1 - \prod_{o \in S_i}(1 - I(o, t_j))$. Finally, we define the influence of a billboard set S_i as the sum of influence from S_i to all trajectories: $I(S_i) = \sum_{t \in \mathcal{T}} I(S_i, t)$.

Billboard Cost. All companies such as LAMAR and JCDecaux do not provide the exact cost of billboards. As reported in the latest studies [26, 29], a billboard's cost is proportional to its influence, so we follow the same setting here: $o.w = \lfloor \tau \times I(o)/10 \rfloor$, where τ is a factor randomly chosen from 0.9 to 1.1 to simulate the fluctuation, and $I(o)$ is the number of trajectories influenced by a billboard o.

7.1.3 Key Parameters.

All key parameters are summarized in Table 6, including the demand-supply ratio α, the average-individual demand Ratio $p(\overline{I^{\mathcal{A}}})$, the unsatisfied penalty ratio γ used in Equation 1, and the distance threshold λ that determines the maximum distance in which a billboard could influence a trajectory. In each set of experiments, we vary only one parameter and set the remaining parameters to their default values (highlighted in bold).

Demand-Supply Ratio α. It refers to the proportion of the global demand over the host's supply, i.e., $\alpha = I^{\mathcal{A}}/I^*$, where $I^{\mathcal{A}} = \sum_{a \in \mathcal{A}} \mathcal{I}_i$ represents the global demand, and $I^* = \sum_{o \in \mathcal{U}} I(\{o\})$ is the host's supply. We simulate five different situations of α, i.e., **low, normal, high, full**, and **excessive** global demand. The corresponding α is set to 40%, 60%, 80%, 100% and 120%, respectively.

Average-Individual Demand Ratio $p(\overline{I^{\mathcal{A}}})$. It is the percentage of average individual demand of advertisers over the host's supply, i.e., $p(\overline{I^{\mathcal{A}}}) = \overline{I^{\mathcal{A}}}/I^*$, where $\overline{I^{\mathcal{A}}} = I^{\mathcal{A}}/|\mathcal{A}|$ is the average individual demand of advertisers. By controlling its value, we could adjust the demand of individual advertisers.

Advertiser's Demand $\mathcal{I}$. Once the Average-Individual Demand Ratio $p(\overline{I^{\mathcal{A}}})$ is fixed, the average demand of advertisers $\overline{I^{\mathcal{A}}}$ can be easily derived as $\overline{I^{\mathcal{A}}} = p(\overline{I^{\mathcal{A}}}) \cdot I^*$. For example, when $\alpha = 100\%$ and $p(\overline{I^{\mathcal{A}}}) = 1\%$, we will have 100 small advertisers with each having a low average individual demand equivalent to 1% of the supply; when $\alpha = 100\%$ and $p(\overline{I^{\mathcal{A}}}) = 20\%$, we will have five big advertisers, and each has a high average individual demand equivalent to 20% of the supply. Subsequently, we generate the demand of each advertiser based on $\mathcal{I}_i = \lfloor \omega \cdot I^* \cdot p(\overline{I^{\mathcal{A}}}) \rfloor$, where ω is a factor randomly chosen from 0.8 to 1.2 to simulate different demand of advertiser.

Advertiser's Payment L. Following a widely adopted setting in marketing studies [1, 2, 4], we set each advertiser's payment to be proportional to her demand: $L_i = \lfloor \epsilon \cdot \mathcal{I}_i \rfloor$, where ϵ is a factor randomly chosen from 0.9 to 1.1 to simulate a various payment.

Unsatisfied Penalty Ratio γ. Recall Equation 1, $\gamma \in [0, 1]$ controls the fraction of payment penalty when the advertiser is not satisfied. At one extreme (i.e., $\gamma = 0$), the host cannot receive any payment if the required influence is not satisfied. At the other extreme (i.e., $\gamma = 1$), the host can receive the fraction of payment as the same fraction of influence that has been satisfied (i.e., $I(S_i)/\mathcal{I}_i$).

7.1.4 Other Setups.

Experiment Environment. All codes are implemented in Java. Experiments are conducted on a machine with Intel Core i7-8750U CPU and 32GB memory running Windows 10.

Performance Metrics. The effectiveness metrics include the total regret, as well as the portion of excessive influence and unsatisfied penalty in composing the total regret. The efficiency metric is the running time, which is evaluated by the average result of five runs.

Table 6: Parameter Settings

Parameter	Values
α	40%, 60%, 80%, **100%**, 120%
$p(\overline{I^{\mathcal{A}}})$	1%, 2%, **5%**,10%, 20%
γ	0, 0.25, **0.5**, 0.75, 1
λ	50m, **100m**, 150m, 200m

Methods for Comparison. To our best knowledge, this is the first work studying how to minimize the total regret of the host in OOH advertising. We compared four methods proposed in this work.

(1) **G-Order**: The budget-effective greedy method (Section 5.1).
(2) **G-Global**: The synchronous greedy method (Section 5.2).
(3) **ALS**: The advertiser-driven local search method (Section 6.1).
(4) **BLS**: The billboard-driven local search method (Section 6.2).

