Efficient Content Distribution in DOOH Advertising Networks Exploiting Urban Geo-Social Connectivity

Fang-Zhou Jiang*, Kanchana Thilakarathna*, Mahbub Hassan*,
Yusheng Ji†, Aruna Seneviratne*
*Data61, CSIRO & UNSW, Australia, firstname.lastname@data61.csiro.au,
†National Institute of Informatics, Japan, kei@nii.ac.jp

ABSTRACT

Digital out-of-home (DOOH) advertising networks, comprised of pervasive distributed digital signages (screens), are rapidly growing. It is reported that more than 70% of DOOH revenue comes from local ads, while it is especially challenging to decide when and where to deliver the most suitable ad due to the spatio-temporal dynamics of human mobility and preferences. Understanding urban geo-social connectivity in terms of people movement would greatly benefit ad content distribution, and could potentially be utilized by a large number of mobile applications and geo-social services. However, existing DOOH ad distribution systems are designed to target individuals, which might not be the best choice in public spaces, and do not consider the preferences of "cohort of users". In this paper, we propose an alternative approach to target cohort of users extracting urban geo-social connectivity through large-scale mobile network data and existing geo-social service data. We construct a dynamic urban geo-social connectivity graph, and formulate the problem of distributing ads for maximum exposure to the "right" users under a constrained budget. Hence, we propose a heuristic algorithm. Simulation results show that our system targeting "cohort of users" achieves a maximum 300% improvement compared to naive distributing method in displaying ads to the "right people" when user preferences are completely known, while a minimum of 25% improvement when the knowledge of user preferences is limited.

Keywords

Spatio-Temporal Dynamics; Digital out-of-home Network; Content Distribution

1. INTRODUCTION

Digital signage, as a form of an electronic display, is becoming ubiquitous and is a very important form of out-of-home (OOH) advertising in public places. Almost to be found everywhere, digital out-of-home (DOOH) advertising expanded rapidly. From 2009 to 2014, it grew at a rate of approximately 30% [7], representing over 40% of total OOH industry in the year 2015 [8]. The pervasiveness of DOOH is partially due to its advantage of displaying multimedia content with more flexibility and has the ability to adapt to different contexts [4]. However, the challenge has been to reach the "right people" with the "right ads" and delivering ads at the "right time" [21, 24]. Our hypothesis is the spatial and temporal dynamics of urban users can be exploited to deliver "right ads" to "cohort of users". Our work attempts to bridge the gap between urban social connectivity and better ad delivery in physical web leveraging users' location information.

Advertisers have attempted to apply the successful model of online personalized ads to DOOH. This approach has its limitations in public space especially related to privacy [13]. Delivering ad content based on "cohort of user" interest has attracted some attentions [13], but surprisingly, do not take into account the spatial and temporal dynamics of user preferences. Currently, the approach is to distribute content via commercialized content management systems [20, 5] via numerous user information collected by advertisers. These systems generally consist of ad generators, ad servers and digital screens. The advertisers send digital content to content servers for scheduling content on the digital screens. However, despite studies [23] in the area of urban modeling, they do not take into account the spatial and temporal dynamics of user preferences. Furthermore, as the digital screen is design to reach large number of people, whether and how personalized ads in DOOH would improve efficiency is still not answered.

The pervasive use of mobile devices enables us to have a much finer and complete understanding of the user dynamics and their interest than traditional methods such as customer surveys. In this paper, we attempt to answer the question whether it is possible to better serve content to given set of digital screens by leveraging urban geo-social connectivity graph built via large-scale mobile user mobility data. We first analyze a large-scale dataset collected by mobile network provider, and show the strong spatio-temporal dynamics of users. We then model urban connectivity as cohort of user movement in a dynamic graph. In addition, an optimization problem of where to distribute a given ad at a given time is formulated assuming full knowledge of user preference. We further propose a heuristic algorithm exploiting spatio-temporal user correlation and spatial similarity when knowledge about users is limited. Finally, we evaluate the improvement of targeting "cohort of users" with

WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3051156

real-world data-driven simulation, and the results shows up to 300% improvement in ad eyeballing by the "right people" compared to distribution scheme that does not take into account dynamics of "cohort of users" preferences. Our constructed fine-granularity spatial similarity graph in urban area could potentially be used for other services and applications such as city planning and location-based recommendation systems. To the best of our knowledge, this is the first work attempting to improve DOOH content distribution efficiency leveraging spatio-temporal dynamics of users and urban geo-social connectivity, and being evaluated with large-scale mobile dataset. This paper makes the following contributions;

- We analyze fine-grained spatio-temporal dynamics of urban users with large-scale datasets, and propose a method of describing the user movement in form of geo-social connectivity graph.
- We formulate ad distribution in DOOH as an optimization problem, and propose a novel spatial similarity based heuristic algorithm that provide the best solution possible when knowledge about spatial user preferences are limited.
- We evaluate both the idea of targeting "cohort of user" and proposed heuristic algorithm with data-driven simulation, and show an up to 3 times improvement in eyeballing by people who are interested.
- Finally, we identify our work's theoretical and practical implications as well as potential applications and services.

The rest of paper is organized as follows; We first present the sets of data used in Section 2. In Section 3, we analyze the spatio-temporal dynamics of the collected dataset. We then formulate an optimization problem and describe the solutions in Section 4-5. Section 6 evaluates our hypothesis and proposed heuristic algorithm, and Section 7 summarizes the related work. Finally, Section 8 provides our discussion and Section 9 concludes the paper.

