Tourismo:A User-Preference Tourist Trip Search Engine

Klaus Arthur Schmid[1], Gregor Jossé[1], Andreas Züfle[1], Georgios Skoumas[2], Matthias Schubert[1], Dieter Pfoser[3]

[1] Ludwig-Maximilians-Universität München, Germany
{schmid,josse,schubert,zuefle}@dbs.ifi.lmu.de,
[2] National Technical University of Athens, Greece
gskoumas@dblab.ece.ntua.gr,
[3] George Mason University, USA
dpfoser@gmu.edu

Abstract. In this demonstration we re-visit the problem of finding an optimal route from location A to B. Currently, navigation systems compute shortest, fastest, most economic routes or any combination thereof. More often than not users want to consider "soft" qualitative metrics such as popularity, scenic value, and general appeal of a route. Routing algorithms have not (yet) been able to appreciate, measure, and evaluate such qualitative measures. Given the emergence of user-generated content, data exists that records user preference. This work exploits user-generated data, including image data, text data and trajectory data, to estimate the attractiveness of parts of the spatial network in relation to a particular user. We enrich the spatial network dataset by quantitative scores reflecting qualitative attractiveness. These scores are derived from a user-specific self-assessment ("On vacation I am interested in: family entertainment, cultural activities, exotic food") and the selection of a respective subset of existing POIs. Using the enriched network, our demonstrator allows to perform a bicriterion optimal path search, which optimizes both travel time as well as the attractiveness of the route. Users will be able to choose from a whole skyline of alternative routes based on their preference. A chosen route will also be illustrated using user-generated data, such as images, textual narrative, and trajectories, i.e., data that showcase attractiveness and hopefully lead to a perfect trip.

1 Introduction

Nowadays, social networks are a great source of rich geo-spatial data. Almost every social network allows users to incorporate geo-social features into their data stream. The different features include, amongst others, geo-tagged pictures (e.g. Flickr), geo-descriptive text (e.g. travel blogs), and tracked movement (e.g., runners' trajectories). For this demo, we rely on all these kinds of user-generated data to define attractiveness on a real world road network. Our aim is to reflect human fondness according to the crowd by using qualitative information and making it measurable. We present Tourismo, a tourist search engine, which computes attractive paths along points of interest (POIs), tailored to the interest of the user issuing the query. Based on this enriched spatial network, which has information about the attractiveness of locations, we aim at answering *attractive path queries*. Currently, navigation systems, i.e., machines, perform this task for us, computing routes such as the shortest route, the fastest route, the most economic route [1], or some combination of such quantitative measure on a spatial network [2]. In all of these cases, the employed algorithms optimize cost measures inherent in the underlying road network. What is rarely reflected, however, is user preference on subjective measures, such as attractiveness and interestingness of a route. Often users are willing to take a suboptimal detour, a deviation from quantitative optimality (shortest, fastest, etc.), in order to improve the quality of their route. In order to see more attractions, for instance, a tourist may be willing to take a moderate detour from a fast, but not very attractive, highway.

How can we measure a subjective concept of "quality"? How to measure attractive, scenic, recreative routes? As machines are not (yet) capable to reflect this concept, we rely on the crowd to answer this question, i.e., we propose to use crowdsourced data to estimate the attractiveness of an area. Relying on different datasets, image data (from Flickr[1]), textual narratives (from travel blogs), and trajectory data (from Endomondo [2]), we investigate the applicability of different data sources as cost measures for the underlying road network. More precisely, we enrich the road network by quantitative scores of qualitative statements as follows:

– areas having a large density of Flickr images indicate a particularly attractive area, increasing the attractiveness score;
– locations mentioned in the positive context of travel blogs increase attractiveness scores;
– routes commonly used by other users are also considered more attractive.

Furthermore, we incorporate meta-information from OpenStreetMap[3], in order to categorize POIs and, using the aforementioned popoularity score, propose routes according to the user's preferences and the fondness of the crowd. Tourismo presents solutions to enrich the underlying road network using the aforementioned data sources. We show an initial approach to map these *attractiveness scores* to a cost measure, which allows one to apply existing routing algorithms which aim at minimizing edge-labeled cost metrics. We apply an adapted algorithm for bicriterion pareto-optimal route search, to find paths which are optimal in both travel time and attractiveness. Our framework allows to specify origin and destination, computes and displays the skyline of pareto-optimal paths. Furthermore, the reasons for attractiveness of each path are illustrated: Flickr images along the way, travel blog entries mentioning locations on the way, and historical trajectories which share the same route. Our demonstrator, which we would like to present to the community at SSTD'15, is an extension of a demonstrator that we recently presented at ICDE'15 [3].[4] The new demonstrator has two major features: First, our demonstration allows to specify the interest of a user, thus allowing to return routes that contain POIs which are of particular interest to the user issuing the query. Second, this version allows to consider a third type of data to enrich the underlying road network with attractiveness information: In addition to geotagged images, and texts containing geospatial references, we allow to learn attractiveness from an existing base of historic trajectory data.