7.2 Effectiveness Study

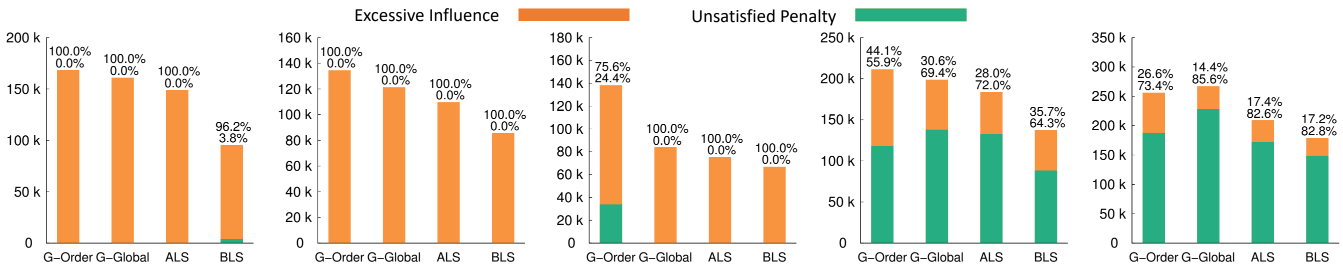
We first evaluate how the varying α and $p(\overline{I^{\mathcal{A}}})$ impact the regret. According to Equation 1, the regret consists of two components: *the unsatisfied penalty from the unsatisfied advertisers,* and *the excessive influence.* Hence, we report the total regret as a result of each experiment as well as both components of the total regret.

As stated previously, α and $p(\overline{I^{\mathcal{A}}})$ are introduced to answer the two key questions we asked at the beginning of the experiment section. To answer Q1, we vary α from 40% to 120%, corresponding to various demand-supply ratios (gradually from low to excessive global demand). When α is small, the global demand is low and all the advertisers can be satisfied; as α becomes much larger, the global demand approaches or even exceeds the supply and some of the advertisers will NOT be satisfied. To answer Q2, we vary $p(\overline{I^{\mathcal{A}}})$ from 1% to 20%, representing different individual demands from advertisers (gradually from low to high individual demand).

To ease our discussion, we **cluster α values into two categories** (i.e., low global demand vs. high global demand), and we also **cluster $p(\overline{I^{\mathcal{A}}})$ values into two categories** (i.e., low individual demand vs. high individual demand). Combining α and $p(\overline{I^{\mathcal{A}}})$, we have in total four different cases:

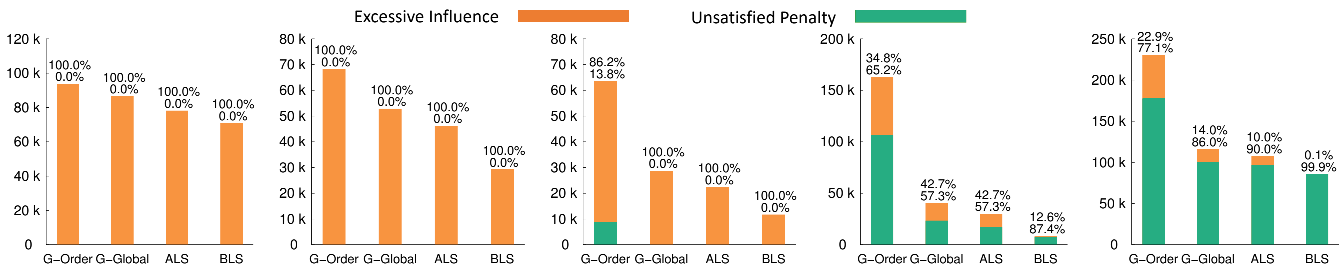
Individual \ Global	Low demand ($\alpha \le 80\%$)	High demand ($\alpha \ge 100\%$)
Low demand ($p(\overline{I^{\mathcal{A}}}) \le 2\%$)	Case 1	Case 3
High demand ($p(\overline{I^{\mathcal{A}}}) \ge 5\%$)	Case 2	Case 4

In the following, we will present the results on NYC dataset under the above four different cases. The findings on SG dataset are similar. Due to space limit, we only report the results of SG under the default settings. For the effectiveness study, we use stacked bars to represent the total regret, and use two numbers on top of each bar to indicate the percentage of excessive influence and that of unsatisfied penalty, respectively. It is worth noting that, the percentage of certain components could be zero in some cases (e.g., when all the advertisers are satisfied, there is *no* unsatisfied penalty). Therefore, some stacked bars may contain one, instead of two, segments.



