GroupTravel: Customizing Travel Packages for Groups

Sihem Amer-Yahia, Shady Elbassuoni, Behrooz Omidvar-Tehrani, Ria Mae Borromeo, Mehrdad Farokhnejad

GroupTravel: Customizing Travel Packages for Groups

Sihem Amer-Yahia[1], Shady Elbassuoni[2], Behrooz Omidvar-Tehrani[1],
Ria Mae Borromeo[3], Mehrdad Farokhnejad[1]
[1]Univ. Grenoble Alpes, CNRS, LIG (France), [2]The American University in Beirut (Lebanon),
[3]The University of the Philippines (Philippines)
[1]firstname.lastname@univ-grenoble-alpes.fr, [2]se58@aub.edu.lb, [3]riamae@gmail.com

ABSTRACT

We present GroupTravel, a framework that generates customized travel packages (TPs) for a group of individuals. GroupTravel implements different consensus functions proposed in group recommendation to reach agreement among members. Given a group whose members provide a travel query, GroupTravel returns k Composite Items (CIs) of Points Of Interest (POIs) that are *valid*, *representative*, *cohesive* and *personalized*. Validity is achieved by satisfying the query expressed by the group. Representativity ensures good coverage of a city. Cohesiveness reflects geographic proximity of POIs forming a CI. Personalization is achieved by choosing POIs that best match the travel preferences of group members. Additionally, group members can interact with generated TPs to customize them. With extensive synthetic experiments and user studies, we examine the benefit of personalization and the impact of different group consensus on user satisfaction. We also show that providing the ability to interact with TPs and reflecting that in the consensus yields better TPs.

1 INTRODUCTION

The ability to generate a travel package (TP) that best fits a traveler's profile is a longstanding problem that has been studied for years (e.g., [1–3]). In this work, we develop GroupTravel, a framework that generates TPs for a group of individuals traveling together. A TP is a set of k Composite Items (CIs), each of which is formed by Points of Interest (POIs) in a city. GroupTravel personalizes TPs based on a group profile that is computed as an aggregation of its members' preferences. GroupTravel allows travelers in a group to further customize the proposed TP via interaction. GroupTravel extends recent work that focused on a single traveler at a time [4] to reach consensus between multiple travelers [5, 6].

Travel Packages as Composite Items. A TP is a set of k CIs. CIs are useful in planning a city tour, selecting books for a reading club, or organizing a movie rating contest [1, 7–15]. Each CI satisfies a query that specifies desired POI categories and a budget constraint [13]. An objective function is defined to build a TP containing k *valid*, *representative*, *cohesive*, and *personalized* CIs. Validity ensures that each CI satisfies the query. Representativity enforces that the k CIs "cover" the city. Cohesiveness forms CIs containing geographically close POIs. Finally, personalization ensures that the CIs contain POIs that match the members' travel preferences. We use a fuzzy clustering algorithm to find the best k CIs forming a TP.

For example, a group wishing to visit Paris may request a TP consisting of five CIs, one per day. The group specifies a query which dictates CI validity: each CI must contain an accommodation, a restaurant, three attractions and one transportation mode, and such that the overall cost of visiting POIs in a CI is no more than \$100. Figure 1 shows the TP returned by GroupTravel. Each CI contains a set of co-located POIs that can be visited in one day. In addition, the TP formed by the set of 5 CIs, provides a good coverage of Paris and the POIs in each CI match the group members' travel preferences.

In this paper, we make three contributions. We formalize the problem of building a TP for a group of travelers. We define how group members can interact with the generated TP to further customize it. We run synthetic experiments and user studies to validate GroupTravel's effectiveness in generating a satisfying TP for groups.

First contribution: TPs for groups. GroupTravel takes as input a query and individual travel profiles. It outputs a *personalized* TP for the group. The preference of a group for an item must *reflect the degree to which the item is preferred by all group members.* The group preference must also *capture the level at which members disagree or agree with each other.* All other conditions being equal, an item that draws high agreement should have a higher score than an item with a lower overall group agreement. The different ways of aggregating *group preference* and *group disagreement* result in different *group consensus methods* ranging from *average preference*, to *least misery* and *disagreement-based* methods [16–18]. We leverage those definitions to generate a group travel profile from individual preferences.

Second contribution: interactive TPs. In [4], we examined the benefits of letting a user interact with a TP to customize it. The rationale is that even though the k CIs are valid, representative, cohesive, and personalized, a user may still want to intervene after seeing the travel options available in a city. We showed that providing interaction primitives to the user enabled expressing additional contextual preferences such as exploring some neighborhoods in a city or requesting a new CI which contains a specific POI. In this paper, we examine the benefit of interactivity in GroupTravel, i.e., for a group of individuals traveling together. To achieve that, we define the impact of each operation on a TP and on the profile of a group.

Figure 2 illustrates the flow of GroupTravel. Given a group of travelers and a consensus function, a group profile is generated from individual profiles. Our fuzzy clustering algorithm admits a geographic region (e.g., a city), a query and the group profile. It generates a TP that is shown to the travelers who can modify CIs, delete CIs, or generate new CIs. This interaction is reflected in the group's profile by updating the overall group preferences according to the

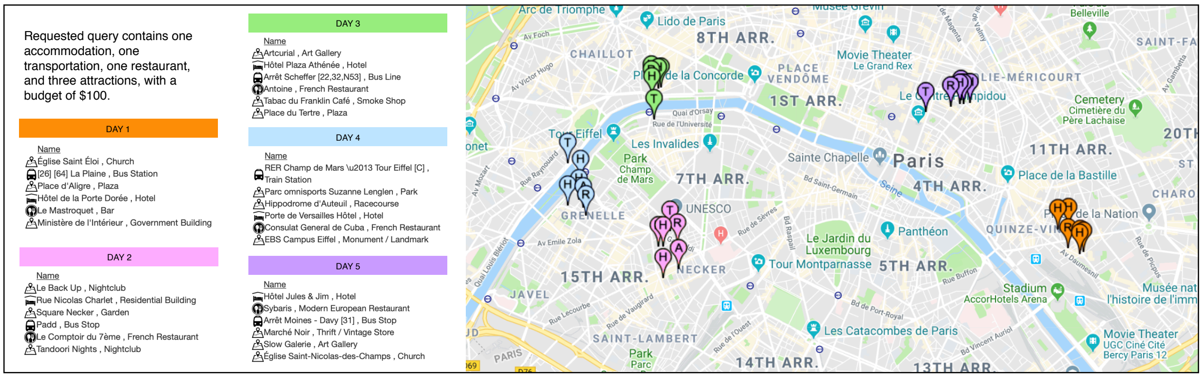

Figure 1: A 5-day travel package (TP) in Paris consisting of 5 Composite Items (CIs) of POIs for the group query $\langle 1\ \textit{accommodation}, 1\ \textit{transportation}, 1\ \textit{restaurant}, 3\ \textit{attractions}, \$100\rangle$**. Letters A, T, R, and H on POIs represent categories of accommodation, transportation, restaurant, and attraction, respectively.**

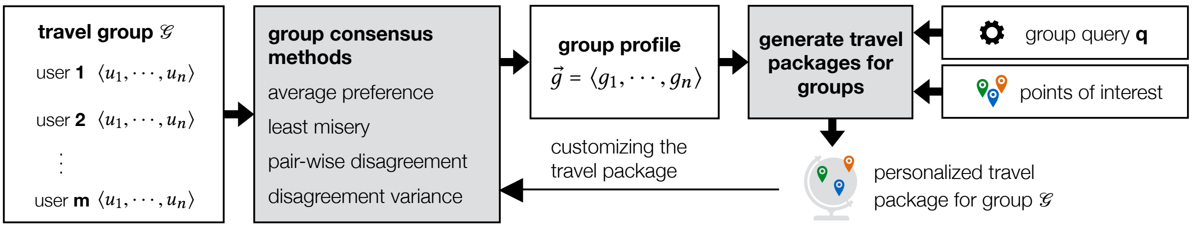

Figure 2: GroupTravel framework

requested changes. The new group profile can then be used to generate other TPs in the same or in a different city and test the "robustness" of the updated profile across cities.

