

# Strategic and Crowd-Aware Itinerary Recommendation

Junhua Liu<sup>1,2</sup>(✉), Kristin L. Wood<sup>1,3</sup>, and Kwan Hui Lim<sup>1</sup>

<sup>1</sup> Singapore University of Technology and Design

<sup>2</sup> Forth AI

<sup>3</sup> University of Colorado Denver

j@forth.ai, {kristinwood, kwanhui\_lim}@sutd.edu.sg

**Abstract.** There is a rapidly growing demand for itinerary planning in tourism but this task remains complex and difficult, especially when considering the need to optimize for queuing time and crowd levels for multiple users. This difficulty is further complicated by the large number of parameters involved, i.e., attraction popularity, queuing time, walking time, operating hours, etc. Many recent works propose solutions based on the single-person perspective, but otherwise do not address real-world problems resulting from natural crowd behavior, such as the Selfish Routing problem, which describes the consequence of ineffective network and sub-optimal social outcome by leaving agents to decide freely. In this work, we propose the Strategic and Crowd-Aware Itinerary Recommendation (SCAIR) algorithm which optimizes social welfare in real-world situations. We formulate the strategy of route recommendation as Markov chains which enables our simulations to be carried out in poly-time. We then evaluate our proposed algorithm against various competitive and realistic baselines using a theme park dataset. Our simulation results highlight the existence of the Selfish Routing problem and show that SCAIR outperforms the baselines in handling this issue.

**Keywords:** Tour Recommendations · Trip Planning · Recommendation Systems · Sequence Modelling

## 1 Introduction

Itinerary recommendation has seen a rapid growth in recent years due to its importance in various domains and applications, such as in planning tour itineraries for tourism purposes. Itinerary recommendation and planning is especially complex and challenging where it involves multiple points of interest (POIs), which have varying levels of popularity and crowdedness. For instance, while visiting a theme park, the visitor's route can include POIs such as roller coasters, water rides, and other attractions or events. The itinerary recommendation problem can be modelled as an utility optimization problem that maximizes the number of facilities visited and the popularity of these facilities<sup>4</sup>, while minimizing the queuing time and travel time from one facility to the other. Facilities in a theme park come with different properties such as popularity, duration, location and a dynamic queuing time. Visitors are often constrained by a time budget that limits

---

<sup>4</sup> The terms "POIs", "attractions" and "facilities" are used interchangeably.



Fig. 1: Existing itinerary recommendation problems leverage data-driven approaches with a single-person perspective. In real life, this will result in the Selfish Routing problem, where leaving all agents free to act according to their own interests results in a sub-optimal social welfare. As illustrated, the recommended path performed sub-optimally, where the closer the POIs are to the start of the route, the more crowded they would be, while leaving all other POIs (in grey) not utilized.

the number of facilities one could visit in a single trip. While many algorithms have been developed [40,18,4,7,26], they mostly aim to recommend itineraries for individual travellers, whereas a real-life itinerary is also affected by the actions of other travellers, such as lengthening the queuing time at a facility.

Many works focus on constructing a single optimal path for the individual traveller, solely based on historical data. While this approach works for the individual traveller, it leads to a sub-optimal itinerary when all travellers are given the same recommendation. Consider a recommender system that recommends an itinerary comprising the most popular POIs with the least queuing time based on such historical data. In a real-life scenario with multiple travellers, all travellers will follow the same recommended itinerary with the shortest historical queuing time, resulting in an expected queuing time that would grow with each new arrival, as illustrated in Figure 1. In other words, the later an agent<sup>5</sup> arrives to the system, the longer her expected queuing time will be. As a result, the social welfare or the collective utility of all agents has failed to be optimized. As an individual traveller, it is extremely difficult for an agent to gain knowledge of the system state, i.e., the people who are visiting the park and their respective paths. As a

<sup>5</sup> We use the terms "travellers", "visitors" and "agents" interchangeably.

result, letting the agent find an optimal strategy that maximizes her expected utility is unrealistic without considering the actions of other agents.

To address this problem, we propose the Strategic and Crowd-Aware Itinerary Recommendation (SCAIR) algorithm, which is a recommender system that maintains an internal information of all recommended routes and leverages on this internal information to make routing recommendations to its arriving agents. In other words, we take a game-theoretic approach to address the problem and formulate a crowd-aware itinerary recommendation algorithm having in mind the Selfish Routing problem [32], i.e., allowing agents act freely results in a sub-optimal social welfare. Concretely, we model the itinerary recommendation problem into a strategic game [25], where the system, i.e., a theme park, defines a set of allocation rules to allocate route to each player in the system, instead of leaving the agents a high degree of freedom to choose their own path. Experiments show that our approach is effective in optimizing utility of all agents.

## 2 Main Contributions

Our main contributions are as follows:

- We introduce and formulate the crowd-aware itinerary recommendation problem as a social welfare optimization problem that considers the actions of multiple travellers, in contrast to existing works that only consider the perspective of the single traveller (Section 4).
- To address this crowd-aware itinerary recommendation problem, we propose the SCAIR algorithm which utilizes a game-theoretic approach to recommend itineraries for multiple travellers (Section 5).
- Using a theme park dataset, we compare our SCAIR algorithm against various competitive and realistic baselines and show how SCAIR outperforms these baselines with a large reduction in queuing times (Sections 6 and 7).