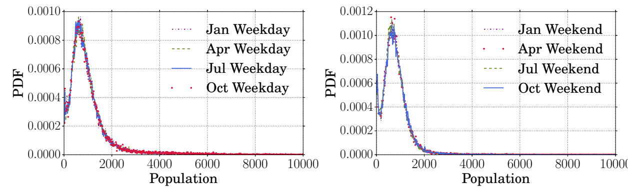
2. SETS OF DATA

The following datasets are used in this paper to gain insights about geo-social connectivity of urban population. We further study the effect of targeting "cohort of users" with simulation driven by these datasets.

2.1 Mobile Urban Dynamics Dataset

We use population dynamic of a city derived from number of mobile devices connected to mobile base stations collected by NTT DoCoMo[1], the largest Mobile Network Operator in Japan. In the first dataset (*DS1*), the city, where the data was collected, is split into 9,367 250m*250m inner city grids in urban areas and 7,160 500m*500m grids for the rural areas. The data covers morning (5am), noon (12pm) and evening (6pm) snapshot on both weekdays and weekends across 4 seasons. The registered home addresses of users are also available, e.g., 20 people currently at grid i come from suburb ID j. Additionally, data is available for a hourly snapshots over 2 days (48hrs) of four hot-spots (2km radius each includes over 150 grids) for fine-grained temporal analysis. We denote this dataset as *DS2*.

[1]www.nttdocomo.co.jp/english/



(a) Weekdays (b) Weekends

Figure 1: Seasonal Spatial Population Distribution

2.2 Yelp Dataset

We obtained business names, rating, reviews and geo-locations of 499,738 businesses in the same geographical area by crawling the popular business review website Yelp[2]. The dataset (*DS3*) provide us further insight into characteristics of fine-grained geographical areas.

3. SPATIO-TEMPORAL DYNAMICS

In this section, we present the basics characteristics of the datasets both spatially and temporally, and show that despite the observed strong dynamics, there are certain spatio-temporal patterns of "cohort of user" mobility and behavior that could be exploited to improve content distribution efficiency.

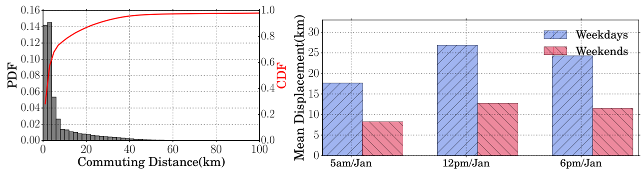
3.1 Spatial Dynamics and Characteristics

In Figure 1, we first show spatial distribution of population on both weekdays and weekends across different seasons in a year, where x-axis represents population per grid. It could be seen that spatial distribution differs on weekdays and weekends, while seasonal variation is subtle. In addition, a higher proportion of grids with population density per grid over 2000 on weekdays is observed. Although the PDF distribution show little seasonal variation, it does not give a full picture of fine-grained spatio-temporal dynamics. We then compare detailed spatial and temporal urban population dynamics on both weekdays and weekends in a selected month. Overall, spatial distribution in the morning is identical on both weekday and weekends. We suspect that is due to most people still sleep at home during that time, which presents a baseline distribution of city population. Moreover, we observe a clear human flow from suburbs into central CBD (center circle) on weekdays from 5am to 12pm, while a similar trend is not obvious on weekends. We also verified the diurnal pattern during the other times of year and did not observe significant difference.

Furthermore, we use Google Map API[3] to decode the 1804 suburb ID in *DS1* to geo-location with latitude and longitude (generally center of the suburb). The home locations along with users current locations could be used to find the physical distance that people travel away from home and urban geo-social connections. Figure 2a illustrates the PDF and CDF distributions of away from home distance. It could be seen that 90% of people are within 20km away from their home at all times. Furthermore, very few people travel over 60km away from home. Figure 2b further examines the temporal dynamics of away from home distance in a bar chart for the month January. As expected, people are generally closer to their homes on weekends than weekdays. Moreover, we observe a highly dynamic spatial variation across

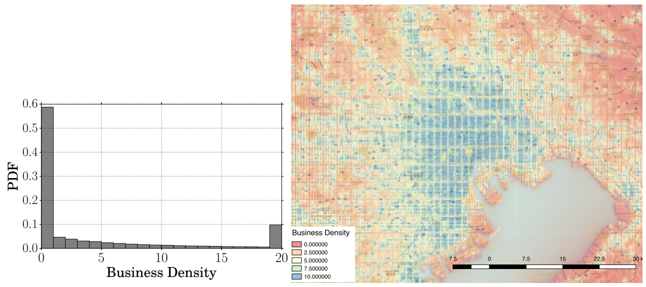
[2]www.yelp.com

[3]https://maps.googleapis.com/



(a) Distribution (b) Temporal Dynamics

Figure 2: Away from Home Distance



(a) Distribution (b) Tokyo Bay Area

Figure 3: Business Characteristics

the city even at the same time. Again, as overall statistics does not tell the full story of fine-grained spatio-temporal dynamics, we further study urban spatial characteristics.

We show urban geographical characteristics in Figure 3, where business density is used to describe business activity in a given area. Figure 3a shows the PDF distribution of business density of the whole city, while Figure 3b display the density map. Land usage of certain area, i.e. residential or business, could potentially be identified by *DS3*. In this paper, we mainly use business density as a feature, however, more features could be collected to further describe characteristics of an area, i.e. social economy [11].