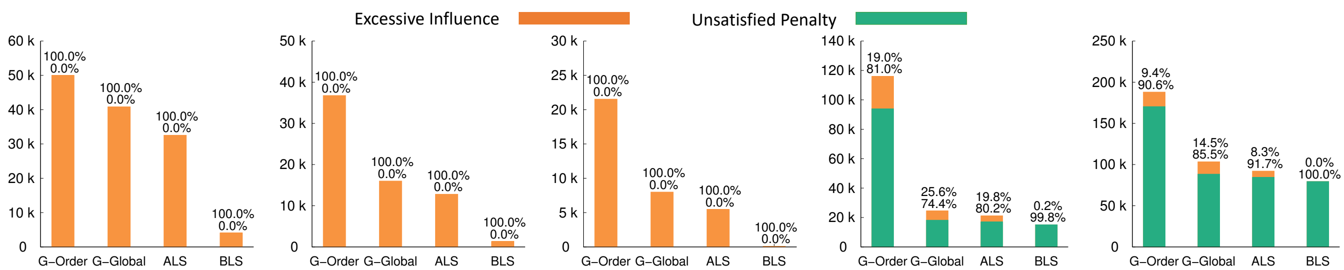
(a) $\alpha = 40\%$ **(b)** $\alpha = 60\%$ **(c)** $\alpha = 80\%$ **(d)** $\alpha = 100\%$ **(e)** $\alpha = 120\%$

Figure 2: Regret of varying the demand-supply ratio α when $p(\overline{I^{\mathcal{A}}}) = 1\%$ (NYC, $|\mathcal{A}| = 100$))



(a) $\alpha = 40\%$ **(b)** $\alpha = 60\%$ **(c)** $\alpha = 80\%$ **(d)** $\alpha = 100\%$ **(e)** $\alpha = 120\%$

Figure 3: Regret of varying the demand-supply ratio α when $p(\overline{I^{\mathcal{A}}}) = 2\%$ (NYC, $|\mathcal{A}| = 50$)



(a) $\alpha = 40\%$ **(b)** $\alpha = 60\%$ **(c)** $\alpha = 80\%$ **(d)** $\alpha = 100\%$ **(e)** $\alpha = 120\%$

Figure 4: Regret of varying the demand-supply ratio α when $p(\overline{I^{\mathcal{A}}}) = 5\%$ (NYC, $|\mathcal{A}| = 20$)

7.2.1 *Experiments over the NYC Dataset.*

Case 1: low α, low $p(\overline{I^{\mathcal{A}}})$ (parts (a)-(c) of Figures 2-4). Corresponding to Case 1, we have $\alpha \leq 80\% \wedge p(\overline{I^{\mathcal{A}}}) \leq 2\%$. This refers to the situation where both the global demand and the individual demand are low, e.g., a host has a small number of small advertisers. As α is small, the global demand is much smaller than the supply and all the advertisers are satisfied. Consequently, the regret consists of only excessive influence, except for G-Order and BLS. We have three main observations.

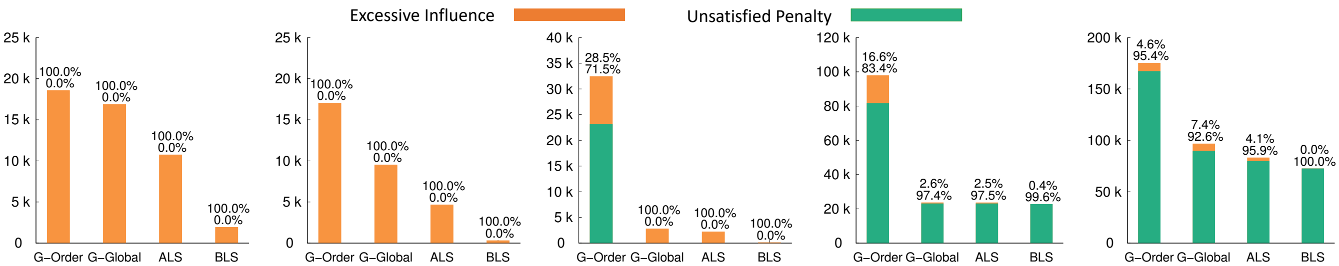
First, as α increases, the excessive influences of all algorithms decrease. This is because, when both α and $p(\overline{I^{\mathcal{A}}})$ are small, the required influence $\mathcal{I}$ of each advertiser is small. In contrast, billboards in NYC are those with high influences. Thus, it is easy to fulfill and exceed the demand and leads to a high excessive influence penalty.

Second, in most experiments, ALS and BLS control the excessive influence better. This is because ALS and BLS are able to satisfy all the advertisers with fewer billboards. ALS tries to exchange all billboards between advertisers based on the solution of G-Global, so it is able to achieve the regret that is equal to or smaller than that of G-Global. While all the advertisers can be satisfied, BLS actually explores more finer-grained exchanges to further reduce the gap between $I(S_i)$ and $\mathcal{I}_i$. Consequently, BLS averagely outperforms G-Order and G-Global by about 200% and 50%, respectively.

Third, in the extreme case (i.e., $\alpha = 40\%$ and $p(\overline{I^{\mathcal{A}}}) = 1\%$) where all the advertisers have very small demand, BLS will sacrifice some advertisers to achieve a smaller regret, if the excessive influence of satisfying advertisers is higher than the unsatisfied penalty.