[1] www.flickr.com

[2] www.endomondo.com

[3] www.openstreetmap.org

[4] Since the ICDE proceedings are not publicly accessible at this time, we have attached the demonstration proposal at the end of this submission. Clearly, this footnote and the attached paper will be removed for a potential camera ready.

2 State of the Art

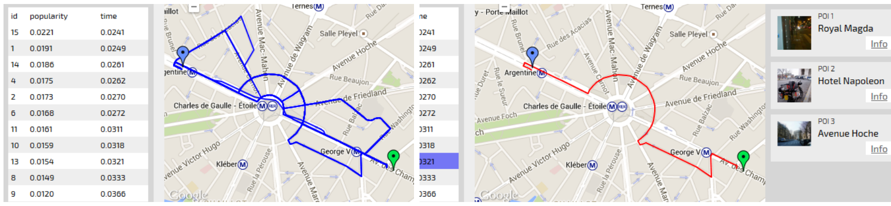
Recently, a lot of interesting research has been done in the context of finding scenic, interesting or popular routes. The first set of related work focuses on providing paths which are easier to memorize, describe, and follow. For example, the authors of [4], [5], and [6] try to tackle the problem by introducing cost criteria that allow for a trade-off between minimizing the length of a path while also minimizing the complexity in terms of instructions or turns along the path. Furthermore, an existing research direction covers the problem of defining tourist routes, which maximize the subset of a set of pre-defined POIs which can be visited in a tourist tour that has a time-constraint [7,8]. In these works, the set of interesting POIs is given, and the main conceptual contribution of is to automatically extract interesting locations, as well as a quantitative estimate of the popularity of this location from a variety of data sources.

The approach most similar to the one presented in this work is [9], which proposes a method for computing beautiful paths, as the authors phrase it. However, in order to quantify quality, the authors rely on explicit statements about the beauty of specific locations, obtained from a crowd-sourcing platform which collects user opinions on photos of specific locations. In contrast, we propose to mine this kind of information from existing crowd-sourced data, which does not require any monetary investment to aquire. Thus our approach has the crucial advantage that it is scalable as the used data is already available globally available, while having local expert users rate photos one by one can hardly be extended to a global scale.

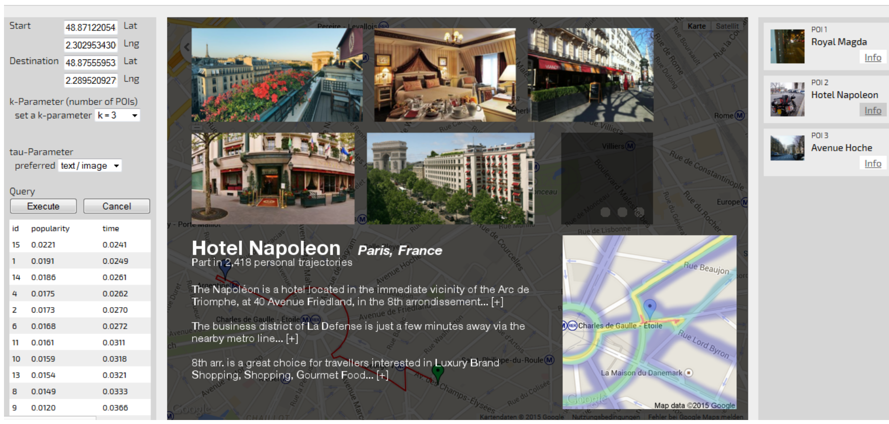
Another important research direction is the *stitching* existing trajectories in order to obtain new trajectories which guarantee that each sub-trajectory is used by other users, and is thus, "popular" following the definition [10] of Chen et al. This, however, only reflects a notion common usage, not taking into account, why a specific sub-trajectory has been favored. For instance, when mining trajectories of commuters, the fastest path is most likely to be chosen by most users. Hence, we propose mining trajectories specific to recreational use and merging this information with the attractiveness scores we derive from other user-generated data sources.

