

# Xenia: A Context Aware Tour Recommendation System Based on Social Network Metadata Information

Michalis Korakakis  
Department of Informatics  
Ionian University  
Corfu, Greece  
p12kora@ionio.gr

Phivos Mylonas  
Department of Informatics  
Ionian University  
Corfu, Greece  
fmylonas@ionio.gr

Evaggelos Spyrou  
Institute of Informatics and Telecommunications  
NCSR – “Demokritos”  
Athens, Greece  
espyrou@iit.demokritos.gr

**Abstract**—Tour planning and point-of-interest (POI) recommendation are two challenging and time-consuming tasks for tourists, predominately due to the large number of POIs a travel destination may contain and the complex constraints and parameters associated with the trip itself (e.g., time, budget, etc.). In this paper we present Xenia, a context-aware platform aiming to construct travel routes (i.e., ordered visits to various POIs that maximize the user’s travel experience) that adhere to the aforementioned limitations by modeling and solving the tour planning dilemma through the Orienteering Problem (OP). To achieve this, we use geo-tagged photos, collected from Flickr and exploit their metadata (e.g., time-stamps, geolocation and user-generated tags). By utilizing these spatio-temporal data, we are able to identify the trajectory patterns of tourists during their vacations and determine the most popular POIs in any given city, along with the tourists sequential POIs visits and their corresponding durations. Finally, we evaluate the effectiveness of the proposed system against a set of typical baseline approaches.

## I. INTRODUCTION

It is rather true that the modern digital era is characterized by the production and consumption of huge amounts of user-generated digital multimedia content. This content is typically shared via social networks that have almost completely replaced traditional means of exchanging multimedia, e.g., e-mail or personal websites. Dominant role in this new market is taken by both web and social networking services alike such as Facebook<sup>1</sup>, Flickr<sup>2</sup>, Twitter<sup>3</sup>, Picasaweb<sup>4</sup> etc. Users tend to upload their photos so as to share them with their digital “friends” or contacts. These photos are typically “tagged”, i.e., annotated in terms of the location/date they had been taken, people present, depicted events, etc.

The latter information, typically referred to as “metadata” plays a crucial role in today’s research activities, since huge, yet weakly annotated datasets are now offered. This means that even though users do not tag their digital content having in mind the needs of the research community, the information

they provide may often be characterized as “social knowledge,” as some kind of knowledge for a specific domain/application may often be extracted. Significant research efforts have been turned towards Flickr [16], since it provides a powerful API<sup>5</sup> and the majority of its hosted photos and their accompanying metadata may be used for non-profit activities<sup>6</sup>. Also its photos are usually *geotagged*, i.e., the location of the depicted content has been added either automatically (e.g., by the camera/smartphone used) or manually (e.g., by the photographer). Since the majority of cameras used are smartphones and consumer or entry level SLR cameras<sup>7</sup>, we may argue that the majority of Flickr users use it as a means of storing and sharing their personal photos. Using this information, one may extract knowledge about users’ whereabouts, interests or even recommend them additional, semantically related information towards covering their information needs.

In this work we propose a novel system, namely *Xenia*, which is a context-aware platform, aiming to automatically construct travel routes. A travel route is an ordered set of POIs, built upon various constraints and parameters that a user imposes explicitly or indirectly. Our platform generates these routes based on socially-generated knowledge derived from a dataset collected from Flickr and containing approx. 130K geotagged photos along with their metadata. We discard any visual information (i.e., the visual content of photos) and work solely on the metadata. Our goal is to identify trajectory patterns generated by users. Since Flickr typically contains touristic photos, we assume that these patterns are generated by tourists within their visit in a given city.

Upon a clustering procedure on the geospatial data of photos, we are able to discover “Areas-of-Interest” (AOIs), without any prior knowledge of the given urban area. An AOI constructed this way is assumed to be an area attracting a large number of visitors (thus “containing” a large number of photos). Each AOI should contain one or more “Places-of-Interest” (POIs). A POI may be a single attraction either limited to a relatively small geographic area (e.g., a statue, or a small building) or to relatively large one (e.g., a museum or a monument). We try to identify POIs contained within an AOI,

<sup>1</sup><http://www.facebook.com>

<sup>2</sup><http://www.flickr.com>

<sup>3</sup><http://www.twitter.com>

<sup>4</sup><https://picasaweb.google.com/>

<sup>5</sup><https://www.flickr.com/services/api/>

<sup>6</sup><https://www.flickr.com/creativecommons/>

<sup>7</sup><http://blog.flickr.net/en/2015/01/13/camera-ownership-on-flickr-2013-2014/>

by ranking textual metadata of photos, cross-checked to geo-information related databases, e.g., map services and Wikis.