For the rest of the paper, Section 3 discusses related works and how our research differs from these earlier works. Section 8 summarizes this paper and introduces some future research directions. Next, we introduce the problem formulation of this crowd-aware itinerary recommendation problem.

## 3 Related Work

There have been numerous works that aim to solve the itinerary recommendation problem and other related tourism recommendation problems. In this literature review, we cover various related work from the Operations Research and Information Retrieval communities.

### 3.1 Itinerary and Tourism-related Recommendation

Many works have modelled the itinerary recommendation problem as a variant of the Orienteering problem [4,7,17]. In the Orienteering problem [33], the recommendation

aims to optimize social welfare with a global reward such as popularity, with respect to budget constraints such as travel time or distance among attractions in an itinerary. This approach typically does not take into consideration the trade-off between the duration in a facility and its popularity, which may contribute substantially to the global profit. For more details on the Orienteering problem and its variants, we refer the interested reader to the various survey papers on this topic [8,34].

There have been various approaches proposed to solve this itinerary recommendation problem based on variants of the Orienteering problem. For example, Zhang et al. proposed the use of heuristic approximation to solve a variant of this problem that involves POI opening hours and incorporating uncertainty in different travel modes [41,40]. Others have used variations of the Ant Colony System to solve the itinerary recommendation problem [19], and variants that incorporate the additional consideration of crowd levels [36]. Another approach is to solve this itinerary recommendation problem using integer programming, to optimize for user interests based on the amount of time tourists spend at POIs [18]. Similarly, there are also various works that aim to recommend routes that are deemed more attractive to tourists [42,10,35]. In the Information Retrieval community, a popular research topic is on item recommendations and this problem can be easily extended for recommending POIs. For example, many works have used matrix factorization or collaborative filtering approaches for finding a ranked list of top locations, which is known as top-k POI recommendation [38,39,15,37,14].

### 3.2 Discussion

These earlier works face a major limitation where the recommendation algorithms are constructed based on a single person's perspective. Despite some recent works exploring the effects of group or crowd behavior [36,6,1,11,9], the algorithms treat the system as a static environment where properties such as queuing time only depend on historical data. Simulating an optimal path in such a static environment has a natural disadvantage where self-interested agents prioritize personal objective functions which may result in ineffective social welfare. For instance, when everyone visiting the theme park follow the same recommended path, the queuing time will increase dramatically, and the optimality of such recommendation algorithms will then collapse. Roughgarden's work [32] discusses this problem extensively, defined as Selfish Routing problem, where giving agents the freedom to act according to their own interests results in a sub-optimal social welfare.

The Selfish Routing problem was studied in the area of Game Theory and Mechanism Design [32,12,28]. The inefficiency of achieving the optimize natural objective is quantitatively measured by Price of Anarchy, which was first defined as the ratio between the worst-case Nash equilibrium and the optimum sum of payoffs in game-theoretic environments [12,28]. Braess's Paradox for traffic flow [3] describe the phenomenon where adding a new link to a transportation network might not improve the operation of the system, in the sense of reducing the total vehicle-minutes of travel in the system [30]. To break out from this phenomenon, a system operator can manually interfere with or change agents' actions to provide policies or economic incentives with well designed strategies. Our proposed game-theoretic, dynamic itinerary recommendation algorithm in this paper is an instance of such strategy.

To address these limitations, we propose the Strategic and Crowd-Aware Itinerary Recommendation (SCAIR) algorithm to address the ineffectiveness of welfare optimization due to the lack of centralized control [31]. The proposed recommendation algorithm takes into consideration all visits in an itinerary planning scenario (e.g., a theme park), and makes recommendations to the next visitor with the knowledge of other visitors' paths in the park. Furthermore, the queuing time at all facilities at a certain hour is dynamically modelled according to the expected number of visitors in the same place at the same hour.

## 4 Crowd-aware Itinerary Recommendation Problem

In this section, we first give an overview of our general approach, followed by formulating our crowd-aware itinerary recommendation problem, before showing the NP-hardness of this proposed problem.

### 4.1 General Approach

In this work, we view the itinerary recommendation problem from a global perspective and formulate it as a strategic game where the system designs and distributes the optimal path to every agent on arrival, based on the existing agents in the system and their respective paths. In the context of a theme park, one can think of this entity as the theme park operator that gives out the recommendation of various itineraries to visit the attractions to different visitors. We propose the SCAIR algorithm that dynamically recommends routes taking into consideration all existing agents in the system.

The crowd-aware itinerary recommendation problem aims to maximize the sum of all agents' utility in the system. This turns out to be a social welfare optimization problem that is NP-hard [24]. Furthermore, simulating or solving the problem is also empirically challenging. One has to take into consideration the entire history of existing visitors results in exponential space-complexity with respect to the number of agents, and exponential time-complexity with respect to the number of facilities in a path.

To overcome these challenges, we propose a simplified version which models the recommendation problem as a finite markov chains and is known to be in NC [29] and decidable in poly-logarithmic time [2]. The simplified model makes an assumption that each decision embeds information of the immediate last decision and the model as a result is able to provide a snapshot of the entire history. Next we will discuss the formulation of the problem.