Third contribution: experiments. The purpose of our experiments is two-fold: (1) study the utility of consensus functions from group recommendation in the context of GroupTravel, and (2) examine the benefit of interactive customization for groups. We run two extensive sets of experiments. In the first, we generate synthetic data to examine the relationship between group characteristics (group size, agreement between members, and consensus methods) and optimization dimensions (representativity, cohesiveness, and personalization). In the second, we run an extensive user study with real users from Figure-Eight[1] and Amazon Mechanical Turk[2]. Our study evaluates the usefulness of GroupTravel by asking actual users about their satisfaction with TPs before and after customization.

Our findings extend previous work for single travelers [4] and in group recommendation [5, 6]. In particular, we find that customization makes travel profiles more robust. Additionally, we find that disagreement-based consensus performs best in terms of all optimization dimensions, and for all different group variants (uniform and non-uniform as well as small, medium and large). Least misery, on the other hand, is more successful at satisfying the median user in larger groups with diverse tastes.

We also observe that TPs for non-uniform groups are more cohesive than TPs for uniform groups. This result generalizes previous work where a tension between personalization and cohesiveness was observed for individual users: the more personalized a TP is, the less likely it is to be cohesive, and vice versa [4]. Non-uniform groups contain members with diverse preferences. This diversity dilutes personalization (the aggregated profile expresses lower preferences than individual profiles). Given that, cohesiveness is likely to be higher for non-uniform groups. Similarly, the cohesiveness of uniform groups increases with group size, while their personalization decreases.

Our user study validates our objective function by showing that personalized TPs perform well and are liked better than non-personalized and random TPs. We also find that TPs obtained using average preference and least misery are best for uniform groups, whereas TPs obtained using disagreement-based methods are best for non-uniform groups. Similarly, incorporating inter-member disagreements is shown to be the best way to reach a consensus within diverse groups.

The paper is organized as follows. In Section 2, we describe our data model. Our approach for building group travel packages and interacting with them is described in Section 3. Experiments are reported in Section 4. The related work is reviewed in Section 5. We conclude with a summary of our work and a discussion of future work in Section 6.

[1] *http://www.figure-eight.com/*
[2] *https://www.mturk.com/*

i.id	i.name	i.cat	i.coordinates	i.type	i.tags	i.cost
1	Le Burgundy	*acco*	⟨48.8679, 2.3256⟩	hotel	*luxury suites cognac champagne bar gastronomic restaurant spa*	3.00
2	The Bicycle Store	*trans*	⟨48.8642, 2.3658⟩	bike shop	*accessoires vélo beach cruiser bicycle paris fixed gear*	2.71
3	Un Zèbre à Montmartre	*rest*	⟨48.886, 2.3348⟩	french	*bankers bar brunch café comedy fireplace frat hipsters liquor margaritas*	3.20
4	Les Arts Décoratifs	*attr*	⟨48.8632, 2.3334⟩	museum	*arts contemporary decorative exhibition fashion gallery mode modern museum*	3.86

Table 1: Sample Points Of Interest in Paris

2 DATA MODEL

2.1 Items

Our travel packages are built using Points Of Interest (POIs) in a city. Table 1 shows sample POIs in Paris. In our experiments, we use the TourPedia dataset.[3] It consists of POIs in eight cities which are divided into four main categories (*cat* for short): (1) accommodation (*acco*), (2) transportation (*trans*), (3) restaurant (*rest*) and (4) attraction (*attr*). Each POI or item i has a unique id, a name, a longitude and a latitude. To be able to set the rest of the attributes for the items in our dataset, we augment it with additional information extracted from Foursquare.[4] Using the Foursquare API, we retrieved the type of each item i. For instance, in the case of an accommodation item i, $i.type$ will be set to either a hotel, a hostel, a motel, a college residence hall, etc. Similarly, for a transportation item i, we set $i.type$ to its transportation mode which can be a tram station, a train station, a car rental, a bike rental, and so on. To set $i.tags$ for an item i, we retrieved all the tags provided by users on Foursquare for item i. Finally, there are may ways of setting the cost of visiting an item i or $i.cost$ including declarative data.

2.2 User Profile

Our goal is to provide a group of users with personalized travel packages. To do that, we build a group travel profile which captures the preferences of group members for different types of POIs. We start by defining a single-user profile and then explain how we aggregate the profiles of different users to generate a group profile.

Each user u is associated with a profile for each POI category c (i.e., *acco*, *trans*, *rest* or *attr*), which is a vector defined as follows:

$$\vec{u} = \langle u_1, \cdots, u_n \rangle$$

where n is the number of different POI types in category c and each $u_j, 1 \leq j \leq n$ is a score between 0 and 1.

To simplify our notation, c does not appear in $\vec{u}$. One way to set the vector $\vec{u}$ is to ask the user to state her preferences for the different types of POIs. In case the POI types are not sufficient to capture all the dimensions of travel preferences for users, we can try to learn these other dimensions from the data. For accommodation and transportation, the types are well-defined (e.g., Bicycle, Bus, Tram for transportation, and Hotel, Hostel, Resort for accommodation). For restaurants and attractions, we leverage their tags to capture information such as cuisine and ambiance for restaurants, or type and entrance fee for attractions. Particularly, we rely on Latent Dirichlet Allocation (LDA) applied to tags to identify latent topics for restaurants and attractions [19]. This results in several types such as "art gallery, museum, library" and "garden, park, event hall" for attractions, and "Japanese, sushi" and "beer, wine, bistro" for restaurants.

To set the individual components of the vector $\vec{u}$, we do the following. For the case of transportation and accommodation, we ask the user to provide a rating r_j between 0 and 5 for each accommodation or transportation type s_j. Similarly, for restaurants and attractions, we ask the user to provide a rating r_j between 0 and 5 for each latent topic s_j where each topic is represented by representative tags. Finally, we set the score u_j in the user profile as the normalized rating over all types or topics, i.e.,

$$u_j = \frac{r_j}{\sum_{k=1}^{n} r_k}$$

2.3 Group Profile

Similar to users, a group of users $\mathcal{G}$ is associated with a group profile for each POI category c (i.e., *acco*, *trans*, *rest* or *attr*), which is a vector defined as follows:

$$\vec{g} = \langle g_1, \cdots, g_n \rangle$$

where n is the number of different POI types in category c, and each $g_j, 1 \leq j \leq n$ is a score between 0 and 1. The value g_j reflects the preference of group $\mathcal{G}$ for a POI type by aggregating the preferences of group members.

To compute each g_j, we need to aggregate the preferences u_j of each user $u \in \mathcal{G}$. To do so, we leverage *consensus functions* that were previously proposed in the context of group recommendation [17, 18]. A consensus function is used to aggregate two components: *group preference* and *group disagreement*. Intuitively, to compute g_j, we need to *reflect the degree to which the j^{th} POI type in a given category is preferred by all group members*, and *capture the level at which members disagree or agree with each other about the j^{th} POI type in a given category.* All other conditions being equal, a POI that draws high agreement should have a higher score than a POI with a lower overall group agreement. We revisit the definitions we introduced in [16] to compute group consensus as a combination of group preference and group disagreement.