We then model our problem as a Tourist Trip Design Problem (TTDP) [14] and solve it through a variant of the Orienteering Problem (OP) [18]. This way, we are able to create touristic tours that may satisfy a multitude of different constraints and parameters, imposed either explicitly by the user or indirectly through various trip-related limitations.

The remaining of the paper is organized as follows: In section II we present related work in the fields of recommendations using socially generated metadata, collected from Flickr and recommendation problems that have been tackled as TTDP. Then, in section III we state our problem as a variation of the TTDP and solve it through the use of the OP. The Xenia system is presented in section IV, where we discuss in detail the set of algorithms involved. Experimental results are presented and discussed in section V. Finally, conclusions are drawn in section VI, where we also discuss possible further extensions of Xenia and plans for related future work.

## II. RELATED WORK

It is common sense, and could be easily verified upon browsing the well-known social networks, that a significantly large percentage of multimedia content circulated within them is of touristic travel nature, e.g., photos or videos of landmarks, POIs, events, etc. These could be used to extract some kind of knowledge, typically oriented to reveal tourists' preferences on landmarks and tourist routes. To this goal, many research efforts have focused on the exploitation of such content towards satisfying travelers' and/or tourists' potential needs. A large research area is oriented not only to the recommendation of main attractions, but also to organizing users' trips, aiming to propose efficient organization of their schedules.

In the early work of Popescu and Grefenstette [12] temporal Flickr metadata are exploited so as to estimate expected visiting times for tourist attractions, thus deducing what a tourist is able to visit in a city within a day. Jain et al. [7] created a graph based on photos. For a given location as a starting point they proposed a tour that visits POIs under distance constraints, however without considering time spent at each POI, which has been exploited by Popescu et al. [13], who also managed to generate new trips by combining existing ones. Similarly, Hao et al. [6] presented an approach that created virtual tours recommending popular places, by mining Flickr data. Sun et al. [17] used a spatial clustering approach which allowed them to identify landmarks. They ranked them based on their popularity and provided recommendations based on minimizing distances and maximizing popularity. Majid et al. [11] collected users' travel history using it to recommend locations in unknown cities. Brillhante et al. [1] presented a framework that models the problem of recommendations as a Generalized Maximum Coverage problem, mining information from Flickr and Wikipedia. Yahi et al. [19] developed Aurigo, a system that relies on an user's own preferences along with crowd-sourced reviews and ratings, so as to construct personalized itineraries that also consider suggestions from the user regarding POIs that should be included within the travel path by utilizing an interactive user interface.

Amongst the few research efforts that tackled the problem of recommendations as a TTDP, we should emphasize the original work of De Choudhury et al. [3] who presented an approach that aims to form travel routes by combining heuristics with a solution of the OP. Within more recent efforts, our work is more similar to the one of Lim et al. [9] who presented an approach for personalized tours, exploiting POI popularity and user interests, formulating their problem as an OP, considering time limits and initial/ending points. They [10] also extended the problem of recommendation for groups of tourists. Their approach was to break down this complex problem to a set of more manageable sub-problems, among whom the allocation of a tourist to a group and of a POI to a group. For the former sub-problem they used k-means and hierarchical clustering, while for the latter they used a variation of the OP that considers various constraints. The survey from Souffriau and Vansteenwegen [15] focuses on how models from the field of Operations Research (OR) fit the problem of scheduled recommendations and how the OP and its extensions may be used to model TTDP in practical applications. Moreover, the survey of Gavalas et al. [5] covers algorithmic approaches for solving TTDP problems.

We should herein emphasize that our work is novel, when compared to the aforementioned, since it does not require any type of prior knowledge (e.g., a set of landmarks as in [1] or knowing the category of POIs as in [9]), at any step. All knowledge is automatically extracted using only the available user-generated "noisy" metadata.