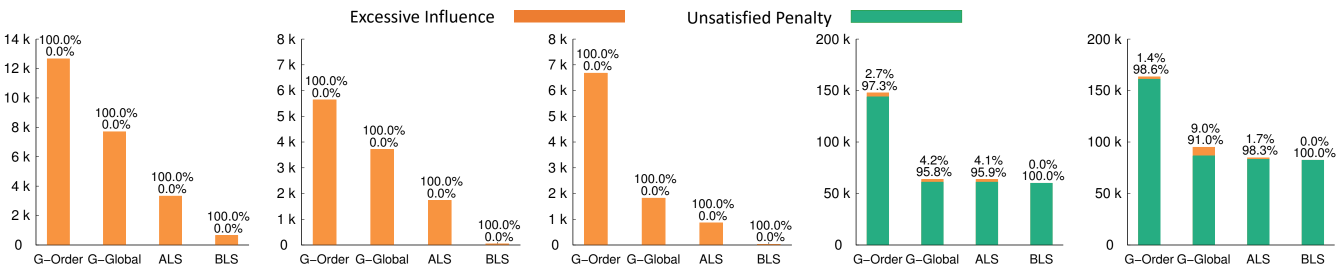
Case 2: low α, high $p(\overline{I^{\mathcal{A}}})$ (parts (a)-(c) of Figure 5 and Figure 6). Corresponding to Case 2, we have $\alpha \leq 80\% \wedge p(\overline{I^{\mathcal{A}}}) \geq 5\%$. The global demand is still lower than the supply but the individual demand is much higher (e.g., the host has a small number of big advertisers). We have two observations.

First, when the global demand remains low, as the individual demand increases, the excessive influence of all the algorithms drops (as compared with Case 1). This is because, with the increase of $p(\overline{I^{\mathcal{A}}})$, the number of advertisers decreases but the individual demand becomes higher. That is, the individual demand becomes



(a) $\alpha = 40\%$ **(b)** $\alpha = 60\%$ **(c)** $\alpha = 80\%$ **(d)** $\alpha = 100\%$ **(e)** $\alpha = 120\%$

Figure 5: Regret of varying the demand-supply ratio α when $p(\overline{I^{\mathcal{A}}}) = 10\%$ (NYC, $|\mathcal{A}| = 10$)



(a) $\alpha = 40\%$ **(b)** $\alpha = 60\%$ **(c)** $\alpha = 80\%$ **(d)** $\alpha = 100\%$ **(e)** $\alpha = 120\%$

Figure 6: Regret of varying the demand-supply ratio α when $p(\overline{I^{\mathcal{A}}}) = 20\%$ (NYC, $|\mathcal{A}| = 5$)

closer to (but does not exceed) the influence offered by billboards, and consequently the excessive influence of all the algorithms drops.

Second, with a higher individual demand, the host could actually deploy more billboards to each advertiser. Hence, BLS is able to explore more neighborhood search space by exchanging billboards. Consequently, BLS hugely outperforms G-Order and G-Global as it is able to reach almost zero regret.

Case 3: high α, low $p(\overline{I^{\mathcal{A}}})$ (parts (d) and (e) of Figures 2-4). Corresponding to Case 3, we have $\alpha \geq 100\% \wedge p(\overline{I^{\mathcal{A}}}) \leq 2\%$. This represents the situation where the global demand is very high (even exceeds the supply) but the individual demand is low, e.g., a host company has a very large advertiser base that is formed by small advertisers with low demand. We have made two observations.

First, given a very high global demand ($\alpha \geq 100\%$), none of the algorithms can satisfy all the advertisers. Hence, the unsatisfied penalty becomes a major component of the total regret. When $\alpha = 100\%$, the global demand is equal to the host's supply. However, since (1) the excessive influence cannot be diminished and (2) multiple billboards may influence the same trajectory, not all advertisers are satisfied. When $\alpha = 120\%$, since the demand exceeds the supply, the unsatisfied penalty becomes even larger.

Second, the advantages of ALS and BLS (especially BLS) become more significant when compared with other approaches in terms of reducing the total regret. It is observed that ALS and BLS are able to satisfy more advertisers and hence suffer from a smaller unsatisfied penalty. This observation also highlights the importance of a proper deployment strategy when the global demand becomes very high. As the host does not have much available supply, the allocation of a billboard to one advertiser actually increases the risk of missing another advertiser and hence each allocation is critical. Obviously, BLS handles this risk the best. It averagely outperforms G-Order and G-Global by about four times and one time, respectively.

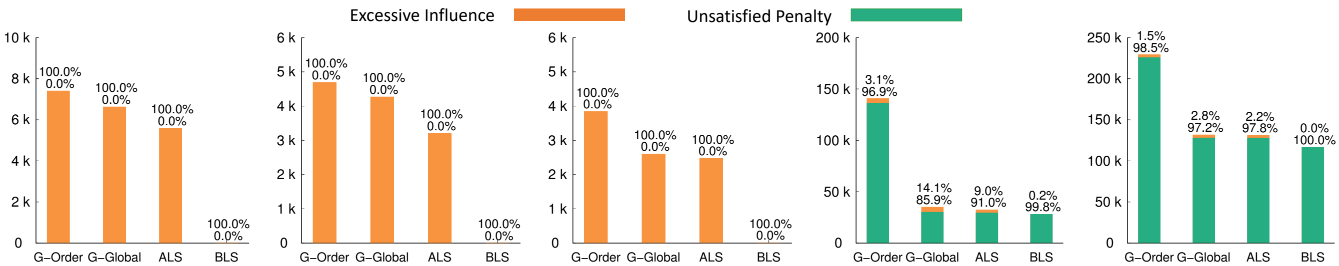
Case 4: high α, high $p(\overline{I^{\mathcal{A}}})$ (parts (d) and (e) of Figure 5 and Figure 6). Corresponding to Case 4, we have $\alpha \geq 100\% \wedge p(\overline{I^{\mathcal{A}}}) \geq 5\%$, representing the situation where both the global demand and the individual demand are very high (e.g., the host is over demanded by a small number of big advertisers). We have three observations.

