



Point-of-interest lists and their potential in recommendation systems

Giorgos Stamatelatos^{1,2} · George Drosatos¹ · Sotirios Gyftopoulos^{1,2} · Helen Briola^{1,2} · Pavlos S. Efraimidis^{1,2}

Received: 21 July 2020 / Revised: 26 December 2020 / Accepted: 15 January 2021 /
Published online: 1 February 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

Location based social networks, such as Foursquare and Yelp, have inspired the development of novel recommendation systems due to the massive volume and multiple types of data that their users generate on a daily basis. More recently, research studies have been focusing on utilizing structural data from these networks that relate the various entities, typically users and locations. In this work, we investigate the information contained in unique structural data of social networks, namely the *lists* or *collections* of items, and assess their potential in recommendation systems. Our hypothesis is that the information encoded in the lists can be utilized to estimate the similarities amongst POIs and, hence, these similarities can drive a personalized recommendation system or enhance the performance of an existing one. This is based on the fact that POI lists are user generated content and can be considered as collections of related POIs. Our method attempts to extract these relations and express the notion of similarity using graph theoretic, set theoretic and statistical measures. Our approach is applied on a Foursquare dataset of two popular destinations in northern Greece and is evaluated both via an offline experiment and against the opinions of local populace that we obtain via a user study. The results confirm the existence of rich similarity information within the lists and the effectiveness of our approach as a recommendation system.

Keywords Points of interest · Foursquare lists · Recommendation systems · Bipartite networks

✉ Giorgos Stamatelatos
gstamatelat@athenarc.gr; gstamat@ee.duth.gr

Extended author information available on the last page of the article

1 Introduction

1.1 User generated content and recommendation systems

Point Of interest (POI) ratings and recommendations are valuable to tourists and enable them to explore new places to visit. It is, therefore, no surprise that there is plenty of active research on recommender systems that, using data from various sources, are able to make personalized venue recommendations. Traditionally, these systems have been relying on text, image or other multimedia content, but with the rapid development of online social networks (OSNs) and location based social networks (LBSNs) (Zheng and Zhou 2011), such as Foursquare and Yelp, link analysis based approaches have gained significant popularity. These LBSNs acted as catalysts to provide an abundance of various forms of data to either inspire the development of novel recommendation systems or enhance existing ones (Jackson and Forster 2010). Typically, this data originates from the actions of the network users themselves and, thus, are often called user generated content (UGC) (Lu and Stepanchikova 2015).

Link analysis based methods for recommendation usually come in the form of explicit or implicit graph structures. In a recommender system where users are being recommended items, UGCs consist of the user–user or user–item structures, or a combination of these. For example, the user–user layout might refer to the similarities in the behavior of users, while user–item might be explicit relations among users and items, such as check-in history.

Although there has been extensive research that is based on user centric (user–user and user–item) link schemes, such as item ratings, GPS trajectories as user–location network, check-in data and similarity among users, there has been a lack of research on link analysis based recommender systems that use an underlying item–item form of relational structure. In this work, we utilize a common feature of social networks and LBSNs, namely the *lists*, to create a collection–item bipartite graph structure and utilize that to infer an item–item topology of POI similarities. These similarities are then used to produce personalized POI recommendations for users by also taking into consideration their profiles.

1.2 Foursquare lists

In this work, we make use of the POI lists information from the Foursquare LBSN, an online portal with rich user generated information, but our method does not rely exclusively on this provider. Our approach is being developed on the plausible assumption that lists are collections of related POIs, an assertion that can be attributed to the features and properties of Foursquare lists. The Foursquare lists feature was implemented in 2011 in order to allow users to keep track of the places they have been or discover places they are willing to go. The lists can be made quickly from a user's check-in history or they can be created from scratch while users can share lists with friends, follow and contribute to public lists. Essentially, this feature

leverages the users' intent to visit places to enable *future check-ins* (or To-Visit lists), a generalization of physical check-ins that is not bound by the users' location history.

Users, naturally, utilize lists for multiple purposes too, but the pattern is that POIs that are listed together are more likely to carry similarities. Another use of lists is for opinionated best places to visit recommendations, often for specific regions, for example "Best places to visit in Chalkidiki". Lists can also portray the favorite places of a user or advertise a certain category of POIs, for example "Sea sports in Kefalonia". The lists contain, in their largest portion, places of entertainment, such as cafes and restaurants, but they also include POIs which fall into different categories, such as landmarks and archaeological sites. Moreover, Foursquare lists capture one more dimension than the traditional check-in history information since users tend to create lists of their check-in histories based on different criteria, usually temporal, spatial or both, for example "Santorini, Summer 2019". This is a further indication that POIs inside a list are related to each other as they are organized based on some criteria that users find reasonable.

Overall, we argue that the users who created the lists, or contributed to them, intentionally grouped together a list of POIs which, in turn, implies that these POIs are, by at least one relevance measure, related to one another. As a result, the coexistence of two POIs in lists is a measure of the likelihood that one user who likes one of the them will like the other one too. The users may be considered as entities that produce sets of relevant POIs according to their own unknown judgement, while our method aims to leverage this phenomenon in order to extract useful information for POI recommendations.

1.3 Our proposed method

Point-of-interest lists, and the assumption that POIs in the same list are related, is the building block of the method that we propose in this paper. Our approach uses this information along with the profile of a user, that specify their preferences with respect to the recommendation context, to offer personalized POI recommendations that correspond to the best places to visit for that particular user. The list data structure can be seen as a bipartite list-location graph, where edges are interpreted as the containment of a POI in a list. Using this bipartite structure, we generate a pairwise similarity matrix of all the POIs, a process that is called *one-mode projection*. Determining, however, when a POI is similar to another may be subject of personal preference and there exist multiple perspectives of how similarity is imprinted in the POI lists. Thus, we use various different weighting methods to generate the similarity matrix and assess the effectiveness of each one. Weighting methods often appear in literature about *link prediction* (Liben-Nowell and Kleinberg 2007; Zhou et al. 2009) or *proximity* (Goyal and Ferrara 2018). Our algorithm, then, analyzes the profile of the user and assigns a relative score on every POI based on their similarity with the user preferences. The final personalized recommendations that are given are considered the POIs with the greatest relative scores. We finally evaluate our approach against a Foursquare dataset in two popular tourist destinations in northern Greece via both an offline experiment and by performing a user survey to obtain the

ground truth for the experiment from inhabitants of these areas. At the same time, we also assess the various weighting methods to uncover their properties and highlight their differences.