## III. FORMULATING THE TOURIST TRIP DESIGN PROBLEM

The most crucial and time-consuming aspect during the arrangement of a tourist's vacation is the design of an optimal trip timetable, mainly due to the fact that it is impossible for her/him to visit every single POI of the visited city, throughout the duration of her/his vacations. Thus, tourists should be able to choose the most beneficial POIs for them, according to their own set of criteria and visit them in an organized and also timely way. As depicted in [14] the task of selecting multiple POIs that adjust to various context-based constraints and parameters (e.g., the weather conditions, hotel selection, transit times, etc.), is typically referred to as the "Tourist Trip Design Problem" (TTDP). Solutions for the TTDP require a sequential set of ordered POI visits that maximize the user's satisfaction, while additionally fully exploiting the available time limit in order to construct the appropriate schedule for the trip. The Orienteering Problem (OP) is considered to be the simplest form of modeling the TTDP.

The Orienteering Problem (OP) is based on the orienteering game, in which several locations with an associated score have to be visited within a given time limit. Within this work we model the TTDP based on the classical formulation of the computationally NP-Hard OP, solving it as an integer problem. The goal is to find a tour that, given a starting and a terminal POIs, maximizes the total score, which is earned upon each visit at a POI, while also adhering to a positive time budget.

The classical OP [18] is typically defined as: "Given a set of vertices, each assigned with a score, determine a path bounded in terms of length, maximizing the sum of scores of visited vertices." We should note that this score is

typically heuristically determined, depending to the context of the application. To formulate it, we first assume the existence of a set of vertices  $\mathcal{V} = \{v_i, i = 1, 2, \dots, N\}$ . At each vertex  $v_i$ , a non-negative score  $s_i$  has been assigned. We define  $v_1$  and  $v_N$  to be the starting and ending vertices, respectively. We also assume that it is not possible to visit all vertices, by imposing a time constraint  $T_{\max}$ , and setting a time cost  $t_{ij}$  between vertices  $v_i$  and  $v_j$ . The goal is to visit a subset of  $\mathcal{V}$ , so as a) to maximize the sum of scores of visited vertices; and b) visit each vertex at most once.

Using this notation, we may formulate the OP as an integer problem. Variable  $x_{ij}$  is equal to 1 when a visit to  $v_i$  has been followed by a visit to  $v_j$ , else is equal to 0. Also,  $u_i$  provides the position of  $v_i$  within the given path. The objective function (score) that has to be maximized is given by Eq. 1.

$$\max \sum_{i=2}^{N-1} \sum_{j=2}^N s_i x_{ij} \quad (1)$$

given the following set of six constraints:

$$\sum_{j=2}^N x_{1j} = \sum_{i=1}^{N-1} x_{iN} = 1 \quad (2)$$

$$\sum_{i=1}^{N-1} x_{ik} = \sum_{j=2}^N x_{kj} \leq 1; \forall k = 2, \dots, N-1 \quad (3)$$

$$\sum_{i=1}^{N-1} \sum_{j=2}^N t_{ij} \leq T_{\max} \quad (4)$$

$$2 \leq u_i \leq N; \forall i = 2, \dots, N \quad (5)$$

$$u_i - u_j + 1 \leq (N-1)(1 - x_{ij}); \forall i, j = 2, \dots, N \quad (6)$$

$$x_{ij} \in \{0, 1\}; \forall i, j = 1, \dots, N \quad (7)$$

The constraints described by Eq. 2 guarantee that the extracted path's starting and ending points are indeed  $v_1$  and  $v_N$ , respectively. Those of Eq. 3 guarantee the connectivity of the path, while also ensure that each vertex is visited at most once. The time constraint  $T_{\max}$  is imposed by Eq. 4. Eqs. 5, 6 are used to eliminate subtours (i.e., directed cycles within the constructed tour). Eq. 7 has already been discussed. Additionally, we assume that  $t_{ij} = t_{ji}$ , i.e., equal travel time among the corresponding vertices.

#### IV. THE XENIA TRIP RECOMMENDER

In this section we shall present our trip recommender system, namely "Xenia." It exploits social knowledge, i.e., user-generated metadata from Flickr to discover Areas-of-Interest, Places-of-Interest and recommend travel routes, aiming to maximize users' overall experience.

##### A. Clustering

The goal of the first step of the proposed system is to discover AOIs of an urban area "in the wild", i.e., without any prior knowledge. To this goal we use geospatial information extracted from photos collected from Flickr. Within the context of this work, we shall assume that an AOI is characterized both by a large number of attracted visitors and by a relatively large number of POIs, contained within. Since the geospatial

information of a photo is expressed as a 2D point (i.e., the corresponding latitude and longitude of the location that it has been taken<sup>8</sup>), the task of AOI identification may be tackled using density-based clustering techniques. These rely on the density attribute (herein the mutual reachability distance) of points contained within the dataset, in order to group them into distinct clusters, which are dense regions, separated by low-density areas, having an arbitrary size and shape. This particular trait is rather useful for our purposes due to the fact that it is typically encountered in the case of urban AOIs.