3 Features

The main feature of this demonstrator is the estimation of attractiveness from text, image, and trajectory data. Details covering text and image data can be found in [3]. In this section, we briefly describe how we enrich the underlying road network using historical trajectory data. For our demonstrator, we use trajectories of walkers, runners and bikers that have uploaded their workouts to Endomondo. Our dataset contains eight million trajectories, which are located all around the world, but have a strong regional focus in Northern Europe. To match each of the GPS trajectories, we apply state-of-the-art map matching techniques, similar to those presented in [11]. In a first step, we perform a basic enrichment: For each edge e of the spatial network, we count the number $\text{tra}(e)$ of historical trajectories that contain this edge. This count can be used as an indication of attractiveness of an edge, following the assumption that runners are, in average, more likely to choose a particularly nice running trail. As mentioned before, following the techniques proposed in [3], we obtain a road network having a attractiveness



(a) Bicriterion Path Skyline (b) Detailed Path Information



(c) Detailed information about selected PoI

Fig. 1. Functionality of the presented framework.

score derived from Flickr image data and travel blog text data. On top of that, we add the trajectory attractiveness tra(e), by introducing a new user-specific weighting factor which is omitted here for reasons of brevity. Relying on the enriched road network, Tourismo supports bicriterion pareto-optimal path queries. Additionally, Tourismo features category-specific path queries. If the user chooses to specify his personal touristic interests, they can choose one or more options from a list containing outdoor activities, cultural sightseeing, culinary interest, and more. In order to provide paths which fulfill these requirements, we mine the meta-information provided by OSM. Thanks to a very active community, OSM data contains well-tended information about POIs, that is named, categorized, and subcategorized. For instance, the meta-categories "food" and "tourist" contain subcategories "bar", "restaurant", "fastfood" and "monument", "museum", "archeological", respectively. Mapping these categories onto the options of user-preferences, we are able to filter POIs which correlate to the particular interest of the user. When querying a route with a specific set of interests, the user is provided a number of pareto-optimal paths, guiding him along POIs tailored to his preference.

4 Framework Description

The demonstrated framework allows users to validate that the notion of attractiveness defined in this paper indeed coincides with the general intuition. The result paths returned to the user yield competitive solutions in terms of travel time while passing

POIs perceived as significant, appealing, and/or recognizable. Hence, we solve the proclaimed task of providing “more attractive” paths to the user. Using OpenStreetMap as a road network, our demonstrator visualizes a map relying on Google Maps. Upon selecting an origin and a destination location on the map, the user is presented with the skyline view as shown in Figure 1(a). In this view, the route skyline is presented to the user, i.e., the set of routes which are pareto-optimal in terms of both popularity and travel time. For each such route, the corresponding travel times are shown in a table in the lower left corner of Figure 1(a). These routes are sorted by both popularity and travel time, which is equivalent by definition of pareto-optimality, i.e., there exists no two routes A, B in the route skyline such that A is both faster and more popular than B. Using this table, the user can select a route which corresponds to the user’s preference between travel time and popularity, yielding the route view shown in Figure 1(b). For the selected route A, this view shows the most popular points of interest on A.

Once a point of interest is selected, the sources of popularity of this POI are shown as in Figure 1(c). For this purpose, Figure 1(c) shows all the pictures relevant for the selected POI, i.e., the set of images having a sufficiently low distance. The bottom-left corner shows all travel blog entries where this entry was mentioned in a positive context. Finally, the lower left corner shows a heatmap derived from all trajectories that share the same trajectory. During the demonstration, users will be able to specify start and target locations (and, if desired, specific categories of interest) on the presented web interface, e.g., their home and their office. Upon being presented with the popular path skyline, the users may browse different paths and inspect the POIs as well as the additional crowd-sourced information, including the images of PoIs on the route, travel blog entries mentioning POIs on the route, and a heat-map of trajectories covering the route.