### 4.2 Problem Formulation

We formulate the crowd-aware itinerary recommendation problem to be a finite markov chain and impose constraints such as (1) fixing the starting point, (2) setting a time budget for the path, and (3) limiting the distance between two stations. These constraints reflect real-life considerations closely, such as fixed starting point near the entrance; visitors having limited time to tour; and dissatisfaction arising with long walking distance among facilities.

Concretely, we model the theme park comprising numerous tourist attractions as a fully connected graph  $G(F, C)$ , where  $F = \{f_1, \dots, f_n\}$  is the collection of  $n$  facilities in the system, and  $C = [c_{ij}]$  is the set of connections from  $f_i$  to  $f_j$ . Each connection  $c_x$  is associated with the properties of distance  $Dist(c_{ij})$  and travel time  $Trav(c_{ij})$  in minutes. Each facility  $f_x$  is associated with a set of properties including coordinates  $(lat_x, long_x)$ , duration of visit  $Dur(f_x)$  in minutes, capacity  $Cap(f_x)$  and popularity  $Pop(f_x)$ .

We formulate the agents' visits as  $m$  states  $S = \{s_1, \dots, s_m\}$ , where each state  $s_x$  is associated with a feasible path  $p_x = [f_1^{(x)}, \dots, f_{n_x}^{(x)}]$  with  $n$  facilities  $[f_1^{(x)}, \dots, f_{n_x}^{(x)}]$ . The total time  $TT_x$  of path  $p_x$  is defined as:

$$TT_x = \sum_{i=1}^{n_x} Dur(f_i^{(x)}) + \sum_{i=1}^{n_x-1} Trav(c_{i,i+1}) \quad (1)$$

We model the utility of the agents with respect to the popularity of each facility visit normalized by the expected waiting time at each facility. Our assumption is that higher popularity of a facility indicates a greater attractiveness to visitors, subjected to how long they have to wait for that facility. Concretely, we define the utility function  $U_x$  for path  $x$  with  $n$  nodes as follows:

$$U_x = \frac{\sum_{f \in p_x} Pop(f)}{Q(p_x | p_{x-1})} \quad (2)$$

where  $Q(p_x | p_{x-1})$  is the expected queuing time at path  $p_x$  given  $p_{x-1}$ , and  $Pop(p_x)$  is the sum of popularity of all facilities in the path. The path's expected queuing time  $Q(p_x | p_{x-1})$  is calculated by summing up the queuing time at all facilities:

$$Q(f_i) = \frac{1}{Cap(f_y)} Dur(f_y) \delta(f_{y,h}^{(x)} = f_{y,h}^{(x-1)}) \quad (3)$$

where  $\delta(f_{y,h}^{(x)} = f_{y,h}^{(x-1)}) = 1$  if the facility appears to overlap between paths  $p_x$  and  $p_{x-1}$  within the same hour  $h$ . Capacity  $Cap(f_x)$  is set to be a constant for simplicity. Finally, the transition matrix  $T$  is defined as:

$$T_{ij} = \frac{\sum_{f \in p_j} Pop(f)}{Q(p_j | p_{j-1}=i)} \quad (4)$$

The transition matrix is then normalized by:

$$T_{ij} := \frac{T_{ij}}{\sum_j T_{ij}} \quad (5)$$

The set of feasible paths, i.e., total search space, is determined by solving an optimization problem, as follows:

$$\begin{aligned} \text{maximize} \quad & TT_x = \sum_{i=1}^{n_x} Dur(f_i) + \sum_{j=1}^{n_x-1} Trav(c_{j,j+1}) \\ \text{subject to} \quad & Dist(c_{j,j+1}) \leq s, \quad TT_x \leq t \end{aligned} \quad (6)$$

for  $n$  facilities in the path, with a constant time budget  $t$ .

Finally, we model the strategic itinerary recommendation problem as a social welfare optimization problem as follows:

$$\begin{aligned} & \text{maximize} && W = \sum_x U_x p_x \\ & \text{subject to} && \sum_x TT_x \leq t, \quad x \in \{1, \dots, n\} \end{aligned} \quad (7)$$

for  $n$  agents and time budget  $t$ .

### 4.3 Proof of NP-Hardness

We further investigate the NP-hardness of various sub-problems and show the respective proofs in this section.

**Theorem 1.** *The path finding problem defined in Equation 6 is NP-hard.*

*Proof.* We prove the NP-hardness of the path finding problem by reduction from the 0-1 Knapsack problem which is known to be NP-hard [23]. Recall that the 0-1 Knapsack problem is a decision problem as follows:

$$\begin{aligned} & \text{maximize} && z = \sum_i p_i x_i \\ & \text{subject to} && \sum_i w_i x_i \leq c \\ & && x_i \in \{0, 1\}, \quad i \in \{1, \dots, n\} \end{aligned} \quad (8)$$

for  $n$  available items where  $x_i$  represents the decision of packing item  $i$ ,  $p_i$  is the profit of packing item  $i$ ,  $w_i$  is the weight of item  $i$ ,  $c$  is the capacity of the knapsack.

Intuitively, the path finding problem is a decision problem of allocating a set of facilities into a path with a capacity of time budget, where each facility comes with properties of profit and duration time.