We choose to cluster these data using HDBSCAN [2], which is an extension of the well-known DBSCAN algorithm [4] that transforms the latter into a hierarchical clustering algorithm. Using the generated dendrogram (i.e., a structural graphical representation of the distances between the connected components of the clustered data) it is then possible to extract flat clusters with varying density thresholds through a top-down process. We should emphasize that HDBSCAN requires a single user-specified parameter as input: the minimum size of points (`minPts`) that a given cluster may contain.

We follow a trial-and-error approach and experiment with various values, using relatively high and low values for `minPts`. Our goal is to identify clusters that correspond to a potential "popular" area (i.e., an area visited by a large number of tourists) while at the same time they do not cover a large part of the examined area, since this way they may contain a very large number of POIs, which is an undesired property.

##### B. Tag Ranking and POI Extraction

For the ranking of tags within AOIs and for the extraction of POIs, we adopt the following approach: First we identify the top  $k$ -representative tags for each cluster. We use the  $N \times N$  co-occurrence matrix  $C$  for each cluster as a means of tag ranking metric.  $C(i, j)$  indicates the number of photos where the  $i$ -th word co-occurs with the  $j$ -th word, i.e., they have been both used to tag it. We should note that  $N$  is the number of the unique words (tags) that are contained within the photos of the examined cluster. Through the use of this matrix, we capable to pair together tags that are semantically related (e.g., *acropolis* and *parthenon*). By employing this technique we are able to accumulate a varied set of representative tags for a geo-cluster while ignoring those with high frequencies.

Finally, we collect relevant information, e.g., names and geographic coordinates, about the POIs of the examined urban area from OpenStreetMap<sup>9</sup> and Wikipedia's Geonames<sup>10</sup>. In order to discern whether a specific POI belongs to a geo-cluster we choose to use their Levenshtein distance [8], between the set of words that characterize it and the top- $k$  representative tags for the geo-cluster. In principle, the Levenshtein distance between two words is defined as the minimum number of edits needed to transform one word into the other, with the allowable edit operations being insertion, deletion or substitution of a single character. The result of this process is the identification

<sup>8</sup>To be more accurate, this geospatial information, when it is manually generated by the users, is prone to errors, since geotagging may in some cases be a subjective task. Thus in some cases it represents the location where a photo has been tagged.

<sup>9</sup><https://www.openstreetmap.org>

<sup>10</sup><http://www.geonames.org>

of a set of POIs within each cluster. The cardinality of this set shall be used as a means to define the cluster’s importance, i.e., a cluster containing a large number of POIs is assumed to be of greater touristic importance.

### C. Travel Sequences

We manually divide a user’s travel history into “sub-streams” through the construction of travel sequences, i.e., consequential visits to POIs. This is the first step for the computation of the average visit duration for a specific POI. We define a travel route  $r$  as a sequence  $\{p_1, p_2, \dots, p_N\}$  of photos. Route members should comply to the following rules:

- $p_1.date \leq p_2.date \leq \dots \leq p_N.date$ , i.e., all members should be consecutive;
- $p_1.user = p_2.user = \dots = p_N.user$ , i.e., all members should belong to the same user;
- let  $p_i, p_j$  denote a pair of consecutive route members. Then  $p_i.date < p_j.date + T_h$ , i.e., a restriction is imposed on the maximum time difference between photos in order to be considered as consecutive within a route.

Using the whole dataset, and for each unique user we are able to construct a set of her/his travel sequences, according to the aforementioned rules. The visit duration per user for a given POI is calculated as the temporal difference of the first and the last photos she/he had taken within it.