Group preference. The degree to which the j^{th} POI type in a given category is preferred by all group members is denoted p_j, and is computed using one of two common preference aggregation functions:

[3] *http://tour-pedia.org/about/*
[4] *https://foursquare.com/*

(1) *Average Preference:* $p_j = \frac{1}{|\mathcal{G}|} \sum_{u \in \mathcal{G}} u_j$

(2) *Least-Misery Preference:* $p_j = \min_{u \in \mathcal{G}} u_j$

Group disagreement. The level at which members disagree or agree with each other about the j^{th} POI type in a given category is denoted d_j, and is computed using one of two common disagreement computation functions:

(1) *Average Pair-wise Disagreement:*
$d_j = \frac{2}{|\mathcal{G}|(|\mathcal{G}|-1)} \sum_{u,v \in \mathcal{G}} (|u_j - v_j|)$,

(2) *Disagreement Variance:*
$d_j = \frac{1}{|\mathcal{G}|} \sum_{u \in \mathcal{G}} (u_j - \mu_j)^2$ where $\mu_j = \frac{1}{|\mathcal{G}|} \sum_{u \in \mathcal{G}} u_j$

The average pair-wise disagreement function computes the average of pair-wise differences in individual preferences for the j^{th} POI type among group members, while the variance disagreement function computes the mathematical variance of individual preferences. Intuitively, the closer the preferences between users u and v, the lower their disagreement.

Group consensus. We are now ready to compute a single group consensus score g_j for the j^{th} POI type in a given category. We do that by combining group preference and disagreement as follows:

$$g_j = w_1 \times p_j + w_2 \times (1 - d_j)$$

where $0 \leq w_1, w_2 \leq 1$, $w_1 + w_2 = 1$, and they specify the relative importance of preference and disagreement in the overall group consensus, respectively. We hence have four possible consensus functions that combine preference and disagreement to compute a single score g_j in the group profile.

Example. Consider a family (a couple with three kids) which forms a travel group $\mathcal{G}$ of size 4. Their preferences for visiting museums are 0.8, 1.0, 0.6, and 0.2, for the father, mother, the teenage child, and the kid, respectively, where 1.0 reflects the highest preference. Using average preference method, the group preference for this POI type is $p = 0.65$. However the group preference towards museums gets as low as 0.2 using the least misery method. Least misery favors the most unhappy user in the group, hence the preference of the kid dominates others'. On the other hand, the average pair-wise disagreement between $\mathcal{G}$'s members is $d = 0.43$. Also, the disagreement variance is $d = 0.088$. Given $w_1 = 0.5$ (hence $w_2 = 0.5$), $\mathcal{G}$'s consensus for museums is $g = 0.61$ by considering average preference and average pair-wise disagreement as the group preference and group disagreement components, respectively.

3 BUILDING GROUP TRAVEL PACKAGES

In this section, we define and solve the problem of building personalized travel packages for groups. We start by introducing Composite Items (CIs) and Travel Packages (TPs), and formulate building travel packages as a fuzzy clustering problem. We then discuss how groups can interact with travel packages to customize them using GroupTravel.

3.1 Composite Items

A Composite Item is a set of POIs of different categories. To be able to define what constitutes a CI, we rely on a group query which is a vector defined as follows:

$$\vec{q} = \langle \#c_1, \cdots, \#c_m, B \rangle$$

where m is the number of POI categories (4 in our dataset), $\#c_j, 1 \leq j \leq m$ specifies the number of items for POI category c_j, and B is a total budget.

A query indicates which categories of POIs, and how many of them, should constitute a CI. For example, the query $\vec{q} = \langle 1\ acco, 1\ trans, 2\ rest, 1\ attr, \$120 \rangle$ represents a CI with 1 accommodation, 1 transportation, 2 restaurants and 1 attraction for a daily budget of \$120.

The query is used to define valid CIs as follows. Given a set of items $\mathcal{I}$ and a query $\vec{q} = \langle \#c_1, \cdots, \#c_m, B \rangle$, a valid $CI \subseteq \mathcal{I}$ is a set of items such that (1) their categories correspond to the requested categories in the group query, and (2) the total budget of items forming the CI is at most B, i.e.,

$$\begin{cases} (i)\ \forall j \in \{1, \cdots, m\},\ \sum_{i \in CI} \mathbb{1}(i.cat, c_j) = \#c_j \\ (ii)\ \sum_{i \in CI} i.cost \leq B \end{cases}$$

where $\mathbb{1}$ is an indicator function which is equal to 1 if the category of item i is c_j and 0 otherwise. We refer to the set of valid CIs as $\mathcal{V}$.

3.2 Travel Packages

We are now ready to define the notion of a group travel package and formulate building travel packages as a fuzzy clustering problem. Given a group $\mathcal{G}$, a set of items $\mathcal{I}$, and a query $\vec{q}$, we define a group travel package as a set of k Composite Items $\mathsf{TP} = \{CI_1, CI_2, \cdots, CI_k\}$ where each $CI_j \subseteq \mathcal{I}$ is a valid Composite Item.

A travel package is formed by *valid* and *cohesive* CIs that are *representative* of the set of available items in the city. The validity of a CI is expressed in terms of a query $\vec{q}$ as defined in Section 3.1. Its cohesiveness must reflect how close the items forming a CI are to each other. The intuition is that each CI represents things to do in a given area of a city and must thus have POIs that are geographically close to each other. Finally, the representativity of a travel package serves the purpose of providing a good coverage of the city [13]. KFC, the algorithm that solves that problem in [13], relies on fuzzy clustering to position k centroids that "cover" the whole dataset. CIs are then formed in the vicinity of these centroids, which ensures that they provide a good summary of the dataset. In the context of this work, we may want to see a given item in different CIs. For example, a user's hotel could belong to multiple CIs. The same applies to a museum if the user wants to go back to the museum (as is the case for the "Louvre museum" in Paris that requires more than one visit). Contrary to *hard* clustering, *fuzzy* clustering allows each data point to participate in multiple clusters [20]. Thus, KFC is a natural choice for us to generate travel packages.

To be able to generate CIs, we define an item vector for each POI i as follows:

$$\vec{i} = \langle i_1, \cdots, i_n \rangle$$

where n is the number of types for the POI category that item i belongs to, and each i_j, $1 \leq j \leq n$ is a score between 0 and 1. The item vector $\vec{i}$ is set based on the category of the item i. For accommodation and transportation items (i.e., $i.cat = acco$ or $i.cat = trans$), we set i_j as follows:

$$i_j = \begin{cases} 1, & \text{if } i.type = t_j \\ 0, & \text{otherwise} \end{cases}$$

where t_j is the j^{th} type in the category that item i belongs to. For restaurants and attractions, the item vector $\vec{i}$ is set to the topic distribution vector for item i obtained from applying LDA.