References

1. Andersen, O., Jensen, C.S., Torp, K., Yang, B.: Ecotour: Reducing the environmental footprint of vehicles using eco-routes. In: MDM13. (2013) 338–340
2. Graf, F., Kriegel, H.P., Renz, M., Schubert, M.: Mario: xxxmulti-attribute routing in open street map. (In: SSTD’11) 486–490
3. Jossé, G., Franzke, M., Skoumas, G., Züfle, A., Nascimento, M.A., Renz, M.: A framework for computation of popular paths from crowdsourced data. In: ICDE15. (2015) 1428–1431
4. Sacharidis, D., Bouros, P.: Routing directions: keeping it fast and simple. (In: ACM SIGSPATIAL GIS13) 164–173
5. Duckham, M., Kulik, L.: Simplest paths: Automated route selection for navigation. (In: COSIT03) 169–185
6. Westphal, M., Renz, J.: Evaluating and minimizing ambiguities in qualitative route instructions. (In: ACM SIGSPATIAL GIS11) 171–180
7. Garcia, A., Arbelaitz, O., Linaza, M.T., Vansteenwegen, P., Souffriau, W.: Personalized tourist route generation. Springer (2010)
8. Gavalas, Konstantopoulos, C., Mastakas, K., Pantziou, G.: A survey on algorithmic approaches for solving tourist trip design problems. Journal of Heuristics 20 (2014) 291–328
9. Quercia, D., Schifanella, R., Aiello, L.M.: The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. CoRR14 (**abs/1407.1031**)
10. Chen, Z., Shen, H.T., Zhou, X.: Discovering popular routes from trajectories. In: ICDE11. (2011) 900–911
11. Newson, P., Krumm, J.: Hidden markov map matching through noise and sparseness. In: ACM SIGSPATIAL GIS 09. (2009) 336–343

Reference [3] added as an additional reference that is not yet publicly available at time of submission.

A Framework for Computation of Popular Paths from Crowdsourced Data

Gregor Jossé♠ Maximilian Franzke♠ Georgios Skoumas♢
Andreas Züfle♠ Mario A. Nascimento♡ Matthias Renz♠

♠Ludwig-Maximilians-Universität München, Germany
{josse,franzke,zuefle,renz}@dbs.ifi.lmu.de
♡ University of Alberta, Canada mario.nascimento@ualberta.ca
♢ National Technical University of Athens, Greece gskoumas@dblab.ece.ntua.gr

***Abstract*—Directions and paths, as commonly provided by route guidance systems, are usually derived considering absolute metrics, e.g., finding the shortest path within the underlying road network. This demo presents a framework which uses crowdsourced geospatial data to obtain paths that do not only minimize travel time but also guide users along popular points of interest (POIs). By analyzing textual travel blog data and Flickr data, we define a measure for popularity of POIs. This measure is used as an additional cost criterion in the underlying road network graph. Furthermore, we propose an approach to reduce the problem of finding paths which maximize popularity while minimizing travel time to the computation of bicriterion pareto optimal paths. The presented framework allows users to specify origin and destination within a road network, returning the set of pareto optimal paths or a subset thereof if a desired number of POIs along the path has been specified. Each of the returned routes is enriched with representative Flickr images and textual information from travel blogs. The framework and its results show that the computed paths yield competitive solutions in terms of travel time while also providing more "popular" paths, making routing easier and more informative for the user.**

I. Introduction

User-generated content has benefited many scientific disciplines by providing a wealth of new data. The proliferation of smartphones and GPS receivers has facilitated contributing to the plethora of available information. OpenStreetMap[1] (OSM) constitutes the standard example in the area of volunteered geographic information. Authoring geospatial information typically implies coordinate-based, *quantitative data*. Contributing quantitative data requires specialized applications (often part of social media platforms) and/or specialized knowledge, as is the case with OSM.

The majority of users contributing content, however, are much more comfortable using *qualitative information*. People usually do not use geographical coordinates to describe their favorite places or their spatial motion. Instead, it is more common for people to rely on adjectives (such as "great" or "cool") to describe their liking relative to certain points of interest (POIs). Hence, there is an abundance of largely unused geospatial information (freely) available on the internet, e.g., in travel blogs. In contrast to quantitative information, which is mathematically measurable (although sometimes flawed by measurement errors), qualitative information is based on personal cognition. Therefore, accumulated and processed qualitative information may better represent the human way of thinking and feeling.

This is of particular interest when considering the routing problem, i.e., computation of paths in road networks. Traditional routing algorithms only take the structure of the underlying road network into account, for instance, in order to compute shortest or fastest paths, i.e., optimizing w.r.t. inherently quantitative measures. However, in real life, users may be willing to find a trade-off between quantitative measures and qualitative benefit. For example:

- a tourist may be willing to take a detour in order to maximize the number and popularity of POIs on their way to the hotel,
- a commuter driving to work may prefer a slight detour if it yields a significantly nicer route,
- a dog walker might want to avoid busy and big roads altogether and favor recreational walks along landmarks and parks.