Regarding the properties of the proposed approach, we argue that utilizing the Foursquare lists for recommendation systems has some inherent advantages in regards to the methodology and the quality of the results. Initially, as a type of information, Foursquare lists create a richer source for personalized POI recommendation since people often want to visit more places than they actually do. In typical user-location methodology schemes, where this information comes from the check-in history, it is not taken into consideration where the user will check-in in the future. Furthermore, often, these types of user-location graphs capture only part of the information, whereas on a list, POIs are arranged and grouped using specific criteria, usually temporal. Another limitation of check-in history systems is that a user's visit to a location does not necessarily imply that the user liked it. As a result, the check-in history (a single list) might contain POIs that are both interesting and uninteresting to the user and, hence, its contents cannot always be considered related. On the contrary, the action of submitting a POI to a list is more conscious and can even be corrected later since the user can remove the POI from that list.

Moreover, an inherent advantage of the proposed method is the ease of applicability to the new users of the network. The data via which the similarities among the POIs are inferred are already publicly available and, thus, a new user only needs to have a profile stating their preferences to receive personalized recommendations. In particular, a new user can easily be introduced into the proposed system, without the need for their friends to participate or a huge user base to be available, as is often the case in collaborative filtering approaches. This property is referred to as *cold start* (Sertkan et al. 2019), a situation related to a common problem in recommendation systems: the system may not be able to generate recommendations until a significant amount of information has been gathered. Despite not being sensitive to new users, our approach is sensitive to new items as new POIs that appear in the LBSN are required to establish list connections before being incorporated into the pool of potential recommendations. This process is expected to take some time, depending on the popularity of the POI and the amount of users engaging in that context. However, the list structure construction does not require any additional effort as the users are already creating and maintaining lists for their own benefits while, at the same time, they are submitting useful information that can be leveraged for suggestions to other users as well.

Lastly, the use of Foursquare lists does not raise any privacy or royalty issues since it is publicly available and accessible via the Foursquare API. For example, the check-in history that is the core of multiple approaches to recommendation is considered personal data while a list might express the same or more information without exposing sensitive data. The approach we describe in this paper is based on purely structural data and it is worth noting that we treat Foursquare lists simply as their ID and no additional information is required, such as the user who created that list or its name. This further extends the field of applications to domains where this information is very difficult to acquire or not available at all, for example where lists are implicit. In addition, our method is easy to reproduce and can be performed at

real time because the similarity matrix among the POIs can be computed ahead of time and the main component of the recommendation is a very simple mathematical operation, a weighted sum.

1.4 Contribution

Our contributions can be summarized as:

- We highlight the aspect of lists of POIs: a generalization of check-in history and a richer source of information for personalized POI recommendation systems.
- We argue that lists of POIs, as primarily information provided by users, are considered *User Generated Content* (UGC) and should be more openly addressed in the literature of tourism applications.
- A rich Foursquare dataset of POIs and POI lists that was acquired over a period of several months is shared with this work.
- The attainment of considerable improvements over the existing modified MI methodology via a comprehensive study of multiple similarity measures, which are also contrasted in the context of tourism.

1.5 Outline

The rest of the article is organized as follows. In Sect. 2, we explore recent literature about recommendation systems based on LBSNs, graph processing recommendation and entity similarity approaches. In Sect. 3, we present the Foursquare dataset of POIs and POI lists that was created as part of this work. Our methodology and the representation of the POI list dataset are presented in Sect. 4. In Sect. 5, we evaluate the system and demonstrate the existence of rich information within the Foursquare lists and the effectiveness of our method as a recommendation system. Finally, Sect. 6 concludes this paper and presents suggestions for future work.

2 Related work

A first attempt that utilizes lists of POIs for recommender systems was presented in Karagiannis et al. (2015). In that work, the similarities among POIs are approximated based on a modification of Mutual Information (MI), while in this paper, we make a complete recommendation system and evaluate it against local inhabitants' opinions. The current system is, furthermore, capable of incorporating user preferences for personalized suggestions and, at the same time, we enrich the similarities with more weighting functions, both set theoretic and graph theoretic, and show that they can outperform the modified MI measure. Moreover, the results among the similarity measures are juxtaposed and useful conclusions are drawn from these that determine the properties of each similarity function. In the rest of this section, we present previous work that is related to the views expressed in this paper.

2.1 The role of LBSNs in recommendation

Developments on mobile devices and the emergence of more advanced online tourism portals has provided researchers with motivation and valuable data to support the investigation of POI recommendation based on LBSNs. Current state of the art is given in Bao et al. (2015), Ravi and Vairavasundaram (2016) and Eirinaki et al. (2018) and the references therein. In general, POI recommendation systems can be categorized by methodology as content based, link analysis based and collaborative filtering based.

This paper constitutes a link analysis method of standalone personalized POI recommendations, where the links refer to the explicit relations among POIs and lists. In particular, we infer a POI–POI network based on this bipartite POI-list network. This structure appears to be a less studied field in the literature, where typically the links refer to user–user or user–POI relations. Several of the recommender systems in the literature can also be applied to a context other than tourism, such as movies, shopping items and books. Similarly, they can be utilized for user recommendations, for example to suggest “people you may know”, “people to follow” or activity recommendation, such as sightseeing, boating and jogging.

2.2 Recommendation using graph processing

The existence of explicit or implicit links in social networks has motivated the use of graph processing and graph theoretic approaches for POI recommendation.

In Wu et al. (2015), the authors studied options of clustering POIs and users based on information from geographic social networks. In particular, they considered the *social distance* between two POIs as the Jaccard index of the users that have checked-in in those POIs and leverage this index to partition the POIs into groups of similar places. In a recommendation perspective, the fact that two commercial places belong to the same cluster indicates that there is a high likelihood that a user who likes one place will also be interested to visit the other.

Two datasets containing geolocation and temporal information concerning users in 11 cities were utilized in Noulas et al. (2012) in order to evaluate a proposed recommendation algorithm in comparison to several known techniques. The methodology was based on a random walk in the graph of users and POIs, on which the edges represent friendships among users or check-in actions among users and POIs, while the recommendations were ranked according to the hitting probabilities of the random walk. The authors mentioned that all versions of collaborative filtering in their experiments, that were supposed to better model users’ preferences, fail to outperform the popularity based baseline.

In Wang et al. (2013), the authors suggested a recommendation algorithm that operates using different factors: (a) past user behavior (visited places), (b) the location of POIs, (c) the social relationships among the users, and (d) the similarity

between users. By analyzing the publicly available data of Gowalla,¹ they showed that more than 80% of the new places visited by a user are in the 10km vicinity of previous check-ins and more than 30% of the new places visited by a user have been visited by a friend or a friend-of-a-friend in the past. These facts imply that geographical and social information significantly affect the choices of a user when deciding which new place to visit. Therefore, it is also desirable for a recommender system to take these components into consideration.

Moreover, in Kefalas et al. (2018), a recommendation system was proposed that incorporates user time-varying preferences as well. In particular, the recommendation was based on a tripartite graph, consisting of users, locations and *sessions*. Sessions can be considered as a more specialized form of POI lists, where the criterion of POI coexistence in the same session is temporal. The authors concluded that the time dimension plays a very important role in recommender systems.