To generate a personalized travel package TP for a group $\mathcal{G}$, we optimize the following objective function:

$$\begin{aligned} &\underset{M,W}{argmax}\ \alpha \sum_{j=1}^{k} \sum_{i \in \mathcal{I}} w_{ij}^{f}\ (1 - Euclidean(i, \mu_j)) + \\ &\sum_{j=1}^{k} \max_{CI_j \in V} \Bigg(\beta \sum_{i \in CI_j} (1 - Euclidean(i, \mu_j)) + \\ &\qquad\qquad \gamma \sum_{i \in CI_j} Cosine(\vec{i}, \vec{g}) \Bigg) \\ &s.t.\ \forall i \in \mathcal{I},\ \sum_{j=1}^{k} w_{ij} = 1 \end{aligned} \tag{1}$$

In Equation 1, we use a normalized geographic Euclidean distance between two items, and Cosine similarity between an item vector and the group profile vector for the category the item belongs to. Euclidean distance is an an approximation of Haversine calculations on a spherical space (to measure the distance in miles/kilometers between two latitudes and longitudes) with Equirectangular calculations on a Euclidean space to gain performance. This approximation makes sense for short distances within a city as we have experimentally observed that our performance gain is 30x with only 0.1% of precision loss. To obtain a normalized Euclidean distance, we divide all distance values by the largest observed distance value. $M = \{\mu_1, \mu_2, \cdots, \mu_k\}$ is a set of k centroids, W is a weight matrix of size $|\mathcal{I}| \times k$ which contains the w_{ij} weights indicating which item belongs to which cluster. α and β are user-dependent parameters controlling the weight of the optimization objectives, and $f \leq 1$ is the weighting exponent used in fuzzy clustering.

The two components of the objective function are inherited from KFC [13] and capture cohesiveness and representativity by choosing items i that are close to the centroid μ_j of each of the k clusters, where closeness is based on the geographic distance. Those components serve to identify cluster centroids μ_j that are representative of the complete dataset, while ensuring that the centroids obtained are close to some valid CI (i.e., $\in \mathcal{V}$). Maximizing the sum of the similarities of all items in a CI to its centroid additionally ensures the cohesiveness of the valid CI considered.

The last component, weighted with γ, captures personalization by comparing the similarity of the group's profile vector $\vec{g}$ to the item vector $\vec{i}$. This allows the algorithm to focus on producing CIs that are valid, cohesive and that contain personalized items that matter to the group, rather than any items.

3.3 Customizing Travel Packages

In order to customize travel packages, we provide group members with a GUI where all the CIs forming a travel package are displayed on an interactive map of the city. We define five atomic operations to allow groups to refine their preferences and produce customized TPs. Our operations are:

1. REMOVE(i,CI): remove POI i from Composite Item CI.
2. ADD(i,CI): add POI i to a Composite Item CI. The user can filter the POIs by category and type and the closest items to CI satisfying the user filter are displayed for the user to choose from.
3. REPLACE(i,CI): replace POI i in CI with another POI. In that case, the system recommends to the user the closest POI j in terms of geographic distance and such that $i.cat = j.cat$.
4. GENERATE(RECTANGL$E(x, y, w, h)$): generate a new CI that is centered in the area enclosed by a rectangle whose upper-left point is (x, y), and with width w and height h. The generated CI is both *valid* and *cohesive*.

Using the above set of operations, group members can customize the generated travel package until they are satisfied with it. For example, a member can drop or add a set of POIs in a given CI. She can also replace POIs with others that the system recommends to ensure that the CI remains as cohesive as possible. Finally, a group member can completely delete a CI by iteratively removing items in that CI until it is empty. Similarly, a group member can generate a new CI by selecting an area in the map. The group interactions with the CIs provide us with additional information about the group travel preferences. Particularly, they are useful in refining the group travel profile.

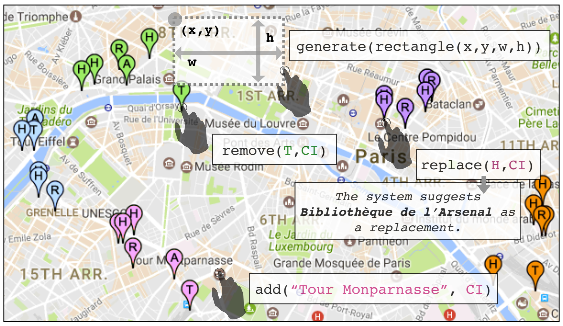

Figure 3: Customization operators

Figure 3 illustrates examples of customization operators in Paris. For instance, a REMOVE operator is requested to discard a bus stop in the area of "Invalides". It is also requested to ADD "Monparnasse tower" to the travel package as an attraction. In response to a REPLACE operator, the system suggests "Arsenal library" to replace "Pompidou library". Also a GENERATE operation is requested by defining an area from "L'église de la Madeleine" to "Palais Royal", where a potential attraction POI is "Place Vendôme".

Refining the group profile. The interactions of group members with the provided CIs serve as implicit feedback that can be used to update the group's travel profile. This refinement serves two purposes: *(1) make the group profile robust so that fewer interactions will be needed in the future including in other cities, (2) build long-lasting profiles for non-ephemeral groups.* We define two strategies for updating the group profile: individual and batch strategies. The *individual strategy* was defined for single travelers [4]. It first refines each group member profile based on that member's interactions with the TP, if that member customized the TP. It then aggregates all individual profiles into a new group profile. The *batch strategy* gathers interactions performed by all group members and directly refines the group profile. We describe the batch strategy that is a direct adaptation of the individual strategy.

Let $\vec{g}$ be the current group profile for POI category c. Furthermore, assume a group member added a set of POIs I^+ that belong to category c. Also, assume a group member removed a set of POIs I^- that belong to category c. Now the group vector $\vec{g}$ for category c can be updated as follows:

$$\vec{g} = \vec{g} + \vec{g}^+ - \vec{g}^-$$

where

$$\vec{g}^+ = \frac{1}{|I^+|} \sum_{i \in I^+} \vec{i}$$

and $\vec{i}$ is the item vector of item i as defined in Section 3.2.

The value $\vec{g}^-$ will be set the exact same way as $\vec{g}^+$ by replacing I^+ with I^- above. Finally, if any of the components of the updated vector $\vec{g}$ falls below 0, the value of this component will be set to 0.

4 EXPERIMENTS

We provide two sets of experiments. First, we generate synthetic data to examine the relationship between group characteristics (group size, uniformity, and consensus methods) and optimization dimensions (representativity, cohesiveness, and personalization). As we do not recruit real participants in this experiment, we focus on dissecting the objective function of GroupTravel. In the second experiment, we describe our user study which evaluates the usefulness of GroupTravel by asking group members about their satisfaction with the generated travel packages, before and after customization.

4.1 Setup

Group composition. We build groups by aggregating profiles of individual users. In our synthetic experiment, user profiles are generated at random. In our user study, user profiles capture the travel preferences of the participants.

We form different groups by varying their *size* (the number of users in a group) and *uniformity*. Intuitively, a group is more uniform if its members have similar preferences with respect to POI types. The uniformity of a group $\mathcal{G}$ is a value between 0 and 1 and is computed as the average pair-wise Cosine similarity between profile vectors of all $\mathcal{G}$'s members, i.e.,

$$uniformity(\mathcal{G}) = \frac{2}{|\mathcal{G}||\mathcal{G}-1|} \sum_{u,v \in \mathcal{G}} Cosine(\vec{u}, \vec{v})$$

We consider three categories of group sizes, i.e., *small* groups having 5 members, *medium* groups having 10 members, and *large* groups having 100 members. We also consider two categories of group uniformity, i.e., *uniform* groups having a uniformity value larger then 0.85, and *non-uniform* groups having a uniformity value smaller than 0.20.