3.2 Temporal Dynamics

We compare the temporal dynamics of four selected hotspots during a period of 48 hours in Figure 4 using finer-grained temporal dataset *DS2*. These hot spots represents exhibition center, entertainment center, inter-city and local train station respectively. From Figure 4a to Figure 4d, hourly box-plots of spatial distributions of population in these hot-spots are shown. The box plot shows the median, 25 and 75 percentile of population among the interested area. Along with the box plot, we also include the temporal dynamics of population spatial entropy. We define S_n^t as spatial population entropy among n grids at time t.

$$S_n^t = -\sum_{i=1}^{n} \frac{p_i^t log_2(p_i^t)}{log_2(n)}, \text{where } p_i^t = \frac{N_i^t}{\sum_{j=1}^{j=n} N_j^t} \quad (1)$$

N_j^t represents the number of people in grid j at time t. An entropy S_n close to 0 indicates that the distribution is extremely skewed, while close to 1 represents a more even distribution. The dynamic entropy change reflects the sparseness variation spatially, and gives us a better understanding of the skewness of distribution in different areas. Furthermore, all selected hot-spots display an clear diurnal pattern both in boxplot and entropy, although the peak of sparseness occurs at different time of the day depending on the function of the area. Despite the strong spatio-temporal fluctuation of entropy, the value is generally over 0.9. This indicates that only limited grids are very different to the remaining ones in the same area, suggesting the existence of spatial correlation for the neighbouring areas.

We then focus more on temporal dynamic of a individual grid. We define the rate of change of grid i at time t, K_i^t, as the percentage change of population compared to the previous observation. (in our case the previous hour)

$$K_i^t = \frac{N_i^t - N_i^{t-1}}{N_i^{t-1}} \quad (2)$$

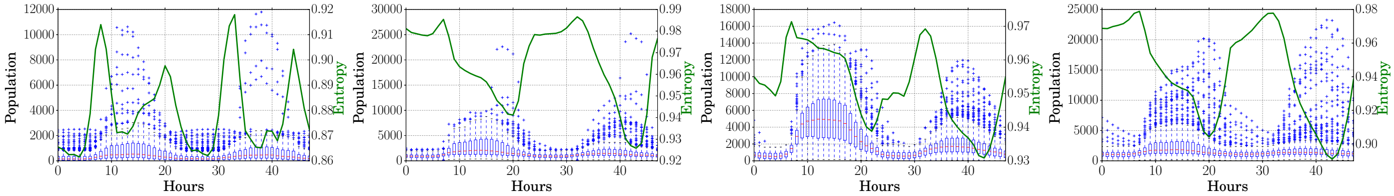
From Figure 4e to Figure 4h, interestingly, some grids present strong daily dynamic patterns (i.e. some grids at Location 3), while other grids does not change significantly throughout the day. More specifically, the rate of change swings significantly higher in the inter-city station, while it is quite stable in most areas of a local major station. The temporal evolution in these selected hot spots gives us a better understanding of the fine grained spatio-temporal dynamics of people.

We constructed a geo-social connectivity graph (details in Section 5), and present some initial findings in this paragraph. Figure 5a further illustrates the CDF of diversity (degree of connectivity) with regards to the number of suburbs that people are from in a given area. Location 1 as an exhibition center, is surprisingly less diverse than the overall Tokyo baseline. Moreover, Location 3 as an inter-city train station is more diverse, with more than triple higher diversity compared to the overall city level observation. A stronger diversity along with higher dynamics could potentially indicate an area more suitable for certain type of ad content. We further study the characteristics and dynamics of connectivity in the following subsection. Moreover, intuitively, popular nodes attract people from various locations. Figure 5b shows the degree of connectivity of the whole city. Degree of connectivity among ranked nodes approximately follows power law distribution, i.e. zipf (linear in log-log plot), which suggests that small number of very popular nodes are accounted for a majority of connections.

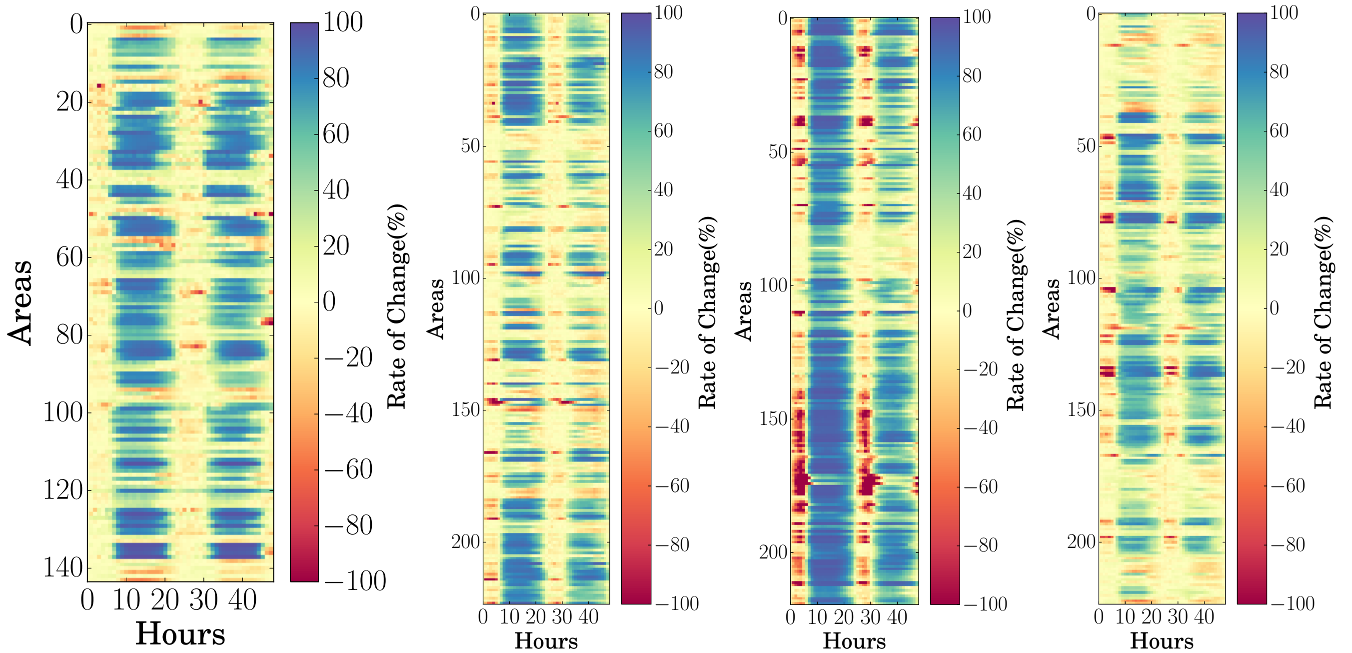
Lastly, we investigate how a certain connection evolve over time, and attempts to understand if the evolution is predictable. We identify an unique connection by ID number of both grid node and home node. We first measure how significant the sets of connections evolve overtime by Jaccard similarity. In total, we observe approximately 1M (992,957) unique connections, totally 4.9M (counting repeated ones) over time. Figure 6 shows the Jaccard similarity of connection sets of different time of day. Quite surprisingly, the pair wise similarity of connection sets at 5am on different days is lower than other times of the day. 12pm seems to be the most stable time of day, with around 40% of shared connections over different days. In addition, weekdays are more similar than weekends, with 5am being an exception.