Finally, an interesting work that aimed to identify and validate the heuristic factors affecting the popularity of “best places to visit” recommendations was presented in Li et al. (2019). This empirical study focused on the explicit best places to visit listings in Qyer.com, a concept that is also a subset of Foursquare lists which contain opinionated recommendations (“Best places to visit in...”).

2.3 Similarity concepts in tourism

Often, similarity measures are utilized in order to express the relations among entities in a link-based system, when these are not explicitly present. For example, in Celik and Dokuz (2018), the researchers utilized data from Twitter to extract the check-ins of users and proposed a methodology for revealing the socially similar users based on their online traces. A bipartite network between tourists and POI reviewers was used in Ahmed et al. (2017) to drive a collaborative filtering approach to recommendation. The user–user similarity weights were inferred using the Jaccard index and the cosine similarity, while the method was evaluated on a Foursquare dataset. Another establishment and tourism portal, TripAdvisor, was utilized in Van der Zee and Bertocchi (2018) as a medium of social network analysis application on the POI reviews. An indirect similarity matrix of POIs was created based on the two-mode network of users and reviewed POIs; the similarity between two POIs was defined as the intersection cardinality between the users that reviewed these POIs. In this paper, we heavily utilize the notion of similarity and we argue that there are multiple perspectives as to what constitutes similarity. As a result, we propose multiple measures to express this similarity among POIs and perform extensive experiments to compare the properties of those functions.

The notion of similarity among entities can be extended, besides POIs or users, to other layers as well. For example, in Preoțiuc-Pietro et al. (2013), the concept of city similarity was studied, where each city is represented by the collection of POIs in it. In particular, a city is represented as a vector of the categories of the POIs within.

¹ A LBSN that operated until 2012, primarily under its mobile application.

Another form of similarity was discussed in Sertkan et al. (2019), the attribute similarity, which defines the pairwise relation among different tourism attributes, such as island, mountains, river, family, diving and others. Finally, in David-Negre et al. (2018), similarities among tourism domains (TripAdvisor, Booking, Trivago) were approximated and denote whether they have been used by the same tourists.

2.4 POI sequence recommendations

An interesting concept in the field of POI recommendations is the conception of trip planning, which refers to the recommendation of a collection of POIs that are bound by a common characteristic. Such systems can be useful for trip planning as the recommendation describes a collection of venues for a trip or a route. Some POI group recommendation systems were the subject of recent papers (Cenamor et al. 2017; Rakesh et al. 2017; Arentze et al. 2018; Wörndl et al. 2017). Trip planning has an indirect connection to POI lists because a trip, as an abstract set of POIs, can be thought as a POI list with the characteristics posed in this paper: a collection of related POIs.

Another interesting concept is the session-based recommendation approaches which are recommendation techniques that aim to predict the user's immediate next actions (Quadana et al. 2018), such as next-item recommendation (Hidasi et al. 2016; Song et al. 2015) and list continuation (Hariri et al. 2012). In Ludewig and Jannach (2018), the authors present the results of an in-depth analysis of a number of complex algorithms, such as recurrent neural networks and factorized Markov model approaches, as well as simpler methods based on nearest neighbor schemes. Their results indicate that the simpler methods are equally well than more complex approaches based on deep neural networks.

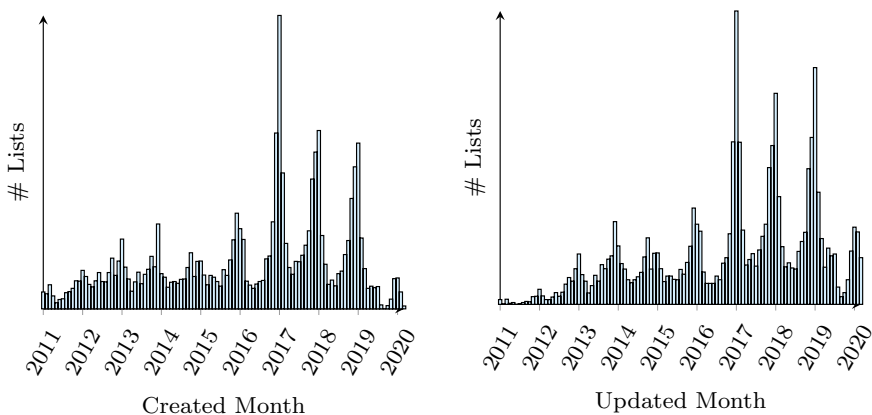
In such systems, it is a common technique to use offline evaluation schemes. In the domain of next-track music recommendation, often only the last element of a sequence is hidden, while in the recommendation of videos in streaming platforms (Hidasi et al. 2016) an approach is taken where the number of hidden elements is incrementally increased. In the case of next-track music recommendation (Hariri et al. 2012), the authors use leave-one-out cross validation since they do not take into account the creation time of the playlists. In this work, we also utilize leave-one-out cross validation for the evaluation of our method during the offline study.

3 The Greek POI list dataset

For the purposes of our project we retrieved and assembled a dataset of POIs and lists from Foursquare covering a large geographic portion of Greece. In particular, the dataset contains 47,745 POIs located in 9 prefectures of tourist interest and 17,000 lists with 115,308 mentions of the POIs. The dataset exhibits a wide diversity of categories regarding the POIs (e.g., restaurants, hotels, car-rentals, archaeological sites, shopping malls, gyms, rivers, churches) while the covered area includes popular touristic destinations (e.g., Rhodes, Santorini, Chalkidiki). The geographical

Table 1 Dataset summary as the number of POIs in each prefecture and their occurrences in lists

Prefecture	POIs	Occurrences in lists
Chalkidiki	7590	14,291
Cephalonia	2612	3671
Cyclades	2171	19,489
Dodecanese	5524	10,160
Drama	1077	245
Kavala	4002	11,129
Serres	1970	400
Thessaloniki	21,266	55,532
Xanthi	1533	391
Total	47,745	115,308

**Fig. 1** Created (left) and updated (right) dates distribution of the POI lists. Each bar corresponds to one month from August 2011 to October 2020

breakdown of the POIs is presented in Table 1. The complete dataset contains data that were retrieved until October, 2020, and is available online.² The dataset can be continuously augmented as long as the retrieval process is performed since new POIs are added in Foursquare and its users create new lists every day.

A temporal examination of the lists in Foursquare reveals interesting characteristics about the users' behavior. Figure 1 presents the histograms of the per month list creation and update dates from 2011 to 2020. It is more than evident that the two distributions bare great resemblance since the vast majority of the local maxima follow the same pattern. The maxima correspond to each August of the respective year, a month that traditionally demonstrates a peak in tourism attention. The years 2017, 2018 and 2019 contain the majority of lists with 52% created and 57% updated

² <https://doi.org/10.17632/dx39jx9v5p.1>.