There is some existing work in this field, although most papers focus on providing paths which are easier to memorize, describe, and follow. For example, the authors of [1], [2], and [3] try to tackle the problem by introducing cost criteria that allow for a trade-off between minimizing the length of a path while also minimizing the complexity in terms of instructions or turns along the path. The approach most similar to the one presented in this work is [4], which proposes a method for computing beautiful paths, as the authors phrase it. However, in order to quantify quality, the authors rely on explicit statements about the beauty of specific locations, obtained from a platform which collects user opinions on photos of specific locations. In contrast, we propose to mine this kind of information from crowdsourced data. This approach has the crucial advantage that it is scalable as the used information is already available. Having local expert users rate photos one by one, however, can hardly be extended to a global scale.

There are two general problems concerning the computation of qualitative paths. First, quality is not easily quantified. Second, the trade-off between quantitative measure and qualitative benefit is subjective and unknown to the framework. This work attempts to close this gap, by quantifying the quality of a POI by mining crowdsourced (or user-generated) data, yielding a popularity estimation. This estimation is then applied to the underlying road network as an additional cost criterion.

[1]https://www.openstreetmap.org/

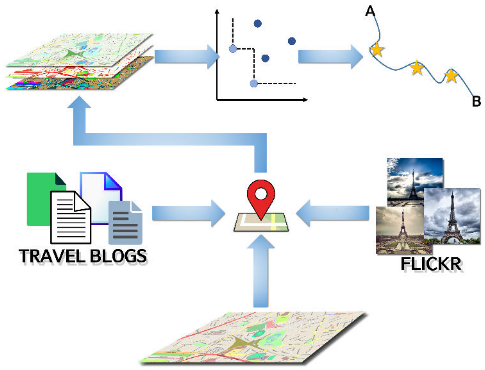


Fig. 1. Data flow chart of the framework, illustrating the data sources as well as the data processing.

Consequently, we obtain a multicriterion graph representing a road network (although for reasons of simplicity we focus on the bicriterion case in our demonstration). Given user input of start and target within the road network, we incorporate state-of-the-art algorithms ([5], [6]) to efficiently compute all pareto optimal paths. Thus, the framework returns all paths between start and target which are optimal under some monotonic function, i.e., the results generated by our framework reflect all practical user preferences. These routes are furthermore enhanced by images and text. The data flow is illustrated in Figure 1.

The challenge of this work is to extract the crowdsourced information (from different sources) and use it to enrich an existing road network. This enriched road network is subsequently used to provide paths that satisfy the claim of high popularity (formally introduced in Section II) while only incurring minor additional travel time.

II. Theoretical Background

In this section we describe the different steps for enriching a road network with crowdsourced geospatial information. First, we describe the sources of information and the respective extraction process. Next, we describe our approach to quantifying quality, i.e., the explicit computation of our measure for popularity estimation. Finally, we discuss our computational methods for path skylines.

A. Quantifying Popularity of Points of Interest

In this work, we choose travel blogs and image datasets as a rich source for crowdsourced geospatial data. This selection is based on the fact that people tend to describe, mention, and photograph POIs that they like or find particularly interesting.

1) Extracting Popularity Information from Text: In a travel blog, tourists express their experiences in relation to journeys taken and places visited. Therefore, places are often associated with qualitative adjectives such as "beatiful","interesting", and "cool". To gather such data, we use classical web-crawling techniques and compile a database consisting of 120,000 texts, obtained from travel blogs as presented in [7]. Extracting qualitative information from text involves the detection of POIs which are mentioned in positive context. The employed approach involves geoparsing, i.e., the detection of candidate phrases containing references to POIs, geocoding, i.e., linking the POIs to geo-coordinates, and sentiment analysis, i.e., the evaluation of such phrases w.r.t. their connotation. Using the Natural Language Processing Toolkit (NLTK), a leading platform for analyzing raw natural language data, we managed to extract 500,000 POIs from the text corpus.

For geocoding, we rely on the GeoNames[2] geographical gazetteer database which contains over eight million POI names and their coordinates worldwide. Whenever possible, POIs extracted from the text corpus are mapped onto geo coordinates. This procedure was successful for 96% of the POIs.