### D. OP Implementation Details

The score function used for the selection of a POI is heuristically determined using the corresponding distinct number of visitors for a POI, i.e., its *popularity*. We consider that a high-traffic POI is of value for a tourist, based on the “wisdom-of-the-crowd”, i.e., the general opinion that a group of people possesses. Moreover, for each POI we consider the *average visit duration*, calculated based on the visit durations of all users that have visited this specific POI. Finally, in order to determine real-life pairwise travel times between POIs we use the Google Maps Distance Matrix API.<sup>11</sup>

## V. EXPERIMENTS

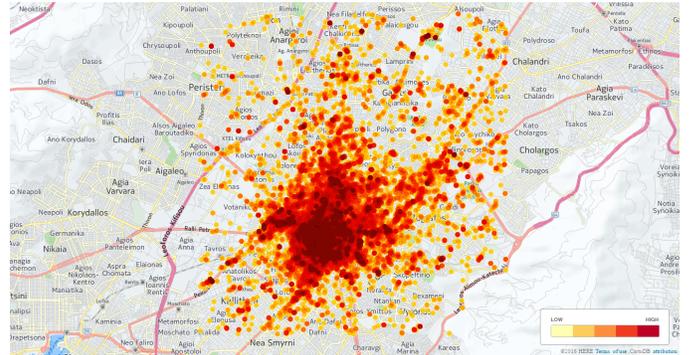
In this section we will present details on the application of Xenia to an urban dataset collected from Attica, Greece.

### A. Dataset

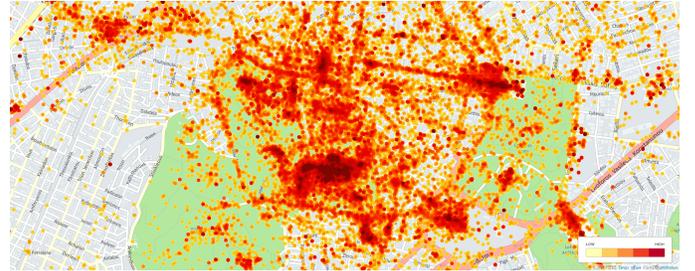
For the experimental evaluation of our approach we used an urban image data set which consists of a total of 130,200 photos collected from the region of Attica, Greece. Considering that a vast amount of popular POIs from this region originate in the city of Athens and more specifically in its center and the surrounding areas, we executed a geo-query so as to collect photos solely from this municipality and its nearby suburbs and exclude popular yet mundane districts for tourists (e.g., major airports, ports, etc.) that are of no value to them. As a result of this process, we ended-up with 102,278 photos. All these photos are geo-tagged, dated between January 2004–December 2015 and collected from Flickr using its public API.<sup>12</sup> Then through the use of a manually created stoplist

and regular expressions, we removed tags that either did not have a semantic relation to the respective photo that they were attached to (i.e., automatically added tags by smartphones and cameras e.g., *iphone*, *android*, etc.) or too generic ones (i.e., tags that are both common and spread to the whole area, thus not providing any useful information, while also tending to be amongst the most popular e.g., *holidays*, *Greece*, *Athens*, etc.). Upon completion of this process and given the initial 43,819 tags, we ended up with overall 35,033 unique ones.

In Fig. 1a we present a density-based visualization of the distribution of the collected photos in the whole urban area used throughout our experiments. Moreover, in Fig. 1b we depict the city center, which is the area that mainly attracts tourists, thus contains the majority of photos. Intuitively, one should notice that high-density areas correspond to places of increased touristic importance.



(a) The whole urban area of Attica, used for our experiments.



(b) The Athens’ city center.

Fig. 1. Density-based visualizations of Flickr photos that lie within the selected geographic bounding box for the broader area of Athens.

### B. Clustering and Travel Sequences Construction

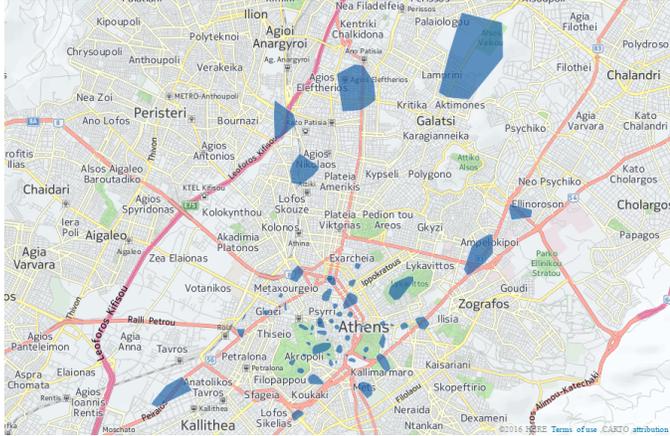
The application of HDBSCAN on the set of geo-tags with  $minPts = 200$ , as described in section IV-A, resulted to a set of 65 geo-clusters. The process of POI recognition and extraction, presented in section IV-B, lead to the identification of 40 places. One may notice that the number of places is smaller than the one of the AOIs (clusters). This happens since we filtered all places that do not correspond to landmarks (e.g., restaurants, bars, cinemas, etc.), yet are frequently visited by tourists. For the construction of the travel sequences we ensued the following procedure. We first queried the dataset for each user separately. Then we ordered all photos per user at ascending date. As we mentioned in section IV-C, we set  $T_h = 8h$ , under the assumption that consecutive photos taken at very large intervals, belong to different sequences. We ended up with this conclusion upon careful inspection of the available