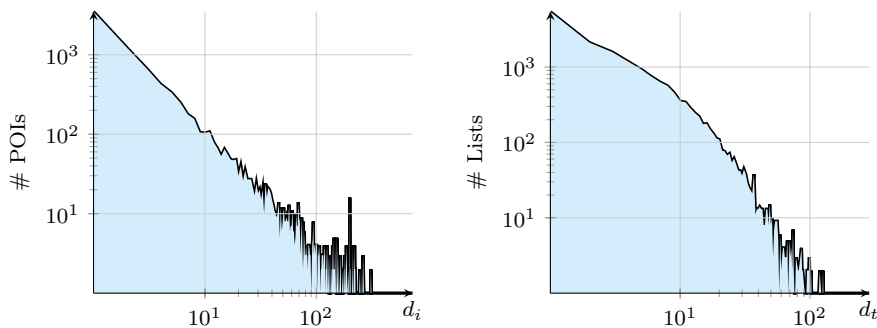


Fig. 2 Log-log density distribution of occurrences for POIs (left) and lists (right) in the full dataset

Table 2 Summary of the case study dataset as the number of POIs and occurrences in lists

Area	POIs	Lists	Occurrences in lists
Thessaloniki	2871	6534	49,448
Kassandra	526	1286	6336
Total	3397	7820	55,784

during this period while 2020 exhibits a significant drop in generated content, possibly related to the COVID-19 pandemic.

The information contained in the two date fields (i.e., the date of creation and the date of the latest update) might potentially be very useful, e.g., for filtering very old lists which could be outdated. In our work, we use the dataset as acquired by the Foursquare API in order to simplify the process and demonstrate the potential of lists without preprocessing or filtering. We argue that most of the dates in the dataset are up-to-date since the majority of the lists were either created or updated recently. Despite this, utilizing this field, for example to filter the most recent lists, could be pursued in the future, possibly in combination with other list metadata, for example their title or their authoring user.

Further analysis of the dataset highlights interesting properties that support our incentive to utilize Foursquare lists in recommendation systems. Figure 2 presents the log-log plot of the occurrences distributions for the POIs and lists in the dataset. The plots provide solid indications about a power law behavior of the distributions, a feature that is commonly observed in datasets of recommendation systems (Goel et al. 2010; Abdollahpouri et al. 2017; Belletti et al. 2019). This phenomenon can be attributed to the fact that people tend to perform additive actions on POIs (i.e., insert into list) with a higher rate for POIs that are already popular (Clauset et al. 2009).

In this study, we examined the performance of our methodology in a case study using a preliminary subset of the complete dataset that was assembled until February, 2020, and focused on two cities with vivid touristic activity, namely *Thessaloniki* and *Kassandra*. These two areas are in northern Greece, contain multiple types of leisure and entertainment attractions and are both established as notable tourist destinations; Thessaloniki is the second largest city in the country, and Kassandra

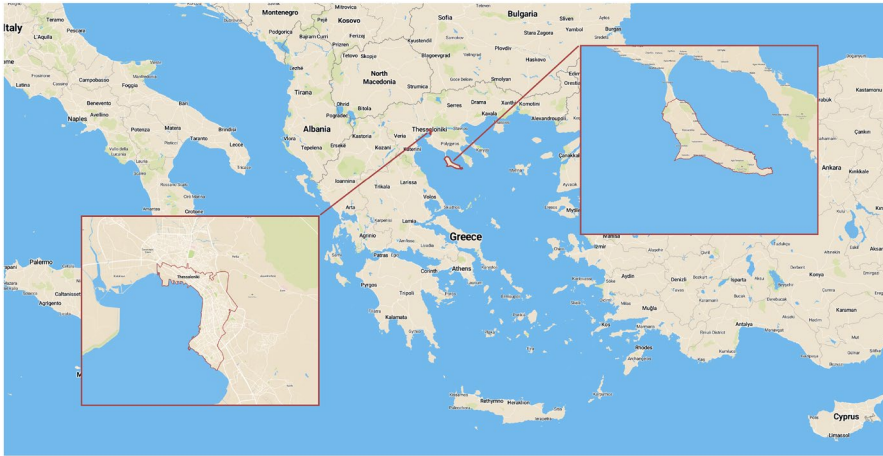


Fig. 3 Map of Greece showing the two areas of our dataset where we perform the case study experiment

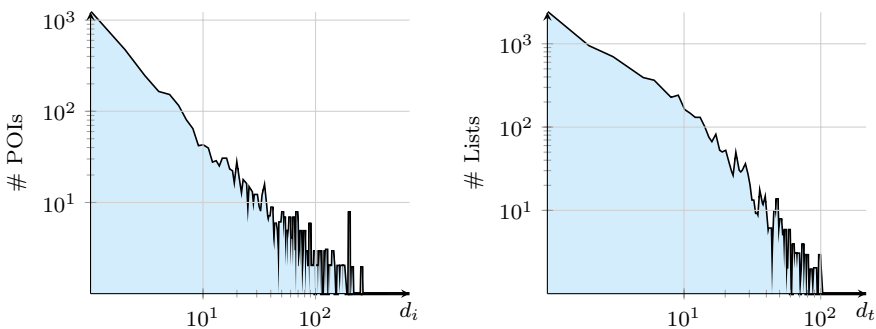


Fig. 4 Log-log density distribution of occurrences for POIs (left) and lists (right) in the case study dataset

is the most visited area of the suburban region of Chalkidiki. Our case study dataset contains 2871 POIs in Thessaloniki and 526 in Kassandra and 7820 lists with 55,784 mentions to POIs forming, thus, a subset of the complete dataset. The geographical breakdown of the data is summarily presented in Table 2 while Fig. 3 shows the geographic areas that are covered by the case study.

Analytical findings on the preliminary assembled dataset confirm the consistency of certain characteristics with the augmented complete dataset. Figure 4 presents the log-log plot of the occurrences distributions for the POIs and lists in the data subset. A prominent observation of the plots, when compared with the respective plots of the complete dataset in Fig. 2, is that the data exhibit a remarkably similar behavior that resembles a power law distribution and indicate a scale free property. Hence, we can safely argue that our case study is performed based on a representative subset of the data and the evaluation of our methodology on this subset provides credible

insight of its performance on the complete dataset. This property is a preliminary positive indication about the sufficiency of the quantity and quality of the available data. However, a more thorough investigation of this issue would be an important future work direction and would facilitate further applications of our approach.