4. EFFICIENT CONTENT DISTRIBUTION IN DOOH NETWORK

In existing DOOH ad content distribution systems, advertisers need to decide when and where to display the "right" ads for the "right" people. Therefore, advertisers are required to either bid or come up with a set of locations when they wish to distribute certain ads. To the best of our knowledge, there has been no work optimizing ads distribution for "cohort of users" in DOOH considering the spatio-temporal dynamics of users. We attempt to solve the problem exploit-

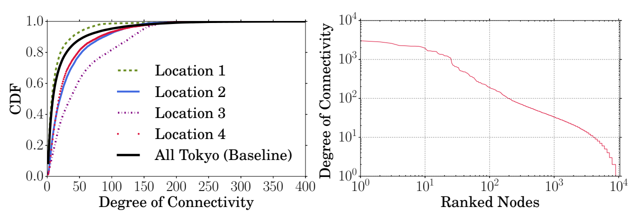


(a) Location 1(exhibition center) (b) Location 2(entertainment) (c) Location 3(inter-city station) (d) Location 4(local station)



(e) Location 1 (f) Location 2 (g) Location 3 (h) Location 4

Figure 4: Temporal Evolution in Selected Hot Spots



(a) Spatial Diversity (b) Spatial Connectivity

Figure 5: Spatial Distribution

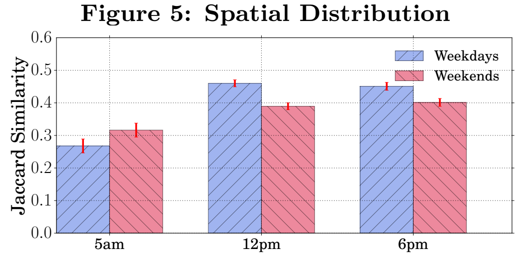


Figure 6: Jaccard Similarity of Connections

ing urban geo-social connectivity constructed from mobile users. The assumption we have made is that there exists spatial correlation of user preference based on local community. For example, people living in city CBD are expected to have different ad preference as people who are living in rural areas. This assumption is partially backed by the spatio-temporal correlation of user interest observed at mobile edge [10] and the spatial-temporal dynamics pattern we showed in Section 3. Thus, we consider the scenario that an advertiser needs to determine a set of locations at each time window t to push their ad content.

5. PROBLEM FORMULATION

We model urban connectivity in the form of user flow as a dynamic graph $\mathbb{G}^t$, consisting of two types of nodes, 1) $\mathbb{L}^t$ - Location nodes representing each location grid in the city and 2) $\mathbb{H}^t$ - Home nodes representing the suburbs of registered homes of people for the considered time window of t. There is a subgraph G_i^t for each location node $L_i^t \in \mathbb{L}^t$ connecting L_i^t to a set of home nodes based on the home locations of the people at L_i^t at time t. As such, we define an edge $e_{i,j}^t \in \mathbb{E}$ between L_i^t and H_j^t if there is at least one user from H_j^t at L_i^t. Edge weight of $e_{i,j}^t$ is defined as the number of people connecting the two node $N_{i,j}^t$.

With the aid of *DS1*, we attempt to model the geo-social connectivity graph $\mathbb{G}^t$. For each unique connection, we constructed a temporal list indicating the existence of connection at different time. We then apply Logistic Regression for predicting the existience of certain geo-social connectivity. Results show a learning score of around 0.7944, and regularization strength does not significantly change the result. Although the existence of connection could be relatively easy to predict with approximately 80% success rate, predicting the strength of certain geo-social connection is quite challenging. Polynomial regression of various degree is trained for *DS1* to predict how strong each connection is temporally, however considerably low R^2 ($< 5 \times 10^{-4}$) could be achieved. Therefore, individual strength of connectivity is hard to predict due to the high dynamics, and data-driven approach is more suitable than direct modeling.

5.1 Known Spatial User Preference

We assume that an advertiser needs to distribute an ad α to a set of locations $\mathbb{L}'$. As node connectivities change over time, we will solve the dynamic problem by solving for each time window t. Further, we presume that individual user preference towards one ad is binary, i.e. people either like or dislike the ad α. Spatial user preference is defined as the percentage of people who are interested in an ad among the total population of the considered area. For each home node $H_j^t \in \mathbb{H}^t$, the spatial user preference for ad α can be denoted as h_j^α, where $h_j^\alpha \in (0,1)$ and $j \in (1,2,...|\mathbb{H}^t|)$.