First, with large $p(\overline{I^{\mathcal{A}}})$ and large α, every unsatisfied advertiser will lead to a high regret. Therefore, all the algorithms suffer from large regrets. Thus, the advantage of ALS and BLS becomes less significant, especially as compared to G-Global.

Second, due to the reason that we mentioned in Section 5.2, G-Order has a much higher unsatisfied penalty than others. Between ALS and BLS, the latter reaches a smaller excessive influence. This is consistent with our expectation as BLS adopts a finer-grained strategy when exploring different deployment plans. Consequently, BLS outperforms G-Global by about three times.

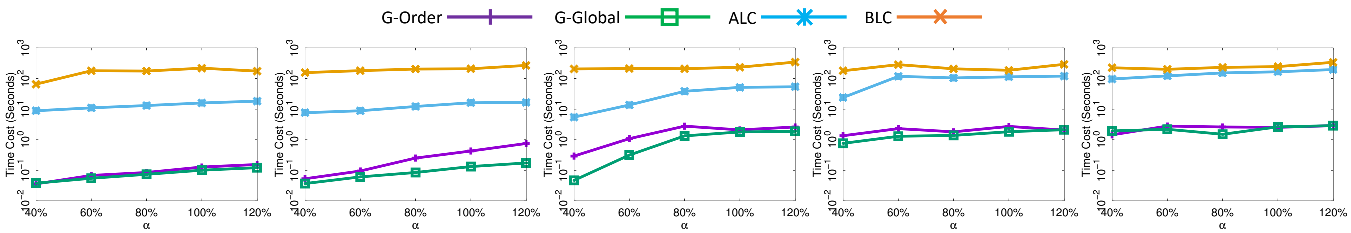
Third, when $p(\overline{I^{\mathcal{A}}})$ increases from 5% to 20%, the total regret becomes much larger, as the individual demand becomes higher. In other words, when the global demand is high, having big advertisers with high individual demand is less beneficial to the host, in terms of regret. This observation is consistent with the second observation we made under Case 2. When the host's supply is sufficient, it is more flexible to have a large number of small advertisers as both the risk and the penalty of missing one advertiser are much smaller.

Revisit Q1 and Q2. Based on the observations we have made from the above four cases, we are able to answer the two key questions Q1 and Q2. (1) When the global demand is low, the total regret is dominated by excessive influence. Under this situation, the host needs to choose advertisers carefully. Based on our experiment, with a high variance of billboards' influence and when the advertisers' average demand is only three times larger than the average



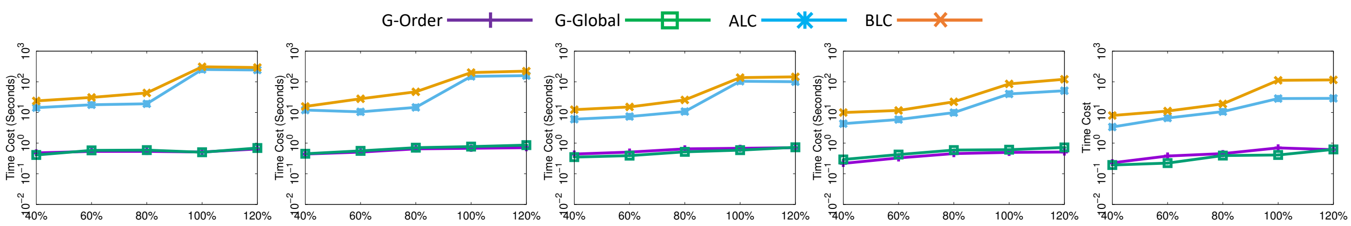
(a) $\alpha = 40\%$ **(b)** $\alpha = 60\%$ **(c)** $\alpha = 80\%$ **(d)** $\alpha = 100\%$ **(e)** $\alpha = 120\%$

Figure 7: Regret of varying the demand-supply ratio α when $p(\overline{I^{\mathcal{A}}}) = 5\%$ (SG, $|\mathcal{A}| = 20$)



(a) $p(\overline{I^{\mathcal{A}}}) = 1\%$ $(|\mathcal{A}| = 100)$ **(b)** $p(\overline{I^{\mathcal{A}}}) = 2\%$ $(|\mathcal{A}| = 50)$ **(c)** $p(\overline{I^{\mathcal{A}}}) = 5\%$ $(|\mathcal{A}| = 20)$ **(d)** $p(\overline{I^{\mathcal{A}}}) = 10\%$ $(|\mathcal{A}| = 10)$ **(e)** $p(\overline{I^{\mathcal{A}}}) = 20\%$ $(|\mathcal{A}| = 5)$

Figure 8: Efficiency Study (SG)