4 Methodology

The method we present in this section comprises a personalized POI recommendation system that uses the Foursquare lists as the only source of knowledge. The system requires a user profile as a preference vector of POIs and is capable of quantifying the preference score of an arbitrary POI that is not present in the profile. Thus, our method is able to predict a relative score that the user would assign to this POI, or otherwise how likely they are to like the POI with respect to other POIs. This method can operate as a personalized recommendation system by assigning relative preferences to all candidate POIs in the context and selecting the ones with the maximum values. On another perspective, it can also be used as a negative recommender to suggest POIs least related with the user profile (where *not to go*).

This section is organized based on the components of our method. In Sect. 4.1, we present the structure of the list dataset and its representation as a bipartite graph. The similarities among the POIs are then approximated in Sect. 4.2 using various similarity functions. In Sect. 4.3, the user profile is introduced into the system while the main component of our method that assigns relative scores to POIs is explained in Sect. 4.4.

4.1 Dataset representation

The dataset that we use to drive our approach consists of the Foursquare user generated lists that refer to the geographic areas of interest, which are the areas that the recommender system is to be deployed and operated. The extent of the lists also corresponds to the geographic extent of the operation of the recommender, such as a city or even a whole country. Every list in the dataset can be portrayed as a collection of one or more POIs, such that the same POI may exist in multiple lists while there could also exist multiple identical lists (with exactly the same POIs).

A natural interpretation of the lists is a bipartite or two-mode graph, where one set of vertices is the lists and the other set of vertices is the POIs. List x and POI y are connected via an edge iff y belongs in list x . An abstract example of the POI lists structure as well as its representation as a bipartite graph is given in Fig. 5. Each list is only described by an arbitrary ID and no additional meta-information is required while each POI is a named location and refers to a specific business or attraction.

4.2 POI similarity matrix

The similarities among the POIs in the dataset are, then, estimated using the lists based on the assumption that every list is a collection of related POIs. This relation

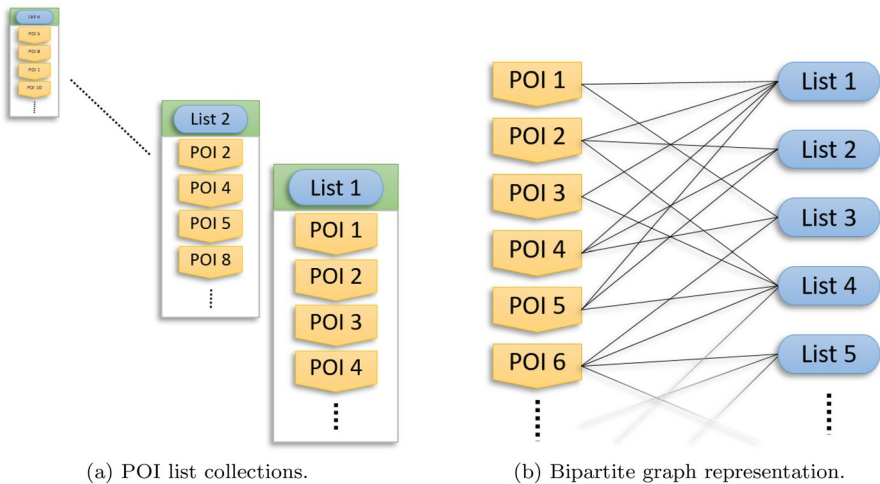


Fig. 5 (a) Abstract example of lists as collections of related POIs and (b) its representation as a lists—POIs bipartite graph, where edges have a “POI is submitted in the list” relation

among the POIs defines the likelihood that a user who likes one POI might like another POI; this likelihood is determined by the strength of the similarity between two POIs. The similarity is imprinted in the social network by its own users and might comprise a variety of components that make this information suitable for recommendation. For example, POIs may be grouped in a collection based on their categories or a characteristic of their categories (type of music for a bar, type of food for a restaurant etc), or they might signify trips with time and distance constraints. Moreover, items in the same list may exhibit complementary characteristics as this is often the desired behavior of tourists in holiday trips. As a result, two POIs with high similarity value may not necessarily have similar features but display similarities based on other criteria.

Overall, users of the social network, using their own judgement, place similar POIs in the same list as a means of grouping them under the same name or context. As a result, these POI lists can provide essential information regarding the homogeneity and diversity of the places contained within them, properties that can be exploited for our recommendation system. Ultimately, a recommendation system is only useful to the social network users so it is sensible to utilize the definition of similarity as portrayed by these same users. These similarities are the building block for our method and are used for creating our personalized recommender system.

Inferring the pairwise similarities from a given bipartite graph is often referred to as projection, an extensively used method for compressing information about bipartite networks (Zhou et al. 2007). The one-mode projection of a bipartite network $G = (X, Y, E)$ onto X (X projection for short) is a weighted, complete, unipartite network $G' = (X, E')$ containing only the X nodes, where the weight of the edge between i and j is determined by a weighting function $\beta_G(X_i, X_j)$. The weighting method may not necessarily be symmetrical but in this work we engage in a simpler approach with commutative weight functions so that

Table 3 Summary of the similarity weighting methods β_G

β	Name	Formula	Description
aa	Adamic	$\sum_{i \in N_{11}} \frac{1}{\log(n'_i)}$	Adamic/Adar index
is	Intersection	n_{11}	Number of common lists
jac	Jaccard	$\frac{n_{11}}{n_{10} + n_{01} + n_{11}}$	Intersection over union
ka	Modified MI	Karagiannis et al. (2015)	Modified MI
ku	Kulczynski-2	$\frac{n_{11}(n_x + n_y)}{2n_x n_y}$	Intersection over harmonic mean
mi	Mutual information	Manning et al. (2008)	Mutual information of sets
cos	Ochiai	$\frac{n_{11}}{\sqrt{n_x n_y}}$	Cosine similarity or intersection over geometric mean
ov	Overlap	$\frac{n_{11}}{\min(n_x, n_y)}$	Intersection over minimum
ρ	Phi	$\frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_x * (n - n_x) * (n - n_y) * n_y}}$	Pearson correlation coefficient
sr	SimRank	Jeh and Widom (2002)	Iterative calculation of SimRank
f1	Sorensen	$\frac{2n_{11}}{n_x + n_y}$	Sorensen–Dice index or F1 score or intersection over arithmetic mean

For two POIs x and y , N_{11} is the set of lists which contain both x and y , N_{10} is the set of lists that contain x but not y , N_{01} is the set of lists that contain y but not x and N_{00} is the set of lists that do not contain either x or y . Their respective cardinalities are denoted as $n_{pq} = |N_{pq}|$. We also set $n = n_{11} + n_{01} + n_{10} + n_{00}$, $n_x = n_{10} + n_{11}$ and $n_y = n_{01} + n_{11}$. Finally, n'_i is the number of POIs contained within list t

$\beta_G(X_i, X_j) = \beta_G(X_j, X_i)$, resulting in an undirected projection. Typically, the weight function expresses a form of similarity among the vertices in order to preserve the semantics of the original graph, which is perfectly suited in this scenario as it identifies our goal: to create the similarity matrix among the POIs.