We define the cost of displaying an ad at location L_i to be C_i and the total cost constraint per an ad as C_{max}^α. Given the number of people in grid L_i^t, N_i^t and its connections with home nodes H_j^t, the problem is to find a set of location $\mathbb{L}' \subset \mathbb{L}^t$ to display the ad α at each time window t that would maximize the exposure to total number of users Γ who are interested in α under the budget constraint C_{max}^α. Thus, we formulate the problem as follows[4];

$$\begin{aligned} \text{Maximize} \quad & \sum_{\forall L_i \in \mathbb{L}'} \sum_{\forall e_{i,j}} N_{i,j} \times h_j^\alpha \\ \text{s.t.} \quad & \sum_{\forall L_i \in \mathbb{L}'} C_i \le C_{max}^\alpha \\ & C_i \ge 0 \end{aligned}$$

$f(L_i) = \sum_{G_i} N_{i,j} \times h_j^\alpha$ computes the total number of people who are interested in α at location L_i, by considering its connections with home nodes. If we consider, $f(i)$ as the value of each location and C_i as the weight of each location L_i, then it is trivial to show that our problem is equivalent to standard Kanpsack problem, where a Knapsack of size C_{max} needs to filled to the maximum value by selecting a subset of items $\mathbb{L}' \subset \mathbb{L}^t$. This immediately follows that our problem is also *NP-Hard* problem and could not be solved in polynomial time even for one time window. However, Knapsack problem is a well studied problem with number of approximation solutions. A reasonable approximation can be found by a local greedy solution that maximizes the value per unit cost. We denote $\Pi' = \sum_{\forall L_i \in \mathbb{L}'} f(L_i)$ as total value of selected set $\mathbb{L}'$.

5.1.1 Multiple knapsack problem

We can further extend to a list of advertisers with ads $\{\alpha, \beta, \gamma...\}$ to be distributed, and optimal locations sets $\{\mathbb{L}_1, \mathbb{L}_2, \mathbb{L}_3...\}$ needs to be determined that will maximize the total value for all advertisers. Similarly, people's interests towards β and γ for all nodes are denoted as $h_1^\beta, h_2^\beta, h_3^\beta, ...h_n^\beta$ and $h_1^\gamma, h_2^\gamma, h_3^\gamma, ...h_n^\gamma$ respectively. As a result, the value of each location L_i is different for each advertiser. We assume that the same digital screen can only display one ad at any given time window. From a DOOH system point of view, the selection of each $\mathbb{L}_k$ might be conflicting goals. The optimization of the whole system is a more complicated problem to study. In this paper, we will focus on the single knapsack problem, but the same methodology can be used to extend to multi-knapsack problem.

[4]For the remainder of the formalization, we drop the time window notation t for brevity

5.1.2 Naive Solution

We assume currently advertisers would choose their location based purely on their budget and total number of people the ad needs to be display to. The higher density of business M_i is generally correlated with higher consumer flow. As business density could be easily determined from yellow-pages or online location based service websites such as Yelp. One simple way is to maximize metrics $\frac{M_i}{C_i}$ (business density per cost) and only display ads in the areas with highest number of consumer per dollar over long term. This solution, however, ignores the dynamics of mobile user and user correlation. We denote this method as "Current Naive Method" to compare the performance with other algorithms.

5.2 Limited Knowledge on User Preference

In the previous subsection, we determined problem of finding the optimal set of location to distribute a given ad α. However, an accurate knowledge of fine-grained spatial user preference is not available. As a result, in this section, we discuss how we could potentially leverage user correlation when limited knowledge about user preference is known. We assume advertisers only have knowledge about the top k home nodes that have the highest preference for an ad α. We propose a heuristic algorithm based on spatial similarity that output a list of potential locations $\mathbb{L}''$ based on the limited knowledge of $h_j^\alpha, j \in (1,2,...|\mathbb{H}^t|)$.

5.2.1 Edge vector Distance

Methods to compare graph similarity has been extensively studied in the literature. i.e. iterative method SimRank [9], feature extraction methods [6] and graph isomorphism method Gromov-Hausdorff distances [16] and Eigenvector Similarity [12]. Some further related work is surveyed in [18]. In our application, we are particularly interest in "if a node in a graph is similar to a node in another graph depends on the neighbor nodes its connected to".

We use edge vector distance as a form of measuring user correlation that best suit our scenario. Specifically, for our particular measure, given $e_1, e_2... \in \mathbb{E}$ are edges shared by graph G_1 and G_2 ($E = E_1 \cap E_2$). We construct vector v_1 and v_2 where v_i is the edge distance in each graph. We define the edge length $\epsilon_e^t = N_{i,k}^t$, where N_{i,k_1}^t is the number of people connecting node N_1 with home node H_k at time t. We compare edges giving each edge e weight ϵ_e^t capturing its local topology.

$$d^t(L_1, L_2) = \frac{\sum_{e \in E_1 \cap E_2} \frac{(\epsilon_{e,1}^t - \epsilon_{e,2}^t)}{\max(\epsilon_{e,1}^t, \epsilon_{e,2}^t)}}{|E_1 \cap E_2|} \tag{3}$$

5.2.2 Spatial Similarity based Heuristic Algorithm

Given the knowledge of top K location set $\mathbb{L}'$, we would like to identify candidate location set $\mathbb{L}''$ under constraint. We do this by leveraging spatial similarity derived by edge vector distance. Detailed algorithm is described in Algorithm 1. In general, we first find point with the shortest mean pair-wise distance to all top K' locations. These are the points that are most similar to the top K' locations with regards to the neighbor nodes that are connected and how they are connected. We then recursively add these points