Formally, we transform the minimization problem in Equation 6 to an equivalent maximization problem. Concretely, the binary variable  $f_i \in \{0, 1\}$  is included, where  $f_i = 1$  if  $f_i$  is in path  $p_x$ , and 0 if otherwise. Furthermore, we define the profit of facility  $f_i$  as  $p_i = -Dur(f_i)$  and set the travel time  $Trav(c_{ij})$  to be a constant. Finally, the distance constant  $cap$   $s$  is set to be infinity. The new problem formulation is represented as follows:

$$\begin{aligned} & \text{maximize} && T'_{path} = \sum_i p_i f_i \\ & \text{subject to} && \sum_i Dur(f_i) f_i \leq t \\ & && f_i \in \{0, 1\}, \quad i \in \{1, \dots, n\} \end{aligned} \quad (9)$$

In this formulation, a path is equivalent to the knapsack in the 0-1 Knapsack problem, where each facility has its profit of  $p_i$ , and its cost of  $Dur(f_i)$  that is equivalent to the profit and weight of an item respectively. The maximization problem is subjected to a constant time budget  $t$  which is equivalent to the capacity  $c$  in a 0-1 Knapsack problem.

As a result, for any instance of the 0-1 Knapsack problem (i.e. item allocation decisions), we are able to find an equivalent instance of the path finding problem (i.e. a facility allocation decisions). Therefore, a solution in the path finding problem yields an equivalent solution to the 0-1 Knapsack decision problem. As such, we have completed the proof of NP-hardness for our path finding problem to be NP-hard.  $\square$

**Theorem 2.** *The social welfare optimization problem defined in Equation 7 is NP-hard.*

*Proof.* Once again, we prove the NP-hardness of our welfare optimization problem by reduction from the 0-1 Knapsack problem.

In Equation 7, the set of paths assigned to agents in the system is equivalent to the set of items in 0-1 Knapsack problem; each path has its utility and total time, which are equivalent to the profit and weight of an item respectively; the maximization problem is subjected to a constant time budget  $t$  which is equivalent to the capacity  $c$  in a 0-1 Knapsack problem.

As a result, for any instance of the 0-1 Knapsack problem decisions, we are able to find an equivalent instance of a path assignment decision that yields a solution to the original Knapsack decision problem. As such, we conclude the proof of NP-hardness and have shown that our welfare recommendation problem is NP-hard.  $\square$

Next, we describe our proposed SCAIR algorithm for solving this crowd-aware itinerary recommendation problem.

## 5 Strategic and Crowd-Aware Itinerary Recommendation (SCAIR) Algorithm

In this section, we describe our proposed SCAIR algorithm, which comprises the main steps of finding feasible paths, generating a transition matrix and simulating traveller visits.

### 5.1 Finding Feasible Paths

Algorithm 1 shows the pseudocode of our path finding algorithm based on a breadth-first strategy. The input is a graph  $G(F, C)$  that represent a theme park with the set of facilities  $F$  and connections  $C$ , time budget  $TT_{max}$ , and distance limit between two facilities  $Dist_{max}$ . This algorithm then generates and returns a collection of feasible paths,  $Paths$ , with respect to the provided input graph  $G(F, C)$ .

We iterate the collection of intermediate  $Paths$ , and call the *FindViableFacilities* function to find viable facilities, where  $f_{-1}^{(i)}$  is the last facility of the path, and  $Dist_{max}$  is the maximum distance an agent wants to travel from one facility to another. We set



**Algorithm 1:** SCAIR - FindFeasiblePaths()

---

**Data:**  $f_i \in F, c_{ij} \in C, TT_{max}, Dist_{max}, f_0$   
**Result:**  $Paths$ : the set of feasible paths

```

1 begin
2    $Paths = [[f_0]]$ ;
3   while True do
4     for  $path_i \in Paths$  do
5        $VF = FindViableFacilities(f_{-1}^{(i)}, Dist_{max})$ ;
6       if  $len(VF) == 0$  then
7          $path_x = path_i + [FindNextNearest(f_{(-1)}^{(i)})]$ ;
8         if  $TT_x < TT_{max}$  and  $path_x \notin Paths$  then
9            $Paths+ = [path_x]$ ;
10           $Paths.pop(path_i)$ 
11        end
12      end
13      foreach  $vf \in VF$  do
14         $path_x = path_i + [vf]$ ;
15        if  $TT_x < TT_{max}$  and  $path_x \notin Paths$  then
16           $Paths+ = [path_x]$ ;
17        end
18      end
19       $Paths.pop(path_i)$ ;
20    end
21    if  $AllPathsMaxTimeBudget(Paths)$  or
22        $AllPathsReachFullLength(Paths)$  then
23      break;
24    end
25  end

```

---

the parameters of total time budget  $T_{max} < 8hours$  and maximum allowed distance between two facilities  $Dist_{max}(f_{current}, f_{next}) < 200m$ . If there are no available facility that meets the distance constraint and the path has available time budget remaining, the agent proceeds to the next nearest facility. We also do not allow an agent to revisit a facility in the same trip.

**Line 2.** The algorithm starts with constructing a 2-dimensional array, where each row represent a path as a sequence of facilities visited. We then conduct a breadth-first search (line 3 to 25), starting with the first row with an element of the initial facility, i.e., the entrance of a theme park.