While the projection allows to capture and quantify the similarity among the POIs that is imprinted by the social network users inside the venue lists feature of Foursquare, it is less informative than the original bipartite graph and, thus, an appropriate weighting method β_G is required, that minimizes this information loss. There exists, however, no universally accepted weighting method of minimizing information loss and, as a result, we proceed with a selection of set and graph theoretic functions that expose the similarity among the POIs, which are given in Table 3. The coarse functionality of these weighting methods is that two POIs with more common lists will be more likely to be related or similar.

Most of the weighting methods that we use are trivial to compute as they simply rely on set theoretic measurements, but others are more computationally demanding. For example, the original SimRank has large time and space requirements (Jeh and Widom 2002). However, the process of creating the similarity matrix needs to be executed only once, or when the primitive list data need to be refreshed. Afterwards, the recommendation algorithm can be performed directly over these projections.

We refer to a projection or a similarity matrix using a weighting method as S or S_β when it refers to a specific projection, with $S(x, y)$ denoting the similarity value between x and y .

4.3 User profile

The user profile is utilized to account for the personalized nature of our approach and allow the personalized suggestions. The profile is a preference vector for a single user, which corresponds to a vector of POIs, each one of which has a numerical preference value attached to it. For these values, we use the 5-level Likert scale ranging from “Strongly Uninteresting” to “Strongly Interesting”, with one neutral level, to encode the preference magnitude. The Likert scale is, then, transformed to the necessary arithmetic scale and in this problem we use an integer scale from -2 to 2 , as it makes more sense in a semantic standpoint and will be justified in the next subsection. It is worth noting that our algorithm can operate under any number of Likert levels as long as it is symmetric and a higher value corresponds to a higher preference. In fact, it can be applied on profiles with decimal preferences as well but, since it is dependant on human input, a 5-level Likert scale is widely used and easier to be interpreted.

Naturally, POIs in the profile need to exist in the dataset with each POI belonging in at least one list, so that the similarities with other POIs can be expressed. However, most of the weighting methods cannot work reliably with just a single list, in particular the ones that rely on the intersection (n_{11}). A POI having very little representation in the lists is easier to have its intersection with other POIs trivialized to zero and, thus, most of its similarities to other POIs also being zero. As a result, having a significant number of POIs with weak list representation can degenerate the dataset into a point where the majority of similarity values will be zero and most POIs unable to be distinguished from one another. For this reason, we suggest using POIs with strong presence in the bipartite graph as profile. In addition, the quality of the profile can be affected by the amount of POIs as well as the diversity of the preference values within it; a user profile that covers the full range of the scale and utilizes all the available levels can possibly better portray the fine differences among the POIs in the profile.

Finally, it is important to distinguish the context of the profile and the recommendation, which may differ. For example, the profile might be referring to the city of residence of the user while the recommendation context will typically be the user’s trip destination. Our method can be applied on this scenario too as long as it is technically feasible with the presence of lists connecting POIs among the contexts.

For a set of POIs p_1, p_2, \dots, p_m , we refer to the user profile as P , where $P(i)$ is the preference association for p_i and, for the purposes of our experiments, is an integer scale in $[-2, 2]$ corresponding to the 5-level Likert scale.

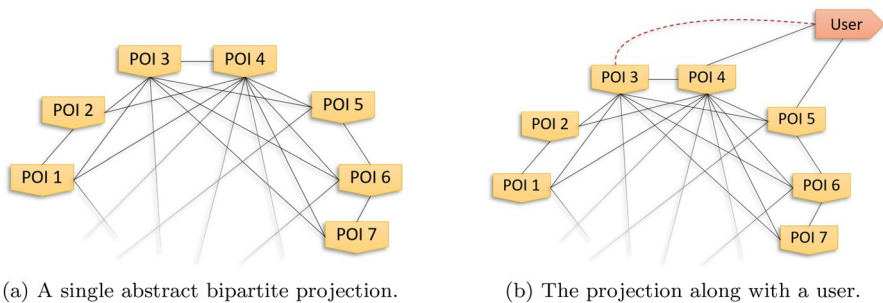


Fig. 6 (a) The transformation of the bipartite graph into its projection using a similarity measure and (b) the same graph with a user attached. The user states preferences for POI 4 and POI 5 and the algorithm assigns a relative preference score to POI 3 (dashed edge)

4.4 Recommendation function

The core of our approach consists of the personalized recommendation algorithm that is capable of assigning relative preference scores to arbitrary POIs based on the similarity matrix and the personal profile of the user. Given a similarity matrix S and a user profile P , for an arbitrary POI $q \notin P$ the relative preference score w predicted by the algorithm is defined to be

$$w(q) = \sum_{i \in P} P(i)S(i, q), \quad (1)$$

which corresponds to the weighted sum of the similarities of the profile POIs with q , and the weights being the profile preferences themselves. The concept of a weighted sum has also appeared in Sarwar et al. (2001), where the pairwise similarities between items and the user preferences are being utilized for recommendations.

The premise of this formula is to provide a higher relevant score w to POIs that have high similarity values with POIs in the profile that have high preference scores. For example, a POI q with high average similarity with the profile POIs ranked high (2 in our scale) will score high in the preference prediction, because of the weighted nature of the equation. Naturally, it is reasonable to assume that POIs that are very similar to “interesting” POIs will likely be interesting themselves. In essence, the algorithm driven by Eq. (1) is based on the intuition “Recommend the most related POIs to the POIs that the user finds interesting”.

The nature of this equation leads to interesting properties regarding its applicability and physical interpretation. First, the preference score w can be interpreted only relatively to other POIs and, hence, defined as *relative preference score*. This is due to the fact that it depends on the values of the similarity scores, which also have magnitudes with meaning only relative to others. As a result, the sign of the relative preference score is also not indicative of the item relevance to the specific user and, hence, a negative score might not necessarily imply a negative relevance. Moreover, the application of this equation across different profiles is not possible due to the absence of normalization and the scores cannot be interpreted relatively

across users. Because of this, it is only meaningful to use this score when comparing two POIs and for a specific user profile, for example to answer the question “Which of these two POIs is more relevant to the interests of this user?”. Finally, we can now clarify why a symmetric scale $[-2, 2]$ used in the profiles is more natural instead of another, such as $[1, 5]$. A symmetric scale guarantees that POIs that have been marked as uninteresting will have a negative impact; the method will be subtracting their most similar POIs while being encouraged to include their least similar ones. Similarly, neutral profile values that correspond to the arithmetic value of zero are irrelevant and can also be omitted in the symmetric scale.