Algorithm 1 Similarity Based Heuristic Algorithm

Input: current graph connectivity, top K' locations $\mathbb{L}'$
Output: set of target locations $\mathbb{L}''$
initialization: current graph $\mathbb{G}^t$, initial top K' locations $\mathbb{L}' = L_1, L_2..L_k$,
forall the *possible nodes pair* $\{L_i,\ L_j\} \in \mathbb{G}^t$ **do**
  Edge vector distance; $d^t(L_i, L_j) \leftarrow Eq.(3)$;
end
distList = list()
for $L_i \in \mathbb{G}^t$ **do**
  $D_i = \sum_{L_c \in \mathbb{L}'} d^t(L_i, L_c)$
  distList.add(D_i)
end
distList.sort()
while $\sum_{L_i \in \mathbb{L}'}(C_i) \leq C^{\alpha}_{max}$ **do**
  $\mathbb{L}''.add(D_i)$ from head of distList
end

Table 1: Summary of Initial Parameters

C^{α}_{max}	1000	C_i	$\in$ [1,100]
α_i	$\in [0, 1]$	K'	3

to the output set based on the similarity until the budget is used up.

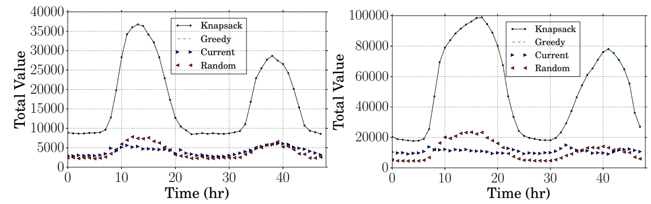
6. EVALUATION

6.1 Known Spatial User Preference

We developed a *Python* simulator to evaluate the performance of system in a real-world setup. First, we assume the spatial variation of user preference is known. In the next subsection, we further evaluate the performance of our proposed algorithm when limited knowledge about spatial user preference is given. A summary of parameters used in simulation are listed in Table 1. We initialize the cost C_i of each location L_i by a random value between 1 to 100 and a user preference α_i by a random value between 0 and 1. Furthermore, if not specified, C^{α}_{max} budget is set to be 1000. The random algorithm is repeated 50 times, where mean value is taken for comparison.

We first show the temporal variation of total value Π' using *DS2* , which indicates the temporal fluctuation of total eyeballing to users of interest. In Figure 7, we show the performance of four algorithms in four hot-spots with an hourly time window. The greedy algorithm performs approximately similar to the optimal solution ($\sim$ 99%) across all four difference locations over time. In addition, advertisers' current naive solution does not always perform better than the random algorithm which is dependent on the time and location. Interestingly, the total value derived by the optimal solution displays a diurnal pattern due to the dynamic changes of urban geo-social graph.

We further test all algorithms for the whole city simulating *DS1* in Figure 8. The knapsack solution again performs very similarly to the greedy solution. Moreover, It could be seen that greedy solution, which considers user correlation performances consistently better than both current naive solution and random algorithm. In fact, on average around 300% improvement in total value Π' comparing to current solution and a 2800% improvement against random algorithm. Furthermore, current naive solution considering both spatial characteristics and cost outperforms the



(a) Location 1 (b) Location 2

(c) Location 3 (d) Location 4

Figure 7: Temporal performance variation

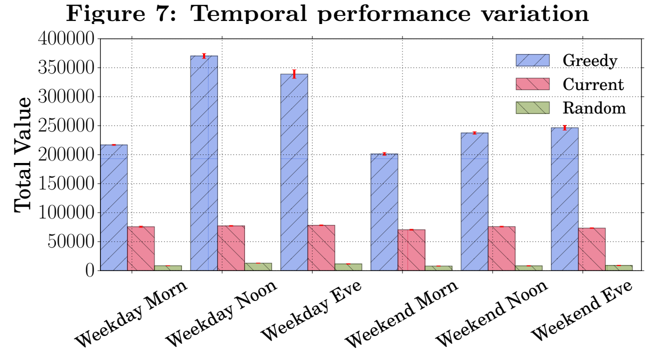


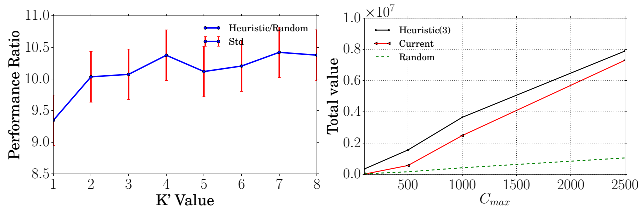
Figure 8: Overall City Performance

random algorithm by almost 8 times, although huge performance gain could be further achieved. This might be a result of the existence of correlation between business density and user density. The performance of greedy algorithm is time variant, while random and current naive solution perform relatively stable over time.

6.2 Performance of Heuristic Algorithm

We then evaluate the performance of our proposed heuristic algorithm when user preference is not completely known. We first vary the value of K', and compare the performance of the proposed similarity based heuristic algorithm to random algorithm. It could be seen that only by knowing the top 3 locations, where the highest number of users like the content, performance could be significantly improved. By leveraging user correlation, distribution efficiency is improved by over 10 times against a random algorithm. In addition, this little information could also improve advertisers' current algorithm performance by approximately 35%. However, higher K' value does not dramatically improve the performance further as shown in Figure 9a. In addition, we also vary the budget amount C^{α}_{max} and compare the performance with regards to total value Π' in Figure 9b. In general, our proposed heuristic algorithm always outperforms the current solution, although the gap narrows as the budget amount C^{α}_{max} increases.