**Line 6 to 11.** If the algorithm is unable to find a facility within the feasible range, it will instead find the nearest facility that is not yet visited, and assign the new path into the  $Paths$  collection if two conditions are met, namely (1) the new path's total time is within the visitor's time budget  $TT_{max}$ , and (2) no identical path exists in the  $Paths$  collection. Eventually we remove the path the iteration started off.

**Algorithm 2:** SCAIR - Simulate()

---

**Data:** *Parks, TimeBudgets, ArrivalIntervals*  
**Result:** Export simulation data to a CSV file

```

1 begin
2   Results = {};
3   for Park ∈ Parks do
4     for SimTime ∈ TimeBudgets do
5       for  $\lambda$  ∈ ArrivalIntervals do
6         Paths = FindFeasiblePaths(Park, SimTime);
7         T = ConstructTM(Park, Paths);
8         Qt, Pop, Utility = RunSimulation(Paths,  $\lambda$ , SimTime);
9         Update(Results, [Qt, Pop, Utility]);
10      end
11    end
12  end
13  ExportCsvFromDict(Results);
14 end

```

---

**Line 13 to 20.** If the algorithm manages to find a set of viable facilities, it will then iterate through the set and execute a similar selection process.

**Line 22 to 24.** The algorithm breaks out from the infinite loop when any one of two conditions is met, namely (1) all paths in the *Paths* collection have maximized its time budget i.e. any additional facility will make the total time of a path to be larger than the visitor's time budget; or (2) every path has included all available facilities.

## 5.2 Transition Matrix

Using the set of feasible paths found (Section 5.1), we now construct a Transition Matrix *T* by calculating  $T_{ij}$  as the costs of taking path *j* given path  $j - 1 = i$ . The output of *FindCost*() function varies based on the arrival interval  $\lambda$  because it affects the expected time of arrival for each facilities at  $path_j$ , which leads to different occurrence of overlapping facilities between  $path_i$  and  $path_j$ .

## 5.3 Simulation

Algorithm 2 shows an overview of the simulation procedure, which involves iterating through the visit data of theme parks *Parks*, a list of time budgets *TimeBudgets*, and an array of arrival intervals *ArrivalIntervals*.

**Line 6 to line 13.** For each step, the *FindFeasiblePaths*() function finds the set of feasible paths which enables the *ConstructTM*() function to construct the transition matrix, with input parameters namely park data *Park* and simulation time *SimTime*. The *RunSimulation*() function then runs the simulation to find the total queuing time *Qt*, average sum of popularity among all facilities visited *Pop*, and the expected utility *Utility* which is calculated as a function of *Qt* and *Pop*. Finally, we update the *Results* dictionary (Line 9) and export the experimental data into CSV files (line 13) after completing the simulations.

## 6 Experimental Setup

In this section, we describe our dataset, evaluation process and baselines.

### 6.1 Dataset

We conduct our experiments using a publicly available theme park dataset from [16]. This dataset is based on more than 655k geo-tagged photos from Flickr and is the first that includes the queuing time distribution of attractions in various Disney theme parks in the United States. In our work, we perform our experiments and evaluation using the dataset of user visits in Epcot Theme Park and Disney Hollywood Studio.

### 6.2 Experimental Parameters

As previously described in Section 5.2, we denote the arrival interval of agents as  $\lambda$  which indicates the time between the arrival of two agents, measured in minute. In this work,  $\lambda$  is set to be a constant for simplicity. For a robust evaluation, we perform our evaluation using multiple values of the evaluation parameters, namely arrival interval  $\lambda \in \{0.01, \dots, 0.09, 0.1, \dots, 1.0\}$ , and simulation time  $T$  between 60 and 360 minutes in 30 minutes intervals (i.e.  $T \in \{60, 90, \dots, 360\}$ ).

### 6.3 Evaluation and Baselines

We compare our proposed SCAIR algorithm against three competitive and realistic baselines. The first two algorithms are based on intuitive strategies commonly used by visitors in real-life [18], while the third is a greedy algorithm used in [41]. In summary, the three baseline algorithms are:

1. Distance Optimization (denoted as *DisOp*) [18]. An iterative algorithm where agents always choose the facility with the shortest distance to the currently chosen one.
2. Popularity Optimization (denoted as *PopOp*) [18]. An iterative algorithm where agents always choose the next most popular facility that satisfies the specified distance constraint from the currently chosen one.
3. Popularity over Distance Optimization (denoted as *PodOp*) [41]. An iterative greedy approach that models utility as the popularity of the POI normalized by the distance from the current one, and iteratively chooses the POI with the highest utility.

Similar to many itinerary recommendation works [17,16], we adopt the following evaluation metrics:

1. Average Popularity of Itinerary (denoted as *AvgPop*). Defined as the average popularity of all attractions recommended in the itineraries.
2. Expected Queuing Time per Visitor (denoted as *AvgQt*). Defined as the average queuing time that each visitor spends waiting for attractions in the recommended itinerary.
3. Expected Utility (denoted as *Uty*). Defined as the average utility score for all users based on the recommended itineraries.