(a) $p(\overline{I^{\mathcal{A}}}) = 1\%$ $(|\mathcal{A}| = 100)$ **(b)** $p(\overline{I^{\mathcal{A}}}) = 2\%$ $(|\mathcal{A}| = 50)$ **(c)** $p(\overline{I^{\mathcal{A}}}) = 5\%$ $(|\mathcal{A}| = 20)$ **(d)** $p(\overline{I^{\mathcal{A}}}) = 10\%$ $(|\mathcal{A}| = 10)$ **(e)** $p(\overline{I^{\mathcal{A}}}) = 20\%$ $(|\mathcal{A}| = 5)$

Figure 9: Efficiency Study (NYC)

influence of billboards (i.e., Figure 2a), there is a high risk of having a great excessive influence; when the average demand is more than ten times of the average influence, it is easy to control the excessive influence. (2) When the global demand becomes much higher or even exceeds the supply, the total regret is dominated by the unsatisfied penalty. Consequently, having a large number of medium-demand advertisers is an ideal balance, as it provides more flexibility when deploying billboards to the advertiser; meanwhile, the penalty of missing one huge advertiser is much lower.

7.2.2 *Experiments over the SG Dataset.*

Figure 7 shows the results over SG. As reported in Figure 1, it is worth noting that: (1) SG has more low-influence billboards and the average influence in SG is smaller; (2) compared with NYC, the influences of billboards in SG are more uniform, and the influence overlaps among billboards are smaller since bus stations are sparse.

We have made two main observations. First, the results of SG are similar to the results of NYC, while the proportions of excessive influence of all algorithms are smaller. The main reason is that the smaller influence of billboards with less overlaps helps the host to efficiently and accurately deploy billboards.

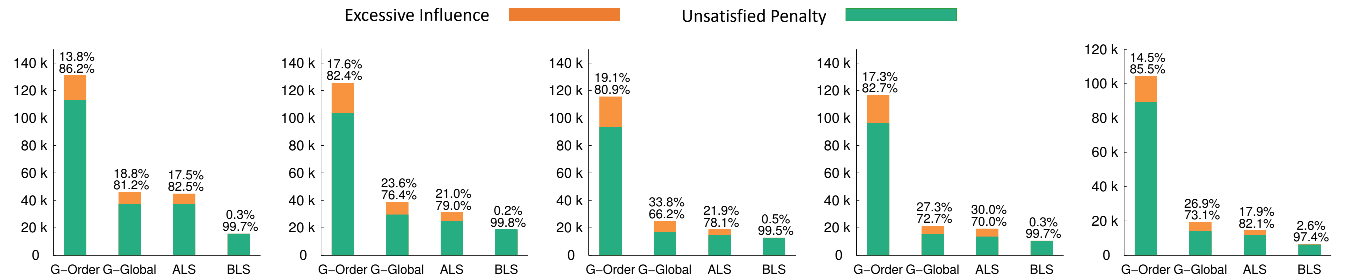
Second, because of a larger number of billboards and a smaller average billboard influence, BLS can explore more possible exchanges between billboards, and hence achieve a finer-grained swapping strategy. Consequently, following the same trend as in NYC, BLS can effectively reduce excessive influence for SG.

7.3 Efficiency Study

The efficiency is important, since a host who owns more than thousands of billboards in a city such as Juping or JCDecaux may have new advertisers every day. Hence, we conduct the efficiency evaluation under various cases of global demand and individual demand, by varying the settings of α and $p(\overline{I^{\mathcal{A}}})$. The results are reported in Figures 8 and 9. We have two main observations.

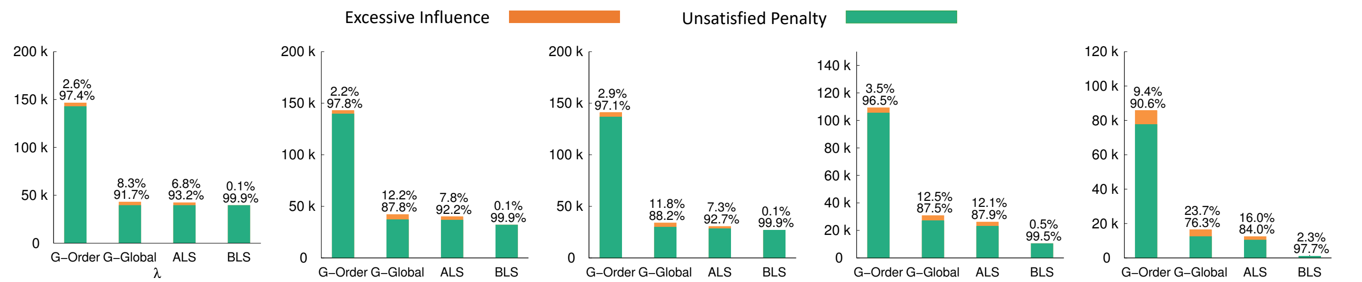
First, both G-Order and G-Global incur a much lower time cost compared to ALS and BLS. This is because both ALS and BLS employ the billboard deployment plan generated by G-Global as the initial plan and try to improve the effectiveness of the plan by exploring the swap of billboards/advertisers. Consequently, G-Order and G-Global provide a trade-off between the effectiveness of the billboards deployment plan and the time required to find the plan.