A performance comparison is shown in Table 2 to compare the current solution and heuristic algorithm with the upper (knapsack optimal) and lower bound (random). The normalized total value Π' is used for comparison. We find

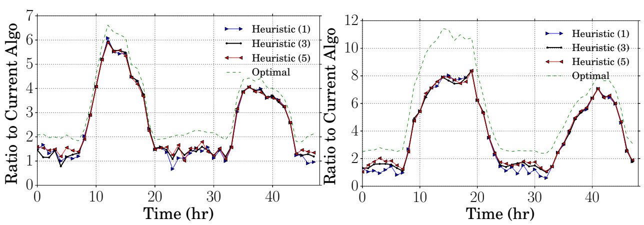


(a) Knowledge of user (b) Varying C_{max}^{α}

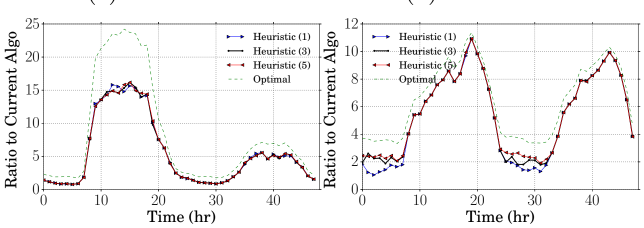
Figure 9: Performance Evaluation

Table 2: Performance Evaluation

Algorithm	Normalized Total Value Π'	Per. gain
Random	1	-
Current Naive	7.5	650%
Heuristic(1)	9.4	25.3%
Heuristic(3)	10.1	7.4%
Optimal	29.2	189.1%



(a) Location 1 (b) Location 2



(c) Location 3 (d) Location 4

Figure 10: Heuristic Performance

that even $K' = 1$ would significantly improve current content distribution efficiency by 25%. A higher K' translate to a higher performance gain, however, the rate of performance gain slows down as K' increases. Furthermore, the last column shows the performance gain comparing to the previous algorithm. Finally, we also evaluate the heuristic algorithm on the four hotspots by varying the amount of known user information K' in Figure 10. It could be seen again that performance of heuristic algorithms is spatial and time dependent. Heuristic algorithm could perform really closely to the optimal solution in some locations, while is far from optimal in some other locations. As a result, traditional static optimization method would perform poorly in real-world scenario compared to data driven approaches.

7. RELATED WORK

A DOOH ad delivery system was initially proposed in [5] in a distributed manner, where demographic data is tracked by individual display. Phan et al. [20] present a content management system for delivering both advertising and non-advertising content for digital signage system. Their system is capable of receiving, storing and scheduling contents on a location-based out of home advertising network. The claimed scheduling algorithm determines available inventory slot (screen location and time) based on registered user information. Authors in [19] extended the system by creating a network program wheel to manage time slots in the network. The above systems require users connect to the system for registration. AdTorrent [17] is a system for targeted advertisement distribution in a vehicular network. The system targets individual mobile user and integrate search, ranking and ad content delivery in the architecture. In [25], authors designed a location-aware mobile digital signage system (LDSS) based on GPS and wireless infrastructure as contrasted to traditional static digital signage. An advertisement recommendation algorithm was proposed and compared with traditional advertising methods, i.e. region triggered and sequential advertising. However, these works focused on individual user and do not take into account the spatio-temporal dynamics of “cohort of user” interest.

Although targeting individual user has proven successful for online advertising, it has its limitation in DOOH network due to privacy concerns. There has been intensive research both in academia and industry in the field of DOOH attempting to improve ad distribution without targeting individual user. In [13], authors addressed the challenges and limitations of personalized advertising in DOOH, and suggested situationalization, which delivers content relevant to individual or a group of individuals based on the context. They further proposed PERSIT matrix with adaptation strategy between personalization and situationlization. PERSIT differs from our work in that it only provides a guide of when to adapt, but does not consider how to improve ad distribution in a system. Satoh [22] presented a framework for context-aware digital signage. This work is limited to displaying location based content in a single digital screen, rather than a DOOH system. Furthermore, it does take into account the spatio-temporal dynamics of users who would consume the content.

Finally, our optimization problem is related to the “billboard/retail store location selection” problem in the field of land economy [15]. Geo-spotting [11] used location-based social network check-in data (Foursquare) to identify optimal location for new retail store. In [14], authors attempted to tackle the problem combining visualization and data mining using taxi trajectory data. However, both our problem formulation and methodology are significantly different, and user dynamics are not discussed in both works.

8. LIMITATIONS AND FUTURE WORK

In this work, we mainly exploit the spatio-temporal dynamics of users in the form of geo-social connectivity with large-scale mobile dataset, and there are factors could potentially be further studied, i.e. spatial user preference and additional spatial similarity features. Firstly, in our work, user interest/preference in a certain area towards ad α is initialized randomly. The existence of spatial user preference could further improve ad distribution by a similarity based heuristic algorithm. However, details about spatial user preference are not well understood, apart from discussed in a few studies [26]. Questions further about user preference that might be understand through social networks or survey data are;

- Is there spatial user correlation of preference? Does shorter distance between two areas link to a higher similarity of user preferences?
- What other spatial characteristics could potentially contribute to the dynamic similarity graph.

These questions related to spatial user preference, if understood, could be extended into a broad field of study. Hence, the similarity graph could be further extended to include these correlations and characteristics. Secondly, this paper mainly addresses the problem of spatial optimization at certain time window. We intent to further extend the paper into optimizing content distribution system in DOOH, where multiple content with different user preferences needs to be distributed.

Finally, as we have constructed a spatial geo-social similarity graph based on spatial-temporal user dynamics. The graph could be built one off, used by various services, and be updated periodically. We believe the graph could potentially be used by a large number of mobile location based services and applications, i.e. to improve recommendation in location-based systems [2, 1] etc. Furthermore, the knowledge of spatial user preference could also be collected much easier through social networks or surveies compared individual user data. These graphs could be highly suitable for tasks such as urban planning, and land economy planning [3]. Our future work also includes blending data from different sources to improve the geo-social graph.