Having identified and geocoded the POIs, the next step is determining the POIs mentioned in positive context. Sentiment analysis, also referred to as opinion mining, is a Natural Language Processing (NLP) problem, which has been thouroughly studied [8]; existing tools perform well on any kind of data. For a given POI p and every phrase mentioning p, the NLTK sentiment analysis provides a score from 0 to 1, reflecting negative to positive context. These scores are then averaged, providing a *travel blog popularity score* $\text{txt}(p) \in [0,1]$ for every POI p.

2) Extracting Popularity Information from Images: For Flickr image data, we use the geotagged dataset provided by the authors in [9] and employ a straightforward popularity estimation approach. We assume a linear correlation between the number of Flickr images in the vicinity of a POI and its popularity within the Flickr community, and we assume this popularity to be an estimation of its general popularity. Thus, for each POI, we aggregate the number of Flickr images within an ε-range (we use $\varepsilon = 100m$ and Euclidean distance) of the POI. Let n_p denote the number of Flickr images in the ε-range of POI p, and let $N := \max n_p$ denote the maximum number Flickr images associated with any POI. Then $\text{im}(p) := n_p/N \in [0,1]$ defines the *image popularity score*.

B. Popularity Graph Enrichment

In a next step, we want to enrich the underlying road network with a combination of both popularity scores. We investigated several strategies and decided upon a compromise between simplicity and effectiveness. Remember that we are interested in paths which provide a trade-off between a quantitative measure (in this case, travel time) and popularity scores. The main challenge is the fact that popularity is a gain, not a cost. Thus, naïve path finding algorithms, which are designed to minimize cost criteria, are not applicable. The "most popular" path would be the solution to the Traveling Salesman Problem among all POIs with a popularity gain > 0. Therefore, rather than considering vertex-associated gain, we transform the score values into edge-associated costs. We now describe this procedure in detail.

Let $G = (V,E,t)$ denote the graph representing the underlying road network, i.e., the vertices $v \in V$ correspond to

[2] http://www.geonames.org/

crossroads, dead ends, etc., the edges $E \ni e = (u, v) \in V \times V$ represent roads connecting distinct vertices. Furthermore, let $t : E \to \mathbb{R}_0^+$ denote the function which maps every edge onto its travel time. We refer to the graph G as *road network graph*. A set of consecutive, acyclic, and mutually different edges is referred to as a *path*. Obviously, the function t naturally extends to any path r, as $t(r)$ is defined as the summed travel time of its edges. Note that our framework and its theoretical background may equally be applied to any other cost criteria (and combinations thereof). However, for reasons of simplicity, we restrict ourselves to travel time as it is probably the most essential criterion for inner city travel.

Let $\mathcal{P}$ denote the set of all POIs. We assume that $\mathcal{P} \subseteq V$, i.e., each POI is also a vertex in the graph. This assumption comes without loss of generality, as we can easily map each POI to the nearest node of the graph or introduce pseudo-nodes. Also, let $\hat{p}(v) := \tau \cdot \mathrm{im}(v) + (1 - \tau) \cdot \mathrm{txt}(v)$ denote the *popularity score* for a vertex $v \in V$. $\hat{p}(v)$ is zero if v is not a POI.

Now, we transform this vertex-associated gain into an edge-associated cost. For each edge $(u, v) = e \in E$, we define the *popularity cost* $p(e)$ as follows

$$p(e) := \phi^{(\hat{p}(u)+\hat{p}(v))/t(e)}$$

where $\phi \in]0, 1[$ is a scaling parameter. Intuitively, $p(e)$ equals 1, if e connects two vertices u and v with popularity score 0. For high popularity scores per distance of e, $p(e)$ approaches 0. Also, $p(e)$ considers the travel time of edge e, such that edges with high travel time require a higher popularity score to maintain the same $p(e)$. If a given road network graph G is enriched with a popularity cost criterion p, we refer to $G' = (V, E, t, p)$ as the *enriched (road network) graph*. Thus, in an enriched graph, every path r is assigned a two-dimensional cost vector, $(t(r), p(r))$ comprising the summed travel time and popularity costs of its edges.

C. Bi-Attribute Skyline Computation

As motivated in Section I, different users may have different prioritization of the given cost criteria. In our case, depending on the type of user (e.g., tourist, dog walker, commuter), travel time is weighed against estimated popularity. Thus, we argue that without any specific knowledge of a user's preferences, we cannot guarantee any path to be optimal to the user. Therefore, we propose to return a set of alternative paths to the user, such that each alternative is pareto optimal w.r.t. to the quantitative and qualitative measures.