Group consensus. Recall from Section 2.3 that for a group $\mathcal{G}$, we aggregate the individual user profiles to generate a group profile by applying a group consensus as follows:

$$g_j = w_1 \times p_j + w_2 \times (1 - d_j)$$

where p_j represents a group preference for POI type j and d_j represents group disagreement for POI type j. We employ the following variants of group consensus in our experiments:

- *Average preference*, where p_j is *average preference* and $w_1 = 1.0$ (i.e., group disagreement is not considered).
- *Least misery*, where p_j is *least misery* and $w_1 = 1.0$ (i.e., again group disagreement is not considered).
- *Average preference with average disagreement*, where p_j is *average preference*, d_j is *average pair-wise disagreement* and $w_1 = 0.5$. Hereinafter, we call this method "pair-wise disagreement" for simplicity.
- *Average preference with disagreement variance*, where p_j is *average preference*, d_j is *disagreement variance* and $w_1 = 0.5$. Hereinafter, we call this method "disagreement variance" for simplicity.

4.2 Optimization dimensions

Once a TP is computed for a group, we measure each component of our optimization objective, representativity, cohesiveness and personalization (Section 3.2).

Representativity measures the collective coverage of POIs in a TP over a region of interest, e.g., a city. The farther CIs in a TP are from each other, the higher the TP's representativity. Representativity is measured as follows:

$$representativity(\mathsf{TP}) = \sum_{l=1}^{k} \sum_{j=l}^{k} Euclidean(\mu_l, \mu_j) \tag{2}$$

where μ_l is the centroid of the composite item C_l, and $Euclidean(\mu_l, \mu_j)$ measures the Euclidean distance between the centroids μ_l and μ_j. Recall from Section 3.2 that the Euclidean distance is an an approximation of Haversine calculations.

Cohesiveness measures the geographical compactness of CIs in a TP, i.e., how close the POIs in a CI are to each other. It is measured as follows:

$$cohesiveness(\mathsf{TP}) = \mathcal{S} - \left(\sum_{CI \in \mathsf{TP}} \sum_{i,j \in CI} Euclidean(i,j) \right) \tag{3}$$

where $Euclidean(i, j)$ measures the Euclidean distance between the geographical coordinates of POIs i and j. The constant $\mathcal{S}$ defines the maximum possible sum of distances over CIs in a TP. In our synthetic experiment, we set $\mathcal{S} = 221.79$ as the largest observed value for aggregated distances.

While representativity and cohesiveness evaluate TPs in a geographical domain, personalization evaluates them in terms of preferences (using the profile vector $\vec{g}$ for the group $\mathcal{G}$). Personalization is measured as follows:

$$personalization(\mathsf{TP}, \mathcal{G}) = \sum_{CI \in \mathsf{TP}} \sum_{i \in CI} Cosine(\vec{i}, \vec{g}) \tag{4}$$

		average preference			least misery			pair-wise disagreement			disagreement variance		
		R	C	P	R	C	P	R	C	P	R	C	P
uniform groups	small	100%	69%	95%	38%	0%	74%	100%	74%	99%	99%	79%	100%
	medium	94%	70%	94%	75%	57%	73%	95%	77%	98%	96%	80%	98%
	large	85%	73%	72%	76%	76%	68%	96%	87%	97%	97%	84%	96%
non-uniform groups	small	17%	89%	75%	21%	76%	07%	98%	94%	98%	97%	90%	98%
	medium	25%	90%	83%	14%	76%	7%	98%	98%	99%	98%	94%	98%
	large	32%	96%	98%	13%	79%	00%	95%	100%	96%	95%	96%	93%

Table 2: Synthetic experiment for travel groups. Optimization dimensions are abbreviated as R for representativity, C for cohesiveness, and P for personalization.

4.3 Synthetic Data Experiment

Our goal in the synthetic data experiment is to examine the relationship between group characteristics and our optimization dimensions. Preferences of real people will be verified later in the user study (Section 4.4). Groups are characterized by their *uniformity*, i.e., similarity between members, *size*, and the *consensus function* used to aggregate individual preferences.

4.3.1 Setup. We describe how we generate user and group profiles and other settings in the synthetic experiments.

Group profiles. We generate user profiles in an independent roll-and-dice process. Each profile is a vector whose cells contain preference values for different types of POIs. We assign a random value between 0 and 1 to each cell in the user profile vector. A group $\mathcal{G}$ is a matrix of $|\mathcal{G}|$ user profiles, where $|\mathcal{G}| \in \{5, 10, 100\}$. For each combination of group size and group uniformity (uniform and non-uniform), we generated 100 different random groups. For each generated group, we computed a group profile using the four different consensus methods. As a result, we obtained 2400 distinct group profiles in total.

Query and objective function. We generate a TP for each group profile. Each TP contains exactly 5 CIs that are valid with respect to a default query, i.e., ⟨1 *acco*, 1 *trans*, 1 *rest*, 3 *attr*⟩. Also, we specify ***an infinite budget*** to ensure that all popular POIs are included in the CIs. Regarding the weights in our objective function (Equation 1), we always set $\gamma = 1.0$ for personalization, and we assign random values to α and β in the range $[0, 1]$ for representativity and cohesiveness, respectively, in order to prevent bias towards an optimization objective.

Optimization dimensions. For the TPs generated for group profiles, we report representativity, cohesiveness, and personalization as defined in Equations 2, 3, and 4, respectively. The values obtained for all dimensions are normalized in the range $[0, 1]$ in min-max style:

$$normalized\_value(o) = \frac{value(o) - min(o)}{max(o) - min(o)}$$

where $min(o)$ and $max(o)$ are the smallest and highest values of an optimization dimension o, respectively. Before normalization, the values of representativity, cohesiveness, and personalization were spread in the ranges $[0.03, 41.39]$, $[19.29, 221.79]$, and $[0.01, 0.16]$, respectively.

Validation of observations. We validate all our observations on optimization dimensions in terms of statistical significance using the One-way ANOVA procedure, with the $\mathcal{F}$-measure of MSB/MSE[5] and the significance level of $p = 0.05$. ANOVA results are reported as $\mathcal{F}(n, k) = x$ given $p < 0.05$, where n and k are the first and second degrees of freedom, respectively, and x is the value obtained for the $\mathcal{F}$-measure. Fully-independent generation of user and group profiles ensures that the $\mathcal{F}$-measure captures truly significant results.

We also compute the Pearson correlation coefficient (PCC) to validate linear correlations between attributes. PCC has a value between $+1$ and -1, where $+1$ reflects a totally positive linear correlation, 0 means no linear correlation, and -1 represents a totally negative linear correlation.

4.3.2 Summary of results. Table 2 reports the values of the optimization dimensions averaged over 100 generated groups. Overall, we observe that disagreement-based consensus functions, whether pair-wise or variance, perform best in terms of all optimization dimensions, and for all different group variants. Least misery appears to be the worst aggregation method. We also observe that TPs for non-uniform groups are more cohesive than uniform groups. However, the cohesiveness of uniform groups increases with group size, while their personalization decreases. We also note that there is a tension between personalization and cohesiveness where more personalized TPs are less likely to be cohesive.

Additionally, we report the similarity between the TP of a group and its median user (Table 3). Overall, we observe that the similarity decreases in larger groups. For non-uniform groups, the best similarity values are achieved using least misery, while for uniform groups, disagreement-based methods are superior.

4.3.3 Interpretation of results. We discuss the influence of consensus functions, group uniformity, and group size, on the optimization dimensions and the agreement between individuals and groups.