9. CONCLUSION

In this paper, we discuss how mobile data could be exploited to improve content distribution efficiency in DOOH advertising network. More specifically, we first analyze a mobile dataset consists of fine-grained population dynamics and urban geo-social links. We construct a dynamic graph of urban connectivity using the dataset, and formulated an optimization problem of reaching highest number of "right" people at the "right" time. In addition, a heuristic algorithm is proposed to solve the problem when limited user preference is known. We evaluate multiple algorithms with data-driven simulations, and show an over 300% improvement compared to advertisers' current distribution approach. This is achieved by targeting "cohort of user" perference and considering the spatio-temporal dynamics of those user preferences. Finally, our constructed urban similarity graph could potentially be used by many other mobile services and geo-social applications.

Acknowledgment

This research was supported/partially supported by Data61, CSIRO, UNSW, Australian Government Research Training Program Scholarship and NII. The work was partially conducted under MOU program when author was in NII, Japan. We thank our colleagues who provided insights and expertises that greatly assisted the research.

10. REFERENCES

[1] Bao, J., Zheng, Y., and Mokbel, M. F. Location-based and preference-aware recommendation using sparse geo-social networking data. In *Proceedings of the 20th international conference on advances in geographic information systems* (2012), ACM, pp. 199–208.

[2] Bao, J., Zheng, Y., Wilkie, D., and Mokbel, M. Recommendations in location-based social networks: a survey. *Geoinformatica 19*, 3 (2015), 525–565.

[3] Batty, M. Big data, smart cities and city planning. *Dialogues in Human Geography 3*, 3 (2013), 274–279.

[4] Bauer, C., Dohmen, P., and Strauss, C. Interactive digital signage-An innovative service and its future strategies. In *Emerging Intelligent Data and Web Technologies (EIDWT), 2011 International Conference on* (2011), IEEE, pp. 137–142.

[5] Carney, P. J., Pina, J. B., Boyle, J. J., and Perine, C. A. System and method for delivering out-of-home programming, June 18 2002. US Patent 6,408,278.

[6] Cha, S.-H. Comprehensive survey on distance/similarity measures between probability density functions. *City 1*, 2 (2007), 1.

[7] EMarkerter. Propped by Digital Growth, Out-of-Home Advertising Holds Its Own in UK, France. Tech. rep., 2015.

[8] EMarkerter. U.S. Digital Out-of-home advertising. Tech. rep., 2015.

[9] Jeh, G., and Widom, J. Simrank: a measure of structural-context similarity. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (2002), ACM, pp. 538–543.

[10] Jiang, F., Thilakarathna, K., Kaafar, M. A., Rosenbaum, F., and Seneviratne, A. A spatio-temporal analysis of mobile internet traffic in public transportation systems: A view of web browsing from the bus. In *Proceedings of the 10th ACM MobiCom Workshop on Challenged Networks* (2015), ACM, pp. 37–42.

[11] Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V., and Mascolo, C. Geo-spotting: mining online location-based services for optimal retail store placement. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (2013), ACM, pp. 793–801.

[12] Koutra, D., Parikh, A., Ramdas, A., and Xiang, J. Algorithms for graph similarity and subgraph matching. In *Technical report*. Carnegie-Mellon-University, 2011.

[13] Lasinger, P., and Bauer, C. Situationalization, the new road to adaptive digital-out-of-home advertising. In *Proceedings of IADIS International Conference e-Society* (2013), pp. 162–169.

[14] Liu, D., Weng, D., Li, Y., Bao, J., Zheng, Y., Qu, H., and Wu, Y. Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations. *IEEE transactions on visualization and computer graphics* (2016).

[15] Malczewski, J. Gis-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science 20*, 7 (2006), 703–726.

[16] Mémoli, F. Gromov-hausdorff distances in euclidean spaces. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on* (2008), IEEE, pp. 1–8.

[17] Nandan, A., Das, S., Zhou, B., Pau, G., and Gerla, M. Adtorrent: digital billboards for vehicular networks. In *Proc. of IEEE/ACM International Workshop on Vehicle-to-Vehicle Communications (V2VCOM), San Diego, CA, USA* (2005).

[18] Papadimitriou, P., Dasdan, A., and Garcia-Molina, H. Web graph similarity for anomaly detection. *Journal of Internet Services and Applications 1*, 1 (2010), 19–30.

[19] Phan, J. Web-based system and method to implement digital out-of-home advertisements, Dec. 1 2011. US Patent App. 13/115,773.

[20] Phan, M., Woo, D. Q., and Araki, J. Content management in out-of-home advertising networks, 2010.

[21] Ranganathan, A., and Campbell, R. H. Advertising in a pervasive computing environment. In *Proceedings of the 2nd international workshop on Mobile commerce* (2002), ACM, pp. 10–14.

[22] Satoh, I. A framework for context-aware digital signage. In *International Conference on Active Media Technology* (2011), Springer, pp. 251–262.

[23] Shimosaka, M., Maeda, K., Tsukiji, T., and Tsubouchi, K. Forecasting urban dynamics with mobility logs by bilinear Poisson regression. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2015), ACM, pp. 535–546.

[24] Stalder, U. Digital out-of-home media: means and effects of digital media in public space. In *Pervasive Advertising*. Springer, 2011, pp. 31–56.

[25] Yu, K.-M., Yu, C.-Y., Yeh, B.-H., Hsu, C.-H., and Hsieh, H.-N. The design and implementation of a mobile location-aware digital signage system. In *Mobile Ad-hoc and Sensor Networks (MSN), 2010 Sixth International Conference on* (2010), IEEE, pp. 235–238.

[26] Zheng, Y., Xie, X., and Ma, W.-Y. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull. 33*, 2 (2010), 32–39.