Second, with an increase of α, all the algorithms require longer search time. The reason is that, when α increases, more billboards need to be deployed to each advertiser. In addition, the number of unsatisfied advertisers increases together with the increase of



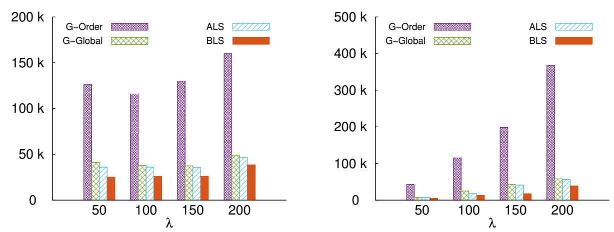
(a) $\gamma = 0$ (b) $\gamma = 0.25$ (c) $\gamma = 0.50$ (d) $\gamma = 0.75$ (e) $\gamma = 1$

Figure 10: Regret of varying the unsatisfied penalty ratio γ (NYC)



(a) $\gamma = 0$ (c) $\gamma = 0.25$ (c) $\gamma = 0.5$ (d) $\gamma = 0.75$ (e) $\gamma = 1$

Figure 11: Regret of varying the unsatisfied penalty ratio γ (SG)



(a) SG (b) NYC

Figure 12: Regret of varying λ

α. Accordingly, ALS and BLS need to execute more iterations to explore alternative deployment plans.

7.4 Other Parameter Study

The Impact of the Influence Range λ. We follow the same setting of the existing work [26, 27, 29] to define λ as the influence range of a billboard. The influence range is modeled as a circle centered on a billboard with a radius of λ meters. An audience can be influenced if she passes through this circle. In the following, we study the impact of varying λ, with the results reported in Figure 12.

We observe that the result of NYC is different from that of SG. In NYC, with an increase of λ, the regret of all algorithms increases. The reason is that when λ increases, billboards can influence more trajectories and the host's supply I^* increases accordingly. While increasing I^* and fixing α and $p(\overline{I^{\mathcal{A}}})$, $\mathcal{I}$ and $I^{\mathcal{A}}$ will increase. However, since $I^{\mathcal{A}}$ and I^* increase, the regret increases proportionally. On the other hand, in SG, when $\lambda \leq 150$, the impact is minor. This is because the billboards are placed at bus stations, and each billboard can only influence audiences of the buses that make a stop at the bus stations. As a result, a fixed group of trajectories are influenced regardless of the choice of λ. It worth noting that the regrets increase when $\lambda = 200$. A possible reason is that, some bus stations are close to intersections. With a larger λ, more trajectories are influenced by the billboards located at the intersections.

The Impact of the Unsatisfied Penalty Ratio γ. We study the impact of varying γ and report the results in Figures 10 and 11. We observe that, when γ increases, the regret of all algorithms drops. Recall Equation 1, γ controls the fraction of payment penalty when the advertiser is not satisfied. Given a smaller γ, the host suffers from a higher payment penalty. On the other hand, when $\gamma = 1$, the host can receive the fraction of payment as the same fraction of influence that has been satisfied. Taking Figure 11 (e) as an example, BLS almost meets the demand of all advertisers.

8 CONCLUSION

In this paper we proposed and studied MROAM, aiming to minimize the regret of the influence provider when dealing with numerous influence purchasers. We proved that it is NP-hard to approximate within any constant factor. Then, we proposed a randomized local search framework with different neighborhood search strategies, and proved that one achieves an approximation factor to a dual problem of MROAM. Our methods can work with any choice of straightforward influence models. Lastly, we conducted extensive experiments on two real-world datasets to verify the efficiency and the empirical effects of our methods.

ACKNOWLEDGMENTS

This research is supported in part by the Australian Research Council (DP200102611), the Ministry of Education, Singapore, under its AcRF Tier 2 Funding (MOE2019-T2-2-116), and the Lee Kong Chian fellowship. H. V. Jagadish was supported in part by NSF grants 1741022 and 1934565.

REFERENCES

[1] Çigdem Aslay, Francesco Bonchi, Laks V. S. Lakshmanan, and Wei Lu. 2017. Revenue Maximization in Incentivized Social Advertising. *PVLDB* 10, 11 (2017), 1238–1249.

[2] Çigdem Aslay, Wei Lu, Francesco Bonchi, Amit Goyal, and Laks V. S. Lakshmanan. 2015. Viral Marketing Meets Social Advertising: Ad Allocation with Minimum Regret. *PVLDB* 8, 7 (2015), 822–833.

[3] BAIDU. [n.d.]. *JUPING.* https://juping.baidu.com.

[4] Prithu Banerjee, Wei Chen, and Laks V. S. Lakshmanan. 2019. Maximizing Welfare in Social Networks under A Utility Driven Influence Diffusion model. In *SIGMOD*. ACM, 1078–1095.

[5] Nicola Barbieri, Francesco Bonchi, and Giuseppe Manco. 2012. Topic-Aware Social Influence Propagation Models. In *ICDE*. IEEE, 81–90.

[6] Song Bian, Qintian Guo, Sibo Wang, and Jeffrey Xu Yu. 2020. Efficient Algorithms for Budgeted Influence Maximization on Massive Social Networks. *Proc. VLDB Endow.* 13, 9 (2020), 1498–1510.