<sup>11</sup><https://developers.google.com/maps/documentation/distance-matrix/>

<sup>12</sup><https://www.flickr.com/services/api/>

data set. We then discarded sequences that consisted of less than 2 or more than 25 members.

In Fig. 2 we illustrate the extracted clusters at two different zoom levels. More specifically, in Fig. 2a we depict the extracted clusters in the whole area, while in Fig. 2b those of the Athens’ city center. The comparison between these clusters and the densities of Fig. 1, is illustrated in Fig. 3 and provides intuitive verification of the clustering process.



(a) A map of Attica portraying all the geo-clusters constructed by HDBSCAN.



(b) High-density geo-clusters around Athens’ city center.

Fig. 2. The set of geo-clusters that were extracted from the Flickr data set using HDBSCAN. The minimum size of points per cluster was set to 200.

### C. Baselines

The main objective of Xenia is to automatically create and recommend travel routes that can actively contribute to the increase of a tourist’s overall travel experience for a holiday destination that is characterized by a large number of POIs. Thus, we assess the performance of our proposed platform against the following baselines, which are in general aligned to the ones used in the state-of-the-art:

- *Random POI Selection (RPOI)*. RPOI randomly selects an unvisited POI at each step and adds it in the route.
- *Greedy POI Selection (GPOI)*. GPOI chooses the next POI according to its corresponding earned score, i.e., the highest score from the set of unvisited POIs.
- *Nearest POI Selection (NPOI)*. NPOI constructs the travel route by adding an unvisited POI to the latter, according to its distance from the POI that was selected in the immediate preceding iteration of the process.

### D. Evaluation

We evaluated Xenia compared to the aforementioned baselines and using the following performance metrics:

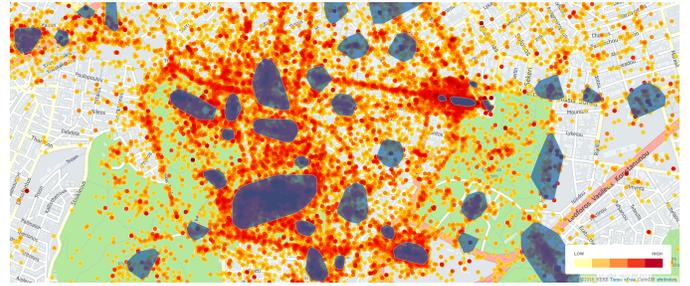


Fig. 3. Extracted clusters (Fig. 2b), overlaid to the density-based visualization of photos’ distribution (Fig. 1b).

TABLE I. PERFORMANCE OF XENIA AGAINST THE APPLIED BASELINE TECHNIQUES ACCORDING TO VARIOUS METRICS.

	Recall	Precision	Popularity
Xenia	<b>.653</b>	<b>.548</b>	<b>.704</b>
RPOI	.315	.462	.318
GPOI	.498	.538	.549
NPOI	.465	.412	.506

- *Recall*. The Recall metric measures the fraction of retrieved documents from a collection that are relevant, in the context of a given query. Herein, we calculate Recall as the amount of visited POIs from a user’s travel history that also appear in the recommended tour.
- *Precision*. We define Precision as the fraction of POIs that appear in a recommended tour from the platform that are also part of a user’s real-life travel sequence.
- *Popularity*. It is defined in section IV-D as the number of distinct users that have visited a specific POI.

We apply the aforementioned evaluation metrics using the following approach. For any given itinerary constructed through the platform we compare and contrast the set of recommended POIs included in it with information found within the examined dataset. Particularly, for Recall and Precision this is expressed through the overall amount of suggested POIs that align with a test subject’s own travel history, i.e., a randomly selected user from the Flickr dataset that satisfies the constraint of visiting a multitude of POIs in Athens, whereas for Popularity we aggregate and then calculate the number of distinct users that have visited the set of POIs that were included in the recommended tour.