Figure 6 shows one bipartite projection, where a weighted edge represents the magnitude of similarity between the adjacent POIs. A user and their profile can be seen in the projected graph as a new vertex with edges across the profile POIs and weights equal to the preference scores. Our algorithm considers the projection as well as the profile to calculate a relative preference score for another arbitrary POI that exists in the dataset.

Based on Eq. (1), the recommender system can be constructed by simply including the top quantities for a similarity matrix S and a user profile P . In particular, the algorithm can make preference predictions for a set of POIs in an area, rank them with respect to their predicted relative weights and offer the top- N recommendations. The recommendations can also be filtered based on categories, areas, or other user defined criteria. These calculations are simple to process and in certain cases can be performed in real time, as S is already computed once. The complexity of the recommendation is, hence, proportional to the filters assigned to the recommender system.

An inherent limitation to our algorithm is due to the relevant nature of the preference predictions. In an extreme scenario where all of the candidate POIs in the recommendation context are irrelevant to the user and their interests (i.e., they do not like any of them), our recommendation algorithm will result in personalized suggestions that are also irrelevant. The reason is that a high preference prediction does not necessarily mean an interesting POI but rather a POI that is more interesting than the rest. This restriction is a common aspect of ranking algorithms and, typically, can be circumvented by expanding the recommendation context in such extreme cases.

5 Evaluation results

The effectiveness of our methodology is evaluated against the case study dataset described in Sect. 3. In particular, we use the two areas *Thessaloniki* and *Kassandra* and evaluate our method using POIs and POI lists in these regions. The evaluation is threefold and consists of the *preference evaluation scheme* (Sect. 5.1), the *recommendation evaluation scheme* (Sect. 5.2) and the *offline evaluation* (Sect. 5.3). During the preference evaluation, we assess the relative preference predictions of Eq. (1) on arbitrary POIs, which can have positive or negative impression on the experiment users. For the recommendation evaluation we use the same method that was used in the TREC 2016 Contextual Suggestion Track (Hashemi et al. 2016), which

evaluates the system as a recommender, only includes top preferences for each user and is performed in two stages. Lastly, for the offline experiments, the system is cross validated by using some of the POI lists as virtual profiles in order to complement the online survey responses. During all evaluation schemes, we also assess the effectiveness of the different similarity measures and make observations about the properties of each one. As for the ground truth, we incorporated a user study in the form of a questionnaire with participants familiar with the attractions of the use case areas.

The implementation of most of the similarity measures on the dataset can be done by simply treating the POIs as the set of lists within which they are contained and applying the formulas in Table 3. There are only two special cases of measures that are worth mentioning in an implementation perspective: the Adamic/Adar index and SimRank. Regarding the Adamic/Adar index, we consider the quantity n'_t (number of POIs within list t) to be the number of POIs that the Foursquare API advertises as belonging to list t , which is the intended way of using this index. This quantity might not be in agreement with the number of POIs in list t in our dataset due to the way that the data were acquired, which was biased towards the POIs, and is, therefore, smaller than the true n'_t . For the implementation of SimRank, we treat the dataset as a bipartite graph as shown in Fig. 5b and use the original computation algorithm. In addition, SimRank has a parameter c which is usually set to 0.8 (Li et al. 2010), but we use the value 0.6 too as a means of relative comparison and denote them as $sr8$ and $sr6$, respectively.

5.1 Preference evaluation

The preference evaluation scheme aims to assess the effectiveness of the relative preference scores that are returned by Eq. (1), which is the basis of our algorithm. Simultaneously, we are doing a comparative analysis of the results and performance of all projection weighting methods.

Initially, we select the POIs that will participate in this experiment based on several criteria and conclude with 19 POIs in Thessaloniki and 11 POIs in Kassandra. These POIs are attractions that match the tourism type of these areas, namely cafes, bars, beaches or restaurants. They are selected with Foursquare rating uniformity in mind in an attempt to collect answers that contain a more balanced preference distribution. The ratings are ranging from 6 to 10, as there were very few to no venues with rating less than 6. For POIs that were around the same rating, we intentionally considered the ones with the most amount of lists so that the users participating in the survey were more likely to have visited them or to have a developed opinion about those. We refer to these 19 and 11 POIs as *survey POIs*.

The survey participants were asked to rate each of these POIs based on how interesting they personally find them. The survey options were compatible with the 5-level Likert scale and were:

1. Very Uninteresting
2. Uninteresting

How would you rate these tourist attractions in Thessaloniki?

In each question you are given one attraction in the city of Thessaloniki and you are asked to rate the attraction based on how interesting "you" find it. Select "1" for a very uninteresting attraction and "5" for a very interesting one according to your preferences. You can select "No Opinion" if you are uncertain about a specific attraction. The attractions can be any type of tourist destination, such as cafe, restaurant and beach.

The responses will be used to assess the effectiveness of attraction recommendation algorithms.

* Required

Estrella *

<https://foursquare.com/v/estrella/513c8a8e4b0e75b7b9a5051>

Choose

1 - Very Uninteresting

2 - Uninteresting

3 - Neutral

4 - Interesting

5 - Very Interesting

No Opinion

[2f48be4b0f6f503ac6888](https://foursquare.com/v/estrella/513c8a8e4b0e75b7b9a5051)

[stores/5a1acfa183e3804e8aca355b](https://foursquare.com/v/estrella/513c8a8e4b0e75b7b9a5051)

How would you rate these tourist attractions in Chalkidiki?

In each question you are given one attraction in the city of Chalkidiki and you are asked to rate the attraction based on how interesting "you" find it. Select "1" for a very uninteresting attraction and "5" for a very interesting one according to your preferences. You can select "No Opinion" if you are uncertain about a specific attraction. The attractions can be any type of tourist destination, such as cafe, restaurant and beach.

The responses will be used to assess the effectiveness of attraction recommendation algorithms.

* Required

On The Rocks *

<https://foursquare.com/v/on-the-rocks/501d6275e4b0a80051c352>

Choose

1 - Very Uninteresting

2 - Uninteresting

3 - Neutral

4 - Interesting

5 - Very Interesting

No Opinion

[4de19b4488773cb686313917](https://foursquare.com/v/on-the-rocks/501d6275e4b0a80051c352)

[pck1aib-har/55434a07499e38ff84b71d0](https://foursquare.com/v/on-the-rocks/501d6275e4b0a80051c352)

Fig. 7 Screenshots of the user survey in the two regions of Thessaloniki and Kassandra

3. Neutral
4. Interesting
5. Very Interesting

We also included “No Opinion” to account for cases where users wanted to refrain from expressing an opinion and ignore these responses in our analysis. Overall, we received 31 user responses for Thessaloniki and 16 responses for Kassandra but we only considered the responses that had at least half of the POIs answered, which were 28 and 14, respectively. Cropped screenshots of the survey for both areas are displayed in Fig. 7.