Definition 1 ((Pareto Optimal) Popular Paths): Given start and target nodes $s, t \in V$ in an enriched road network graph, the set of pareto optimal paths consists of all paths r between s and t which are non-dominated in the following sense: For each r there exists no other path r' (between s and t) with lower travel time and lower popularity cost, i.e., $\nexists\, r' : t(r') < t(r) \wedge p(r') < p(r)$.

This definition is, of course, a special case of the general definition of multicriterion pareto optimality. To find all Pareto Optimal Popular Paths, we use the framwork for Multi-Attribute Routing in OpenStreetMap (MARiO) [10], where the popularity cost, as defined above, is incorporated as a new cost criterion.

D. k-Constrained Pareto Optimal Popular Paths

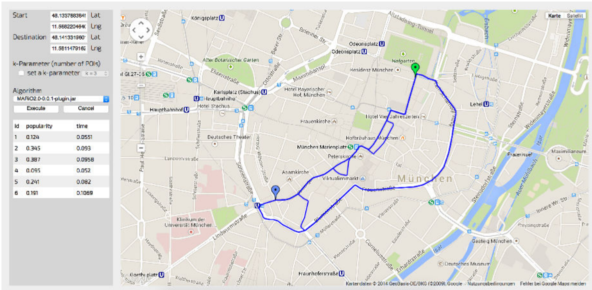
A problem with pareto optimal path computation is the often extensive number of results. Especially in inner city road networks, there exist an abundance of different routing possibilities, each resulting in a slightly different cost vector. As the number of pareto optimal paths is generally unrestricted, we propose to limit the result set in order to improve usability. Also, the POIs along a path may be interpreted as path descriptors. Therefore, it may be of interest to the user to explicitly specify the number of "popular waypoints" along their desired path. Hence, we introduce the *k-constrained pareto optimal popular paths*. Choosing an appropriate k for a specific search task depends on the distance of start and target. Experimentally, a value between 3 and 6 has proven reasonable.

Definition 2 (k-Constrained (Pareto Optimal) Popular Paths): Given an integer k as well as start and target nodes $s, t \in V$ an enriched road network graph, the set of k-constrained pareto optimal popular paths consists of all popular paths that visit at least k POIs.

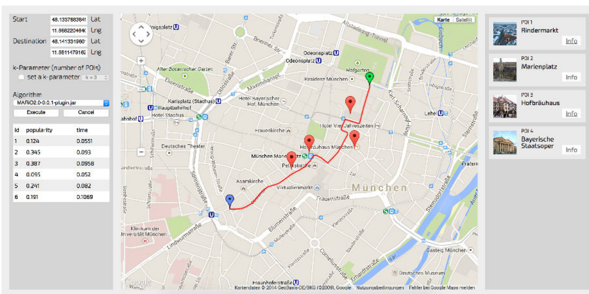
If a path encounters more than k POIs, it is easy to determine the k most popular POIs (simply by comparing scores). Thus, a path which fits the desired parameter is obtained and returned to the user. Concludingly, by returning the k-constrained pareto optimal popular paths, we provide the user with a selection of non-dominated paths which typically encounter highly significant POIs within the respective query city.

III. Framework Description

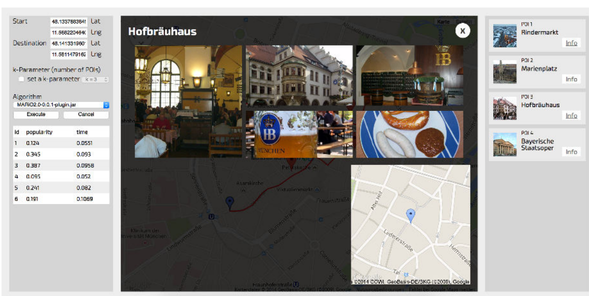
The demonstrated framework enables users to validate that the notion of popularity defined in this paper indeed coincides with the general intuition. The result paths returned to the user yield competitive solutions in terms of travel time while passing POIs perceived as significant, appealing, and/or recognizable. Hence, we solve the proclaimed task of providing "more popular" paths to the user. The two supported query types are described in Section III-A. For both queries, the result paths are presented as a list, each with its associated costs, as well as visualized on a map relying on Google Maps. When selecting a path, the user is provided with a), selected images associated with POIs on the respective path, and b), selected travel blog entries referring to POIs on the respective path. Of course, the availibility of such images and text is dependent on the crowdsourced data and can therefore not be guaranteed. Some features of our framework are shown in Figure 2. Figure 2(a) depicts the main view of the framework which allows one to browse the pareto optimal bicriterion paths w.r.t. user-specified start and target nodes. The lower left corner shows a visualization of these paths, the so-called path skyline, where each path is represented by its two-dimensional cost vector. The list on the left allows to select a specific path which is then displayed on the map with a list of its respective



(a) Illustration of a Bicriterion Path Skyline



(b) Detailed Path Information



(c) Detailed POI Information

Fig. 2. Functionality of the presented framework.