Influence of consensus functions. We observe in Table 2 that TPs are generally more personalized when their associated group profile is built using a disagreement-based consensus (variance disagreement and pair-wise disagreement). Least misery is the worst consensus method for personalization. This shows that optimizing towards one single group member is not an effective personalization strategy. Incorporating inter-member disagreements is therefore the

[5] *MSB: Mean Square Between, MSE: Mean Square Error*

		average preference			least misery			pair-wise disagreement			disagreement variance		
		R	C	P	R	C	P	R	C	P	R	C	P
uniform groups	**small**	99%	31%	93%	38%	98%	75%	98%	26%	99%	98%	20%	99%
	medium	86%	55%	85%	93%	66%	75%	87%	47%	99%	98%	45%	99%
	large	99%	71%	94%	98%	68%	71%	98%	57%	99%	97%	61%	99%
non-uniform groups	**small**	15%	73%	59%	83%	76%	64%	13%	74%	38%	14%	72%	37%
	medium	21%	67%	28%	83%	81%	88%	21%	59%	14%	21%	63%	14%
	large	5%	26%	2%	48%	44%	54%	2%	22%	1%	2%	23%	5%

Table 3: Agreement between median users and groups, where the value 100% represents the highest degree of agreement. The symbol R stands for representativity, C for cohesiveness, and P for personalization.

best way to obtain POIs that satisfy everyone in a group, regardless of group uniformity. We also observe that average preference and disagreement-based methods result in similar representativity values, validating that fuzzy clustering achieves good representativity overall by strategically placing centroids on a map.

Influence of group uniformity. We observe in Table 2 that TPs for non-uniform groups are always more cohesive than TPs for uniform groups. This result generalizes previous work where a tension between personalization and cohesiveness was observed for single-member groups: the more personalized a TP is, the less likely it is to be cohesive [4]. Non-uniform groups contain members with diverse preferences. This diversity makes personalization weaker (the aggregated profile expresses lower preferences than individual profiles). Given that, cohesiveness is likely to be higher for non-uniform groups.

Influence of group size. In Table 2, we observe that regardless of the consensus function, cohesiveness increases as uniform groups grow in size. The values of PCC are +0.98, +0.73, +0.73, and +0.99 for average preference, least misery, variance disagreement, and pair-wise disagreement, respectively. As groups grow in size, their uniformity decreases yielding a weaker personalization effect, which in turn favors cohesiveness. We also observe an inverse correlation between personalization and group size for uniform groups. The values of PCC are −0.99, −0.99, −0.89, and −0.89 for average preference, least misery, variance disagreement, and pair-wise disagreement, respectively. That is explained by the fact that in larger groups, the preferences of individuals fade out and personalization decreases. This is also compatible with the previous single-user study [4].

Agreement between individuals and groups. Table 3 reports the similarity of individuals (the median user in each group) and groups they belong to. The goal is to measure the *sacrifice of individuals when joining groups.* For this aim, we compute a median profile for each of the 600 previously generated groups. The sum of Cosine values between the profile of the median user u and all other members of u's group is the highest. We generate a TP for each median user and compute the optimization dimensions for that TP. Table 3 reports the similarity between the values of optimization dimensions of a group and its median user. The higher that value, the better it is from that median user's perspective.

A general observation is that group size and group uniformity play an important role. In large groups, preferences of individuals fade out and returned TPs are farther from the the median user's preferences. Concerning cohesiveness, the highest similarity is obtained with least misery. It is also the case for personalization. Least misery yields higher similarity between the median user and the group for non-uniform groups. Both findings are consistent with previous work on group recommendation [5, 6], where least misery is more successful at satisfying the median user in larger groups with diverse tastes. For uniform groups, disagreement-based methods are best for personalization.

4.4 User Study

The goal of our user study is to observe how GroupTravel helps users obtain and refine a TP when traveling with others. The study consists of two parts. First, we focus on personalization aspects of GroupTravel and compare personalized and non-personalized TPs together. Second, we shed light on customization and observe how enabling interaction with TPs and their refinement can help improve the group travel profile, which consequently means more satisfactory TPs.

4.4.1 Setup. We recruited 3000 participants for our user study. To ensure diversity, we gathered 2000 of them from Figure-Eight platform[6], and the remaining 1000 from Amazon Mechanical Turk[7]. After pruning profiles with invalid email addresses and/or identifiers, we retained 90.1% and 96.6% of the participants in the aforesaid platforms, respectively. We based our choice for the number of study participants on Equation 5, that uses the central limit theorem [21].

$$Sample\ size = \frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + (\frac{z^2 \times p(1-p)}{e^2 N})} \quad (5)$$

We describe the parameters in Equation 5 as follows.

- $N = 200,000$ is the population size, i.e., the number of contributors on Figure-Eight and Amazon Mechanical Turk platforms [22].
- e = 3% is the margin of error, i.e., the percentage of deviation in result in the sample size compared with the total population.
- z = 95% is the confidence level, i.e., if the job is repeated 100 times, 95 times out of 100 the result would lie within the margin of error.

[6] *http://www.figure-eight.com/*
[7] *https://www.mturk.com/*

		random	non personalized	average preference	least misery	pair-wise disagreement	disagreement variance
uniform	**small**	3.42	3.59	3.54	3.53	3.77	3.65
	medium	3.43	3.48	3.69	3.47	3.56	3.65
	large	3.52	3.58	3.72	3.62	3.78	3.70
non-uniform	**small**	3.01	2.68	3.28	3.28	3.23	3.19
	medium	3.01	2.94	3.17	3.19	3.21	3.26
	large	3.05	3.00	3.14	2.84	3.09	3.12

Table 4: Independent evaluation of user study

- $p = 50\%$ is the percentage value, i.e., the expected result value of the experiment. It is advised to put it at 50% when the result is not known.

Our sample size rounded up to at least 1062 participants based on the above formula. We recruited almost three times more participants to allow flexibility in forming groups and account for contributors who might quit the study before fully completing it.

We built travel profiles for the recruited participants by asking them to state their preferences on POI categories using Google Forms[8]. We then used the generated user profiles to build groups with varying characteristics, i.e., size and uniformity. For uniform groups, we generated 5 groups of each size (small, medium, and large). We gathered assessments from all members of small and medium groups, and from 30 random members for large groups. For non-uniform groups, we generated 3 groups for each size, and gathered assessments from all members of small and medium groups, and between 19 and 30 members for large groups. Each participant was paid \$0.01 for profile collection and \$0.50 for evaluating TPs.

4.4.2 Summary of results. Regarding personalization, we observe that participants liked personalized TPs more than non-personalized and random TPs. We also observe that TPs associated with average preference and least misery are the best performers for uniform groups, and TPs associated with disagreement-based methods are highly appreciated by members of non-uniform groups. Regarding customization, we noticed the supremacy of the batch strategy over the individual strategy in almost all cases.

4.4.3 Exploring personalization. In this part of the study, we aim to evaluate how satisfied users are with personalized TPs. We build personalized packages in the city of Paris. Similarly to the synthetic data experiment, each TP contains exactly 5 CIs that are valid with respect to a default query, i.e., $\langle 1\ acco, 1\ trans, 1\ rest, 3\ attr\rangle$. Also, we specify *an infinite budget* to ensure that all popular POIs are included in the CIs. We conduct two evaluations, *independent* and *comparative*.

Independent evaluation. We asked members of the formed groups to evaluate TPs. The TPs under evaluation were either *non-personalized*, or *personalized* with one of the four group consensus methods. Non-personalized TPs were generated by setting the weight of the personalization dimension to 0 in the objective function. In addition, to filter undesired participants, we injected a random TP which included *invalid* CIs, and discarded input from participants who preferred that TP (23 participants). For each TP out of the 6 TPs to be evaluated (random, non-personalized, and personalized with the four different consensus methods), we asked the remaining 326 participants to indicate their interest in visiting POIs in the TP *with other members of the group*, using a score between 1 and 5. A score of 1 means that there are very few POIs that the participant is interested in, and a score of 5 means that the participant is interested in almost all of the POIs. To prevent bias, we did not share with participants any details about the characteristics and members of the group they are involved in.