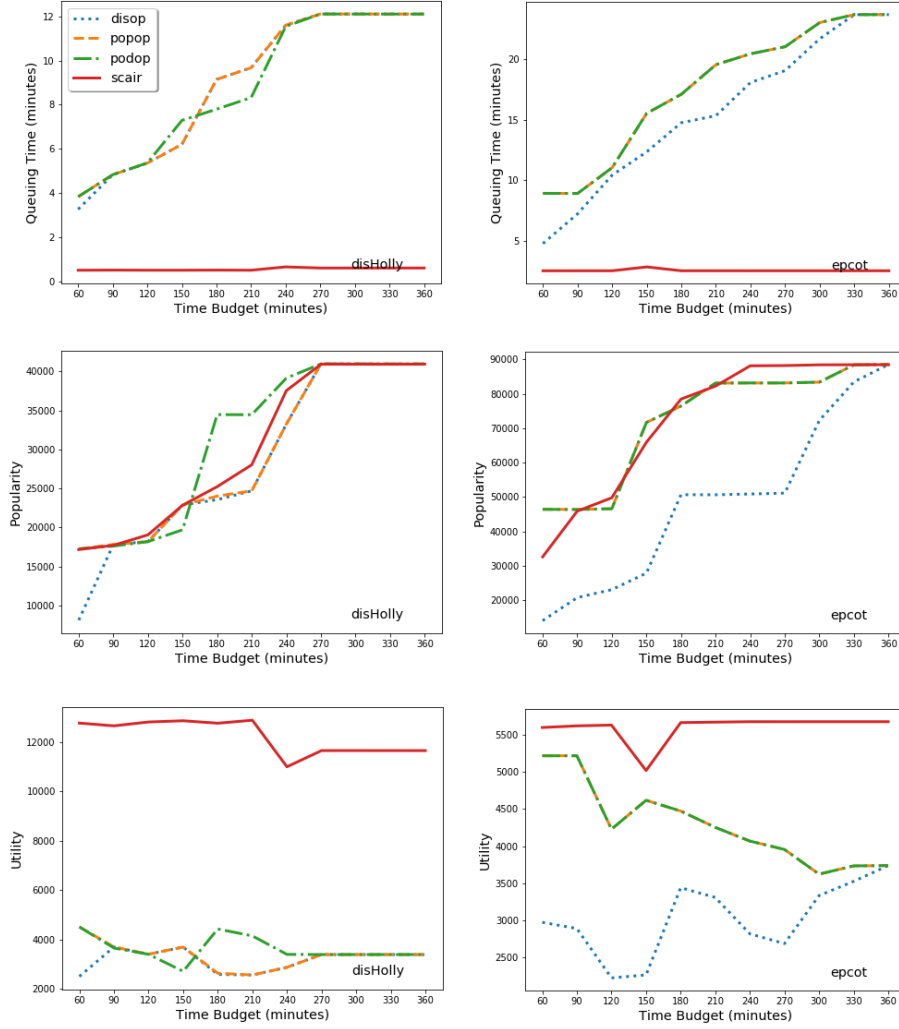


Fig. 2: The plots show how the queuing times, popularity and utility change with respect to simulation time  $T$ , over two theme parks data (Disney Hollywood and Epcot Theme Park). We observe that: (1) SCAIR's queuing time is consistently and significantly lower than the baselines. (2) Popularity of all 4 algorithms perform similarly for DisHolly, while DisOp performs significantly poorer than the others for Epcot. (3) SCAIR's utility consistently outperforms the baselines.

## 7 Results and Discussion

Figure 2 shows the experimental results of our proposed SCAIR algorithm compared to the three baseline algorithms. The x-axis indicates the time budget of visits and the y-axis indicates the queuing time, popularity and utility. To examine the effects of different user arrival frequency, multiple experiments are conducted based on different arrival intervals  $\lambda$ , i.e., from 0.01 to 0.1 with a step size of 0.01, and from 0.1 to 1.0 with a step size of 0.1. The values in the graph are averaged across all  $\lambda$ .

Table 1: Queuing Time Ratio (Smaller values are better)

	Disney Hollywood Epcot Theme Park	
	(DisHolly)	(Epcot)
DisOp	$0.045 \pm 0.221$	$0.076 \pm 0.414$
PopOp	$0.046 \pm 0.215$	$0.092 \pm 0.368$
PodOp	$0.045 \pm 0.211$	$0.092 \pm 0.368$
<i>SCAIR</i>	$0.003 \pm 0.010$	$0.016 \pm 0.006$

### 7.1 Queuing Time

In relative terms, we observe that SCAIR outperforms the baselines for both the queuing time and utility in both theme parks. SCAIR is able to maintain a low queuing time with different time budgets, while the baseline’s queuing time increases with the growth of time budget. The observation is consistent for both theme parks. Table 1 shows the ratio of queuing time and time budget of visitors. SCAIR produces a queuing time ratio that is 78.9% to 93.4% shorter than that of the baselines, across both DisHolly and Epcot theme parks.

### 7.2 Popularity

All four algorithms perform similarly for the DisHolly dataset, while PopOp, PodOp and SCAIR remain similar but outperform DisOp for the Epcot dataset. We observe that PodOp achieves a relatively high Popularity when time budget is equal to 180 min and 210 min. We observe that this phenomena is due to the special geographic distribution of the POIs in DisHolly, where the optimal path according to the algorithm includes two POIs that are remote from other POIs but yield very high popularity.

### 7.3 Utility

In terms of Utility, SCAIR outperforms all baselines consistently across all time budgets for both theme parks. The main contributing factor for this result is due to the much

improved queuing time performance that SCAIR is able to achieve, compared to the various baselines. In turn, the reduced queuing time leads to a higher utility score as tourists are able to utilize more of their time budget in visiting attractions, rather than spending excessive amount of time queuing.