[7] Facebook. [n.d.]. *Cost per post engagement.* https://en-gb.facebook.com/business/help/1514627528773502.

[8] M. R. Garey and David S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman.

[9] Qintian Guo, Sibo Wang, Zhewei Wei, and Ming Chen. 2020. Influence Maximization Revisited: Efficient Reverse Reachable Set Generation with Bound Tightened. In *SIGMOD*. ACM, 2167–2181.

[10] Tao Guo, Kaiyu Feng, Gao Cong, and Zhifeng Bao. 2018. Efficient Selection of Geospatial Data on Maps for Interactive and Visualized Exploration. In *Proceedings of the 2018 International Conference on Management of Data, SIGMOD Conference 2018, Houston, TX, USA, June 10-15, 2018*, Gautam Das, Christopher M. Jermaine, and Philip A. Bernstein (Eds.). ACM, 567–582.

[11] Kai Han, Keke Huang, Xiaokui Xiao, Jing Tang, Aixin Sun, and Xueyan Tang. 2018. Efficient Algorithms for Adaptive Influence Maximization. *Proc. VLDB Endow.* 11, 9 (2018), 1029–1040.

[12] Brian Haven. 2007. Marketing's new key metric: engagement. *Marketing* (2007), 1–15.

[13] JCDecaux. [n.d.]. *JCDecaux.* https://www.jcdecaux.com.sg/streetside/engaging-solutions.

[14] David Kempe, Jon M. Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *SIGKDD*. ACM, 137–146.

[15] Juntao Lai, Tao Cheng, and Guy Lansley. 2017. Improved targeted outdoor advertising based on geotagged social media data. *Annals of GIS* 23, 4 (2017), 237–250.

[16] LAMAR. [n.d.]. *LAMAR.* https://www.lamar.com.

[17] Yuchen Li, Ju Fan, Yanhao Wang, and Kian-Lee Tan. 2018. Influence Maximization on Social Graphs: A Survey. *IEEE Trans. Knowl. Data Eng.* 30, 10 (2018), 1852–1872.

[18] Dongyu Liu, Di Weng, Yuhong Li, Jie Bao, Yu Zheng, Huamin Qu, and Yingcai Wu. 2017. SmartAdP: Visual Analytics of Large-scale Taxi Trajectories for Selecting Billboard Locations. *IEEE Trans. Vis. Comput. Graph.* 23, 1 (2017), 1–10.

[19] Reza Lotfi, Yahia Zare Mehrjerdi, and Nooshin Mardani. 2017. A multi-objective and multi-product advertising billboard location model with attraction factor mathematical modeling and solutions. *IJAL* 7, 1 (2017), 64–86.

[20] Xiancai Tian and Baihua Zheng. 2018. Using Smart Card Data to Model Commuters' Responses Upon Unexpected Train Delays. In *IEEE Big Data*. 831–840.

[21] TLC. [n.d.]. *Taxi and Limousine Commission.* http://www.nyc.gov/html/tlc.

[22] China Tower. [n.d.]. *Telecommunications Wower Infrastructure Service Provider.* https://ir.china-tower.com/en/about/profile.php.

[23] Hongyang Wang, Qingfei Meng, Ju Fan, Yuchen Li, Laizhong Cui, Xiaoman Zhao, Chong Peng, Gong Chen, and Xiaoyong Du. 2020. Social Influence Does Matter: User Action Prediction for In-Feed Advertising. In *AAAI*. 246–253.

[24] Liang Wang, Zhiwen Yu, Dingqi Yang, Huadong Ma, and Hao Sheng. 2020. Efficiently Targeted Billboard Advertising Using Crowdsensing Vehicle Trajectory Data. *IEEE Trans. Ind. Informatics* 16, 2 (2020), 1058–1066.

[25] Sheng Wang, Zhifeng Bao, J Shane Culpepper, and Gao Cong. 2021. A Survey on Trajectory Data Management, Analytics, and Learning. *ACM Computing Surveys (CSUR)* 54, 2 (2021), 1–36.

[26] Ping Zhang, Zhifeng Bao, Yuchen Li, Guoliang Li, Yipeng Zhang, and Zhiyong Peng. 2018. Trajectory-driven Influential Billboard Placement. In *SIGKDD*. ACM, 2748–2757.

[27] Ping Zhang, Zhifeng Bao, Yuchen Li, Guoliang Li, Yipeng Zhang, and Zhiyong Peng. 2020. Towards an Optimal Outdoor Advertising Placement: When a Budget Constraint Meets Moving Trajectories. *ACM Trans. Knowl. Discov. Data* 14, 5, Article 51 (2020), 32 pages. https://doi.org/10.1145/3350488

[28] Yipeng Zhang, Zhifeng Bao, Songsong Mo, Yuchen Li, and Yanghao Zhou. 2019. ITAA: An Intelligent Trajectory-driven Outdoor Advertising Deployment Assistant. *Proc. VLDB Endow.* 12, 12 (2019), 1790–1793.

[29] Yipeng Zhang, Yuchen Li, Zhifeng Bao, Songsong Mo, and Ping Zhang. 2019. Optimizing Impression Counts for Outdoor Advertising. In *SIGKDD*. ACM, 1205–1215.