Table 4 illustrates the results of our independent evaluation. The average interest of participants who were not filtered out is reported for groups with different sizes and uniformity categories. The results validate our objective function (Equation 1), because they show that personalized TPs perform well and are liked better than non-personalized and random TPs. We also observe that scores for uniform groups remain fairly stable as groups become larger. However for non-uniform groups, scores decrease by group size. This is in-line with our findings in the synthetic experiment where the preferences of individual members in non-uniform groups fade out as groups grow in size, resulting in less-personalized TPs.

Comparative evaluation. We also presented the participants of our study with a pair of TPs among the 6 aforementioned TPs, and asked them to choose the one that they prefer the most, and to state a reason behind their choice. Table 5 reports results for each pair-wise comparison in terms of the percentage of supremacy. For instance, for small uniform groups, **AVTP** is preferred over **LMTP** in 48% of the time implying that **LMTP** is preferred over **AVTP** in 52% of the time. We observe that TPs associated with average preference and least misery (**AVTP** and **LMTP** in Table 5) are winners for uniform groups, whereas TPs associated with disagreement-based methods (**ADTP** and **DVTP**) are winners for non-uniform groups. This finding is in-line with previous work on group recommendation [5, 6], where recommendations aggregated with either average preference or least misery are preferred for uniform groups where preferences are homogeneous. Similarly, incorporating inter-member disagreements is shown to be the best way to reach a consensus within diverse groups.

We also reviewed the statements that the participants provided to justify their choices. First we focused on cases where personalized TPs are not preferred. In these cases, participants often justified their choice only as a tie-breaker: *"I like this TP a little more"*, *"I think this TP is a bit better"*. For uniform groups, participants mentioned that they prefer TPs with average preference and least misery because those

[8] *https://docs.google.com/forms*

		AVTP vs.				LMTP vs.			ADTP vs.		DVTP vs.
		LMTP	**ADTP**	**DVTP**	**NPTP**	**ADTP**	**DVTP**	**NPTP**	**DVTP**	**NPTP**	**NPTP**
uniform	**small**	48%	56%	64%	48%	56%	56%	64%	36%	36%	40%
	medium	56%	42%	42%	54%	62%	56%	70%	42%	56%	42%
	large	47%	52%	44%	50%	51%	54%	55%	46%	52%	55%
non-uniform	**small**	27%	27%	66%	60%	40%	73%	40%	60%	60%	47%
	medium	43%	73%	46%	53%	73%	66%	67%	46%	43%	67%
	large	54%	42%	48%	49%	51%	36%	42%	64%	57%	48%

Table 5: Comparative evaluation of user study. AVTP, LMTP, ADTP, and DVTP refer to personalized TPs obtained with average preference, least misery, average disagreement, and disagreement variance, respectively, and NPTP refers to the non-personalized TP.

TPs reflect their personal preferences better: *"this TP seems to me more interesting for my taste"*, *"with this TP, I can move around the city more"*. We observe the same type of statements for non-uniform groups in case of disagreement-based methods, e.g., *"there are types of places in this TP that I want to visit"*.

4.4.4 Exploring customization. In this experiment, we aim to validate the benefit of customization by allowing group members to interact with travel packages. More precisely, we want to assess if interacting with personalized TPs will refine the group profile in such a way that subsequent TPs are more relevant. To do so, we displayed personalized TPs on the map of Paris and asked participants to interact with the CIs forming those TPs by adding, removing, replacing POIs or generating new CIs (see the interface in Figure 3). We then refined the group profile based on the interactions of all group members. We compare the *individual* and *batch* strategies defined in Section 3.3.

With the refined group profile in Paris using either strategy, we built a customized travel package in a comparable city, namely Barcelona. We then asked our participants to evaluate the generated TP for Barcelona both in independent and comparative evaluations. Similar to the personalization study case in Section 4.4.3, to filter undesired participants, we injected a random TP which included *invalid* CIs, and discarded input from participants who preferred that TP. Participants were asked to rate the TPs in a scale of 1 to 5. A total of 18 workers participated in this study. That allowed us to build one uniform group with 11 members and one non-uniform group with 7 members. We recruited workers with an approval rate superior to 90%.

Independent evaluation. Participants in each group were asked to evaluate three different TPs in Barcelona: the first one was non-personalized, the second personalized and customized using the individual strategy, and the third personalized and customized with the batch strategy. Table 6 reports average ratings. Results are comparable across TPs. Overall, all travelers were equally satisfied with the POIs in all the TPs.

Comparative evaluation. Participants in each group were asked to compare a pair of TPs built from a non-personalized TP, a personalized and customized TP using the individual strategy, and a personalized and customized TP with the batch strategy. Table 6 reports results for each pair-wise comparison in terms of the percentage of supremacy. For instance, the batch strategy is preferred over the individual one in 82% of the time for uniform groups, implying that the individual strategy is preferred over the batch one only 18% of the time. The batch strategy is by far the best. That is particularly true for uniform groups. That is in-lined with the independent study. The intuition is that refining group profiles directly yields better TPs.

5 RELATED WORK

5.1 Itineraries and Personalization

The extraction of travel itineraries from Flickr photos was first proposed in [1] and their personalization in [2] to build customized city tours. Tailored itineraries are extracted from Flickr using an objective function that combines POI popularity with the actual user preferences over POI categories. This approach is not directly applicable to ours, since the personalization is merely a filtering of extracted trajectories. In our case, it is the POI composition itself that is personalized using the query and travel profile. That makes our problem computationally more challenging. Another difference is that unlike itineraries, POIs forming a CI are not ordered, and hence, their generation relies on a different model and algorithm (clustering instead of graph traversal). Finally, in this work we are also interested in generating travel packages for groups of users traveling together rather than just a single user.

The idea of learning travel packages was recently explored in [3]. This work proposed learning topics conditioned on both the tourists and the intrinsic features (i.e., locations and travel seasons) of landscapes. As a result, preferences on which landscapes to visit, in which season, and how to travel from one point to another (transportation modes), are extracted. This work does not propose interactive refinement of one's travel preferences and does not support group travel.

5.2 Composite Items

Composite retrieval was studied with different semantics in recent work [1, 7–10, 13–15, 23]. Most existing algorithms rely on a two-stage process that decouples the query (e.g., a CI must contain one museum and 2 restaurants) from the optimization goal (e.g., each CI is a set of close landmarks in a city). In [13], it was shown that an integrated approach produces better representative CIs than a two-stage approach. We hence build on that and extend it to build personalized CIs for groups.

TP type	uniform (11 members)	non-uniform (7 members)
individual	3.45	3.69
batch	2.91	3.8
non-personalized	3.37	3.83

Table 6: Independent evaluation of customized travel packages.

	batch vs.		individual vs.
	individual	non-personalized	non-personalized
uniform	82%	63%	54%
non-uniform	72%	57%	14%

Table 7: Comparative evaluation of customized travel packages.

5.3 Recommendation and Interactivity

Out of the multitude of itinerary recommendation approaches [24–27], only a few are interactive. In [24], a user provides feedback on the next set of POIs to visit, the system then recommends the best itineraries and further suggests new POIs, with optimal utility, to solicit feedback for. In GroupTravel, the system recommends POIs to replace those unwanted by a user or to form a CI with some selected POIs. We do not assume any prior knowledge about a city. Additionally, we focus on enabling group-based interaction with TPs.