## 8 Conclusion and Future Work

We now summarize the main findings of our work and discuss some possible directions for future research.

### 8.1 Conclusion and Discussion

Prior works on itinerary recommendation typically aim to make recommendations for the individual traveller and perform poorly in scenarios where multiple travellers use the same recommended itinerary, i.e., the Selfish Routing problem. In this paper, we introduced the crowd-aware itinerary recommendation problem and highlighted this Selfish Routing problem where all self-interested agents aim to maximize their own utility which result in sub-optimal social welfare. For example, when all travellers are recommended the same POIs with a short queuing time based on historical data, those POIs then become congested and suffer from a long queuing time.

To address this problem, we proposed the SCAIR algorithm that takes into consideration crowd behavior and addresses the NP-hard Social Welfare Optimization problem with a finite markov chains, which is in NC and can be solved in poly-logarithmic time. We performed a series of experiments using a theme park dataset. Experimental results show that SCAIR outperforms various competitive baselines in terms of a reduced queuing time and improved utility, while offering similar levels of popularity scores. An investigation into the effects of user arrival rates show that the performance of SCAIR remains competitive, compared to the various baselines, regardless of the arrival rates.

### 8.2 Future Work

We intend to further investigate models that further simulate real-life situations. For instance, we can also locate the entrances and exits of the theme parks to initialize and end paths; we could also use soft-max instead of one-hot to simulate the choices of paths which simulates the probabilistic decisions visitors make in real-life. We will also attempt to improve the formulation of the multi-objective optimization problem, such as by assessing the Pareto efficiency of the two objectives. It is also worthwhile to look into modifying our strategic recommendation algorithm and applying them to other game-theoretic environments, such as knowledge acquisition [20], crisis management [22] and career path planning [5,21]. Finally, we intend to look further into prior works, such as [13,27] to explore Machine Learning approaches in solving time-variant path planning problems and attempt to enhance the solution and simulation performance.

## 9 Acknowledgements

This research is funded in part by the Singapore University of Technology and Design under grants SRG-ISTD-2018-140 and PIE-SGP-AI-2020-01.

## References

1. Anagnostopoulos, A., Atassi, R., Becchetti, L., Fazzone, A., Silvestri, F.: Tour recommendation for groups. *Data Mining and Knowledge Discovery* **31**(5), 1157–1188 (2017)
2. Arora, S., Barak, B.: *Computational complexity: a modern approach*. Cambridge University Press (2009)
3. Braess, D.: Über ein paradoxon aus der verkehrsplanung. pp. 0042–0573. *Physica-Verlag* (1968), <https://doi.org/10.1007/BF01918335>
4. Choudhury, M.D., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R., Yu, C.: Automatic construction of travel itineraries using social breadcrumbs. In: *Proceedings of HT'10*. pp. 35–44 (2010)
5. Dave, V.S., Zhang, B., Al Hasan, M., AlJadda, K., Korayem, M.: A combined representation learning approach for better job and skill recommendation. In: *Proceedings of CIKM'18*. pp. 1997–2005. *ACM* (2018)
6. Garcia, I., Sebastia, L., Onaindia, E.: On the design of individual and group recommender systems for tourism. *Expert Systems with Applications* **38**(6), 7683–7692 (2011)
7. Gionis, A., Lappas, T., Pelechrinis, K., Terzi, E.: Customized tour recommendations in urban areas. In: *Proceedings of WSDM'14*. pp. 313–322 (2014)
8. Gunawan, A., Lau, H.C., Vansteenwegen, P.: Orienteering problem: A survey of recent variants, solution approaches and applications. *European Journal of Operational Research* **255**(2), 315–332 (2016)
9. Gunawan, A., Yuan, Z., Lau, H.C.: A mathematical model and metaheuristics for time dependent orienteering problem. In: *Proceedings of PATAT'14*. pp. 202–217 (2014)
10. Herzog, D., Sikander, S., Worndl, W.: Integrating route attractiveness attributes into tourist trip recommendations. In: *Companion Proceedings of WWW'19*. pp. 96–101 (2019)
11. Hu, F., Huang, X., Gao, X., Chen, G.: Agree: Attention-based tour group recommendation with multi-modal data. In: *Proceedings of DASFAA'19*. pp. 314–318 (2019)
12. Koutsoupias, E., Papadimitriou, C.: Worst-case equilibria. In: *Proceedings of STACS'99*. pp. 404–413 (1999)
13. Lees-Miller, J., Anderson, F., Hoehn, B., Greiner, R.: Does wikipedia information help netflix predictions? In: *Proc. of ICMLA'08*. pp. 337–343. *IEEE* (2008)
14. Leung, K.W.T., Lee, D.L., Lee, W.C.: Clr: a collaborative location recommendation framework based on co-clustering. In: *Proceedings of SIGIR'11*. pp. 305–314 (2011)
15. Li, X., Cong, G., Li, X.L., Pham, T.A.N., Krishnaswamy, S.: Rank-geofm: a ranking based geographical factorization method for point of interest recommendation. In: *Proceedings of SIGIR'15*. pp. 433–442 (2015)
16. Lim, K.H., Chan, J., Karunasekera, S., Leckie, C.: Personalized itinerary recommendation with queuing time awareness. In: *Proceedings of SIGIR'17*. pp. 325–334. *ACM* (2017)
17. Lim, K.H., Chan, J., Karunasekera, S., Leckie, C.: Tour Recommendation and Trip Planning using Location-based Social Media: A Survey. *Knowledge and Information Systems* **60**(3), 1247–1275 (2019)
18. Lim, K.H., Chan, J., Leckie, C., Karunasekera, S.: Personalized trip recommendation for tourists based on user interests, points of interest visit durations and visit recency. *Knowledge and Information Systems* **54**(2), 375–406 (2018)