### E. Results

In Table I we present the results of the evaluation of Xenia in the Athens’ dataset, from which, we are able to assess its performance against all applied baselines. Our proposed methodology managed to outperform all the baselines that were evaluated, essentially fulfilling its goal, i.e., to maximize the overall travel experience for a tourist, given a area that is characterized by a high-density of POIs. In particular, for the Recall metric it achieved the best results in comparison with the other baselines. GPOI scored the second best performance whereas NPOI and RPOI came in third and fourth respectively. For the Precision metric, our approach was again the best, however by a marginal difference compared to the GPOI which was second. Moreover, RPOI managed to outperform the NPOI baseline by a relatively small margin. Finally, in terms of Popularity, our approach significantly surpassed all baselines.

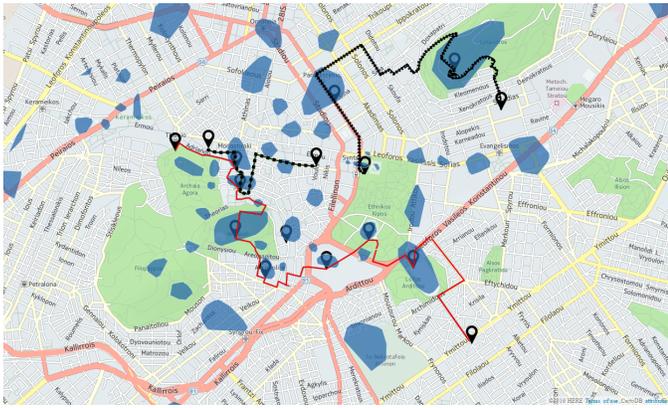


Fig. 4. Travel routes constructed by the platform for different time budgets and starting/ending points. The markers showcase a POI that was identified from a particular cluster and was added to a route.

The results from both Precision and Recall measures indicate the relevance of our platform to real-life travel routes. In the case of the Popularity, the outcome of the experiments showcases that Xenia is able to recommend tours that contain highly-demanded POIs characterized by the attribute of popularity that we measure through the overall number of people that visit an area. In Fig. 4 we depict the recommended routes for different time periods, while details are presented in Table II. Starting/ending points were randomly selected from a set of predefined available POIs in the center of Athens.

TABLE II. DETAILS FOR THE TRAVEL ROUTES DEPICTED IN FIG. 4

<b>1st query (green – dense dots)</b>		
Start: Monastiraki (37.976292, 23.723786)	End: Mitropoleos (37.975254, 23.732320)	Duration: 50 min.
Hadrian's Library → Roman Agora		
<b>2nd query (black – stipples)</b>		
Start: Kolonaki (37.979055, 23.746925)	End: Syntagma Sqr. (37.976670, 23.720750)	Duration: 100 min.
Lycabettus → Academy of Athens → Tomb of the Unknown Soldier		
<b>3rd query (red – straight line)</b>		
Start: Empedokleous (37.964379, 23.744454)	End: Thissio (37.976670, 23.720750)	Duration: 240 min.
Panathenaic Stadium → Zappio Megaro → Temple of Olympian Zeus → Acropolis Museum → Acropolis of Athens → Roman Agora → Hadrian's Library		

## VI. CONCLUSIONS AND DISCUSSION

In this paper, we presented Xenia, a novel platform for recommending touristic routes that satisfy a multitude of different constraints and parameters. Through an unsupervised approach we clustered a set of images collected from Flickr as a means to discover high traffic areas that we defined as AOIs. An AOI can contain multiple POIs that are favored by tourists during their vacations. We modeled and solved the TTDP problem through the integer programming formulation of the OP. We evaluated the proposed platform against a set of typical baselines. Our results indicate that it is able to provide routes that are comparable to real-life ones. We also demonstrated the robustness and the efficiency of our system, considering all the necessary stages, i.e., from data clustering to travel route construction. Since no prior knowledge was necessary at any stage, we proved that our approach is able to successfully work “in the wild”.