POIs, as in Figure 2(b). As mentioned before, for each POI, additional information can be displayed. Figure 2(c) shows this information for the famous brewery Hofbräuhaus in Munich, Germany.

A. Supported Queries

Our framework supports two different queries. The main task featured in the demonstration is the *popular path query*: Given start and target locations within a road network, our framework provides the user with the set of pareto optimal paths w.r.t. to travel time and the popularity cost. The second query allows to constrain the number of desired POIs along the path, the k*-constrained popular path query*. It returns all pareto optimal paths which visit at least k POIs and presents their k most popular POIs to the user. Obviously, the result set of a k-constrained popular path query is a subset of the result set of a popular path query with the same input. Note, however, that while the popular path query always returns at least one result, the same does not hold for the k-constrained query.

B. Extendability

The developed framework has been built to allow easy extension and further implementation, for instance: (i) Other geospatial sources of crowdsourced data in order to measure the popularity of a POI, such as social check-in data, trajectory data, or other sources of textual data such as news articles can easily be incorporated. Currently, Flickr image data und textual travel blogs are supported. (ii) Further scoring functions which map additional data to popularity scores can be added and the existing may be replaced. Currently, the functions presented in Section II-A1 and Section II-A2 are used. (iii) The graph enrichment function which maps scores of POIs to edge-associated costs is easily interchangeable. Currently, the enrichment function of Section II-B is used. (iv) Different path skyline computation algorithms may be incorporated. Currently, the framework relies on algorithms implemented in the the MARiO framework [10].

C. Demonstration Scenario

During the demonstration, users will be able to specify start and target locations on the presented web interface, such as their home and their office. Upon being presented with the popular path skyline, the users may browse different paths and inspect the POIs as well as the additional crowdsourced information.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme - Marie Curie Actions, Initial Training Network GEOCROWD under grant agreement No. FP7-PEOPLE-2010-ITN-264994. Additional funding has been provided by the IKT II program in the Shared-E-Fleet project, which is funded by the German Federal Ministry of Economics and Technology under the grant number 01ME12107. Mario A. Nascimento has been partially supported by NSERC Canada. The responsibility for this publication lies with the authors.

References

[1] D. Sacharidis and P. Bouros, “Routing directions: keeping it fast and simple,” in *ACM SIGSPATIAL GIS13*, pp. 164–173.
[2] M. Duckham and L. Kulik, “Simplest paths: Automated route selection for navigation,” in *COSIT03*, pp. 169–185.
[3] M. Westphal and J. Renz, “Evaluating and minimizing ambiguities in qualitative route instructions,” in *ACM SIGSPATIAL GIS11*, pp. 171–180.
[4] D. Quercia, R. Schifanella, and L. M. Aiello, “The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city,” *CoRR14*, vol. abs/1407.1031.
[5] M. Shekelyan, G. Jossé, M. Schubert, and H.-P. Kriegel, “Linear path skyline computation in bicriteria networks,” in *DASFAA14*, pp. 173–187.
[6] H.-P. Kriegel, M. Renz, and M. Schubert, “Route skyline queries: a multi-preference path planning approach,” in *ICDE10*, pp. 261–272.
[7] G. Skoumas, K. A. Schmid, G. Jossé, A. Züfle, M. A. Nascimento, M. Renz, and D. Pfoser, “Towards knowledge-enriched path computation,” in *ACM SIGSPATIAL GIS14 (to appear)*.
[8] E. Loper and S. Bird, “Nltk: The natural language toolkit,” in *ETMTNLP02*, pp. 63–70.
[9] H. Mousselly-Sergieh, D. Watzinger, B. Huber, and M. e. a. Döller, “World-wide scale geotagged image dataset for automatic image annotation and reverse geotagging,” in *ACM MMSys14*, pp. 47–52.
[10] F. Graf, H.-P. Kriegel, M. Renz, and M. Schubert, “Mario: Multi-attribute routing in open street map,” in *SSTD11*, pp. 486–490.