MOBI is a collaborative itinerary planning framework [28]. Each user provides preferences and constraints in the form of "I want {at most, at least, exactly} [number] {activities, hours} of $\{cat_1, cat_2, \cdots, cat_n\}$", which resembles our query. Users interact with proposed itineraries and are told which constraints remain to be satisfied. In our work, users intervene in a second stage to refine the package. We have shown that helping bootstrap travel package construction is preferred as it induces fewer interactions.

Finally, in our previous work [4], we studied the benefit of interactivity in the generation of customized travel packages for a single user. We found that while personalization helps the selection of relevant POIs to include in a travel package, customization is necessary to allow users to customize their travel packages as they explore the alternatives the city has to offer. Customization has also been shown to help refine users' profiles based on their interactions with travel packages. In this current work, we extend the approaches and techniques we proposed for generating customized travel packages to support group travel.

6 CONCLUSION

We develop GroupTravel, a framework that generates travel package for groups. GroupTravel aggregates individual preferences into a single group preference using consensus methods developed for group recommendation. GroupTravel relies on a fuzzy clustering algorithm to generate k valid, representative, cohesive and personalized composite items that form a travel package. Travelers can interact with the generated travel package to further customize it. We run extensive synthetic and real data experiments and show that our findings are consistent with previous work in generating travel packages for single users [4] and in group recommendation [5, 6].

This work opens several research directions. One immediate challenge is to incorporate different collaboration models into the primitives used to interact with the TPs. We are examining different models such as the star model where a designated traveler moderates all requests from others in the same group, the sequential model where a TP is customized in a pipeline fashion, and a hybrid model where different primitives are requested in parallel by different travelers. This additional expressiveness raises new algorithmic questions and new ways of conducting user studies at scale.

REFERENCES

[1] Munmun De Choudhury, Moran Feldman, Sihem Amer-Yahia, Nadav Golbandi, Ronny Lempel, and Cong Yu. Automatic construction of travel itineraries using social breadcrumbs. In *HT*, pages 35–44, 2010.

[2] Igo Ramalho Brilhante, José Antônio Fernandes de Macêdo, Franco Maria Nardini, Raffaele Perego, and Chiara Renso. Where shall we go today?: planning touristic tours with tripbuilder. In *CIKM'13, San Francisco, CA, USA*, pages 757–762, 2013.

[3] Zhiwen Yu, Huang Xu, Zhe Yang, and Bin Guo. Personalized travel package with multi-point-of-interest recommendation based on crowdsourced user footprints. *IEEE Trans. Human-Machine Systems*, 46(1):151–158, 2016.

[4] Manish Singh, Ria Mae Borromeo, Anas Hosami, Sihem Amer-Yahia, and Shady Elbassuoni. Customizing travel packages with interactive composite items. In *2017 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2017, Tokyo, Japan, October 19-21, 2017*, pages 137–145, 2017.

[5] Sihem Amer-Yahia, Behrooz Omidvar Tehrani, Senjuti Basu Roy, and Nafiseh Shabib. Group recommendation with temporal affinities. In *EDBT*, pages 421–432, 2015.

[6] Sihem Amer-Yahia, Senjuti Basu Roy, Ashish Chawlat, Gautam Das, and Cong Yu. Group recommendation: Semantics and efficiency. *Proceedings of the VLDB Endowment*, 2(1):754–765, 2009.

[7] Sihem Amer-Yahia, Francesco Bonchi, Carlos Castillo, Esteban Feuerstein, Isabel Méndez-Díaz, and Paula Zabala. Composite retrieval of diverse and complementary bundles. *IEEE Trans. Knowl. Data Eng.*, 26(11):2662–2675, 2014.

[8] Albert Angel, Surajit Chaudhuri, Gautam Das, and Nick Koudas. Ranking objects based on relationships and fixed associations. In *EDBT 2009*, pages 910–921, 2009.

[9] Horatiu Bota, Ke Zhou, Joemon M. Jose, and Mounia Lalmas. Composite retrieval of heterogeneous web search. In *WWW 2014*, pages 119–130, 2014.

[10] Alexander Brodsky, Sylvia Morgan Henshaw, and Jon Whittle. Card: a decision-guidance framework and application for recommending composite alternatives. In *RecSys*, pages 171–178, 2008.

[11] Adrian Graham, Hector Garcia-Molina, Andreas Paepcke, and Terry Winograd. Time as essence for photo browsing through personal digital libraries. In *ACM/IEEE JCDL*, pages 326–335, 2002.

[12] Alexander Jaffe, Mor Naaman, Tamir Tassa, and Marc Davis. Generating summaries and visualization for large collections of geo-referenced photographs. In *SIGMM MIR*, pages 89–98, 2006.

[13] Vincent Leroy, Sihem Amer-Yahia, Éric Gaussier, and Seyed Hamid Mirisaee. Building representative composite items. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015*, pages 1421–1430, 2015.

[14] Senjuti Basu Roy, Sihem Amer-Yahia, Ashish Chawla, Gautam Das, and Cong Yu. Constructing and exploring composite items. In *SIGMOD Conference*, pages 843–854, 2010.

[15] Min Xie, Laks V.S. Lakshmanan, and Peter T. Wood. Breaking out of the box of recommendations: From items to packages. In *RecSys*, 2010.

[16] Sihem Amer-Yahia, Senjuti Basu Roy, Ashish Chawla, Gautam Das, and Cong Yu. Group recommendation: Semantics and efficiency. *PVLDB*, 2(1):754–765, 2009.

[17] Mark O'Connor, Dan Cosley, Joseph A. Konstan, and John Riedl. Polylens: a recommender system for groups of users. In *ECSCW*, 2001.

[18] Anthony Jameson and Barry Smyth. Recommendation to groups. In *The adaptive web*, pages 596–627. Springer, 2007.

[19] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Lda. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[20] James C. Bezdek, Robert Ehrlich, and William Full. FCM: The Fuzzy c-Means Clustering Algorithm. *Computers & Geosciences*, 10(2-3):191–203, 1984.

[21] SurveyMonkey. Calculating the number of respondents you need. https://help.surveymonkey.com/articles/en_US/kb/How-many-respondents-do-I-need.

[22] Djellel Difallah, Elena Filatova, and Panos Ipeirotis. Demographics and dynamics of mechanical turk workers. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 135–143. ACM, 2018.

[23] Sihem Amer-Yahia and Senjuti Basu Roy. Interactive exploration of composite items. In *Proceedings of the 21th International Conference on Extending Database Technology, EDBT 2018, Vienna, Austria, March 26-29, 2018.*, pages 513–516, 2018.

[24] Senjuti Basu Roy, Gautam Das, Sihem Amer-Yahia, and Cong Yu. Interactive itinerary planning. In *ICDE 2011*, pages 15–26, 2011.

[25] Igo Ramalho Brilhante, José Antônio Fernandes de Macêdo, Franco Maria Nardini, Raffaele Perego, and Chiara Renso. Planning sightseeing tours using crowdsensed trajectories. *SIGSPATIAL Special*, 7(1):59–66, 2015.

[26] Kwan Hui Lim, Jeffrey Chan, Christopher Leckie, and Shanika Karunasekera. Towards next generation touring: Personalized group tours. In *ICAPS*, pages 412–420, 2016.

[27] Dawei Chen, Cheng Soon Ong, and Lexing Xie. Learning points and routes to recommend trajectories. In *CIKM*, pages 2227–2232, 2016.

[28] Haoqi Zhang, Edith Law, Rob Miller, Krzysztof Gajos, David C. Parkes, and Eric Horvitz. Human computation tasks with global constraints. In *CHI*, pages 217–226, 2012.