Among our plans for future work, we intend to assess our proposed platform in additional cities and integrate temporal and/or weather data as a means to deduce the number of people that would probably visit a particular POI at a given time. Within the process we would also like to incorporate information from other social networks, e.g., Twitter, blogs etc. and other modalities, e.g., sentiment analysis using Natural Language Processing. We also plan to make suggestions for groups of people, instead of individuals and gender/age based recommendations. Finally, we would like to expand the domain of recommendations by adding other types of POIs, e.g., restaurants, bars, and even hotels.

## REFERENCES

- [1] I. Brillhante, J.A. Macedo, F.M. Nardini, R. Perego and C. Renso, *Where shall we go today?: planning touristic tours with Tripbuilder*. In Proc. of ACM Int'l Conf. on Information & Knowledge Management, 2013.
- [2] R.J. Campello, D. Moulavi and J. Sander, *Density-based clustering based on hierarchical density estimates*. In Proc. of Pacific-Asia Conf. on Knowledge Discovery and Data Mining, 2013
- [3] M. De Choudhury, M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel and C. Yu, C., *Automatic construction of travel itineraries using social breadcrumbs*. In Proc. of ACM conf. on Hypertext & Hypermedia, 2010.
- [4] M. Ester, H.P. Kriegel, J. Sander and X. Xu, *A density-based algorithm for discovering clusters in large spatial databases with noise*. In KDD, vol.96, no.34, pp. 226–231, 1996.
- [5] D. Gavalas, C. Konstantopoulos, K. Mastakas and G. Pantziou, *A survey on algorithmic approaches for solving tourist trip design problems*. Journal of Heuristics, vol.20, no.3, pp.291–328, 2014.
- [6] Q. Hao, R. Cai, J.-M. Yang, R. Xiao, L. Liu, S. Wang and L. Zhang, *Travelscope: standing on the shoulders of dedicated travelers*. In Proc. of ACM MM, 2009.
- [7] S. Jain, S. Seufert and S. Bedathur, *Antourage: Mining Distance-Constrained Trips from Flickr*. In Proc. of ACM WWW, 2010.
- [8] V.I. Levenshtein, *Binary codes capable of correcting deletions, insertions, and reversals*, Soviet Physics Doklady 10, pp. 70710, 1966.
- [9] K.H. Lim, J. Chan, C. Leckie, S. Karunasekera, *Personalized Tour Recommendation Based on User Interests and Points of Interest Visit Durations*. In Proc. of IJCAI, 2015.
- [10] K.H. Lim, J. Chan, C. Leckie, S. Karunasekera, *Towards Next Generation Touring: Personalized Group Tours*. In Proc. of Int'l Conf. on Automated Planning and Scheduling (ICAPS), 2016.
- [11] A. Majid, L. Chen, Ling, G. Chen, H.T. Mirza, I. Hussain and J. Woodward, *A context-aware personalized travel recommendation system based on geotagged social media data mining*. Int. Journal of Geographical Information Science, vol.27, no.4, pp. 662–684, Taylor & Francis, 2013.
- [12] A. Popescu and G. Grefenstette, *Deducing trip related information from flickr*. In Proc. of ACM WWW, 2009.
- [13] A. Popescu, G. Grefenstette and P.-A. Moëllic, *Mining tourist information from user-supplied collections*. In Proc. of ACM CIKM, 2009.
- [14] W. Souffriau, P. Vansteenwegen, J. Vertommen, G. Vanden Bergh and D. Van Oudheusden, *A personalized tourist trip design algorithm for mobile tourist guides*. Applied Artificial Intelligence, vol.22, no.10, pp.964–985, 2008.
- [15] W. Souffriau, P. Vansteenwegen, *Tourist trip planning functionalities: Stateoftheart and future*. In Proc. of Int'l Conf. on Web Eng., 2010.
- [16] E. Spyrou and Ph. Mylonas, *A survey on Flickr multimedia research challenges*. Engineering Applications of Artificial Intelligence, vol.51, pp.71–91, 2016.
- [17] Y. Sun, H. Fan, M. Bakillah and A. Zipf, *Road-based travel recommendation using geo-tagged images*. Computers, Environment and Urban Systems, Elsevier, 2013.
- [18] P. Vansteenwegen, W. Souffriau, D. Van Oudheusden, *The orienteering problem: A survey*. European J. of Oper. Research, vol.209, no.1, 2011.
- [19] A. Yahi, A. Chassang, L. Raynaud, H. Duthil and D. H. P. Chau, *Aurigo: An interactive tour planner for personalized itineraries*. In Proc. of Int'l Conf. on Intelligent User Interfaces (IUI), 2015.