19. Lim, K.H., Wang, X., Chan, J., Karunasekera, S., Leckie, C., Chen, Y., Tan, C.L., Gao, F.Q., Wee, T.K.: PersTour: A personalized tour recommendation and planning system. In: Extended Proceedings of HT'16 (2016)
20. Liu, J., Lim, K.H.: Self-evolving adaptive learning for personalized education. arXiv preprint arXiv:2005.02164 (2020)
21. Liu, J., Ng, Y.C., Wood, K.L., Lim, K.H.: Ipod: An industrial and professional occupations dataset and its applications to occupational data mining and analysis. arXiv preprint arXiv:1910.10495 (2019)
22. Liu, J., Singhal, T., Blessing, L.T., Wood, K.L., Lim, K.H.: Crisisbert: a robust transformer for crisis classification and contextual crisis embedding. arXiv preprint arXiv:2005.06627 (2020)
23. Martello, S., Pisinger, D., Toth, P.: Dynamic programming and strong bounds for the 0-1 knapsack problem. *Management Science* **45**(3), 414–424 (1999)
24. Nguyen, T.T., Roos, M., Rothe, J.: A survey of approximability and inapproximability results for social welfare optimization in multiagent resource allocation. *Annals of Mathematics and Artificial Intelligence* **68**(1-3), 65–90 (2013)
25. Osborne, M.J., Rubinstein, A.: A course in game theory. MIT press (1994)
26. Padia, P., Lim, K.H., Chan, J., Harwood, A.: Sentiment-Aware and Personalized Tour Recommendation. In: Proceedings of BigData (2019)
27. Palumbo, E., Rizzo, G., Troncy, R., Baralis, E.: Predicting your next stop-over from location-based social network data with recurrent neural networks. In: RecTour@ RecSys (2017)
28. Papadimitriou, C.H.: Algorithms, games, and the internet. In: ICALP'01. pp. 1–3 (2001)
29. Papadimitriou, C.H., Tsitsiklis, J.N.: The complexity of markov decision processes. *Mathematics of operations research* **12**(3), 441–450 (1987)
30. Pas, E.I., Principio, S.L.: Braess' paradox: Some new insights. pp. 265–276 (1997)
31. Piliouras, G., Nikolova, E., Shamma, J.S.: Risk sensitivity of price of anarchy under uncertainty. *ACM Trans. Econ. Comput.* **5**(1) (2016)
32. Roughgarden, T.: Selfish routing and the price of anarchy, vol. 174. MIT press Cambridge (2005)
33. Tsiligrirides, T.: Heuristic methods applied to orienteering. *Journal of the Operational Research Society* **35**(9), 797–809 (1984)
34. Vansteenwegen, P., Souffriau, W., Oudheusden, D.V.: The orienteering problem: A survey. *Euro. J. of Operational Rsch.* **209**(1), 1–10 (2011)
35. Wakamiya, S., Siriaraya, P., Zhang, Y., Kawai, Y., Aramaki, E., Jatowt, A.: Pleasant route suggestion based on color and object rates. In: Proceedings of WSDM'19. pp. 786–789 (2019)
36. Wang, X., Leckie, C., Chan, J., Lim, K.H., Vaithianathan, T.: Improving personalized trip recommendation to avoid crowds using pedestrian sensor data. In: Proceedings of CIKM'16. pp. 25–34 (2016)
37. Yao, L., Sheng, Q.Z., Qin, Y., Wang, X., Shemshadi, A., He, Q.: Context-aware point-of-interest recommendation using tensor factorization with social regularization. In: Proceedings of SIGIR'15. pp. 1007–1010 (2015)
38. Ye, M., Yin, P., Lee, W.C., Lee, D.L.: Exploiting geographical influence for collaborative point-of-interest recommendation. In: Proceedings of SIGIR'11. pp. 325–334 (2011)
39. Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, N.M.: Time-aware point-of-interest recommendation. In: Proceedings of SIGIR'13. pp. 363–372 (2013)
40. Zhang, C., Liang, H., Wang, K.: Trip recommendation meets real-world constraints: Poi availability, diversity, and traveling time uncertainty. *ACM TOIS* **35**(1), 5 (2016)
41. Zhang, C., Liang, H., Wang, K., Sun, J.: Personalized trip recommendation with poi availability and uncertain traveling time. In: Proceedings of CIKM'15. pp. 911–920 (2015)
42. Zhang, Y., Siriaraya, P., Wang, Y., Wakamiya, S., Kawai, Y., Jatowt, A.: Walking down a different path: route recommendation based on visual and facility based diversity. In: Companion Proceedings of WWW'18. pp. 171–174 (2018)