The survey POIs and the survey responses are utilized both as profile and ground truth via the leave-one-out cross-validation method. According to this method, one POI k is held out from the profile and its rate is attempted to be predicted; the process can be summarized in the following steps:

1. Compute the relative recommendation score $w(k)$.
2. Normalize $w(k)$.
3. Compare the normalized w vector against the ground truth vector.
4. Repeat the process over all users and projections.

For each missing POI k , the relative preference weight $w(k)$ is computed via Formula 1 with k itself excluded from the user profile. Then, we apply a small transformation to $w(k)$, specific to this experiment, due to technical limitations of the cross-validation method. In particular, because each profile with an excluded POI is considered a different profile, the relative scores $w(k)$ cannot be interpreted relatively

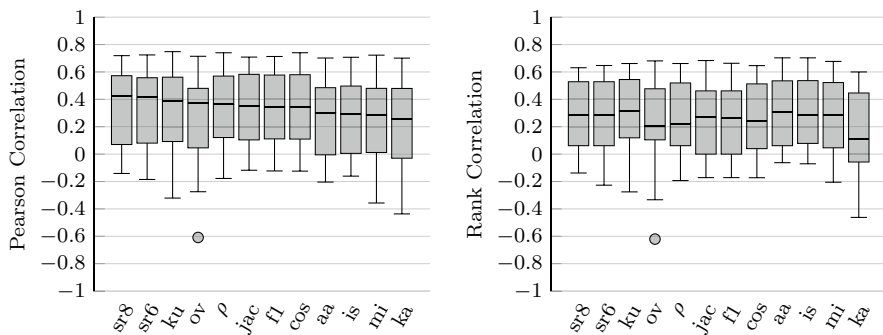


Fig. 8 Preference evaluation scheme results for Thessaloniki (28 data points). Each box displays the 4 quartiles of the user distribution. The measures are ordered by the median of the Pearson correlation without considering the outliers

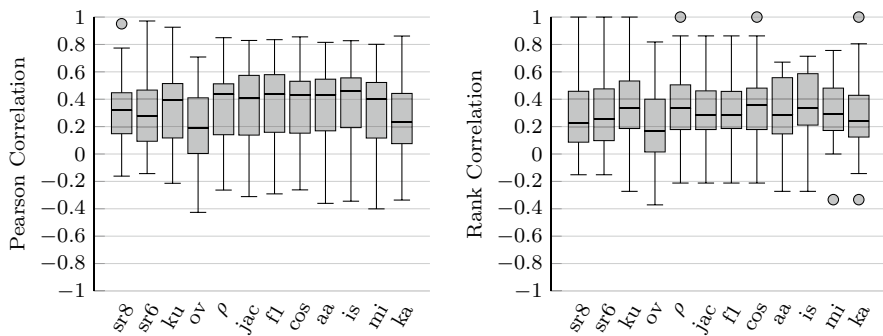


Fig. 9 Preference evaluation scheme results for Kassandra (14 data points). Each box displays the 4 quartiles of the user distribution. The ordering is the same as Fig. 8 for consistency

across these profiles, even if they refer to the same user; more information has been established in Sect. 4.4.

As a simple way to overcome this technical difficulty, we normalize the predicted scores $w(k)$ with the sum of the profile preferences $\sum_{i \in P} P(i)$. Hence, we get a vector of predicted preference values, one for each survey POI which is then compared to the ground truth vector of the preferences stated in the user profile. The comparison is done using the Pearson correlation coefficient, which is a measure of the linear association between these two vectors, and Kendall tau-b correlation (Agresti 2010). Tau-b correlation is a rank correlation measure and a generalization of the Kendall tau-a coefficient that accounts for ties in the input lists, specifically present in the distinct 5-level preferences of the survey. This process is repeated over all users and projections.

Figures 8 and 9 display the evaluation results of Thessaloniki and Kassandra respectively as box plots. The results are grouped by similarity measure to better

convey the differences among them. It should be noted that, regarding the rank correlation, there exists a maximum τ value that the algorithm can achieve due to the discrete survey answers, where two POIs can be tied on the same rank. The same situation is very unlikely to occur in the output of the algorithm due to the floating point nature of the operation, which in practice disallows ties on the same rank. As a result, the maximum correlation can be shown to be equal to the correlation of the ground truth vector with the “best” possible untied vector (the flattened ground truth vector without ties):

$$\tau_{max} = \frac{T(n) - \sum_i T(t_i)}{\sqrt{T(n)}\sqrt{T(n) - \sum_i T(t_i)}}, \text{ where } T(x) = \frac{x(x-1)}{2} \quad (2)$$

and $t = [t_1, t_2, t_3, t_4, t_5]$, with t_1 being the number of POIs in the “Very Uninteresting” group, t_2 being the number of POIs in the “Uninteresting” group etc. For this reason, we only show the ratio of tau-b correlation to τ_{max} on the rank correlation plots. Following Eq. (2), it is also evident that the quantity τ_{max} may be different for each profile. Increasing the number of possible ranks in the survey would diminish this behavior but at the cost of user convenience.

In general, we have mixed feelings about the preference evaluation as there exist both strong and weaker results. In particular, it seems as the values are consistent, with most of the similarity measures firmly around the same restricted range of values. Interestingly, there is a small discrepancy between the Pearson and rank correlation as there does not appear to be a clear winner in both settings. Furthermore, in all settings, the majority of the projections outperformed the modified MI index (ka) and the overlap coefficient (ov), the latter of which also contains the only outlier in Thessaloniki. Both versions of SimRank (with parameters 0.6 and 0.8) are the top performing similarity measures in Thessaloniki and their top two quartiles can be considered as good results (Akoglu 2018). Although this statement is not true for Cassandra too, this, along with the observation that results in Cassandra appear to be better, can be attributed to the smaller sample size and, therefore, the reduced difficulty of the problem. As a result, the two geographic regions cannot be directly compared to one another. We argue that, even with this small size of questionees set, our complex method and its results indicate firm evidence of the fact that lists contain information useful in the context of POI recommendation and prompts us to perform experiments on our system as a recommender.

5.2 Recommendation evaluation

The recommendation evaluation scheme aims to assess the effectiveness of the personalized recommendations that result from assuming the top quantities of a user profile and projection scenario. The major difference between this scheme and the preference evaluation is that now we only assess the POIs that are most relevant to the users: those with the top relative preference scores. We are using the top-5 quantities as this appears to be the most dominant setting. Similarly, as in the preference

evaluation scheme, we are also doing comparative analysis among the weighting methods.

The process is based on the TREC 2016 Contextual Suggestion Track (Hashemi et al. 2016), where the responses of Sect. 5.1 survey are used exclusively as profiles. Specifically, by using each response as profile, we produced the top-5 results for all weighting methods and aggregated them. While some of the recommended POIs were common across the weighting methods, the cardinalities of the recommendations' unions were all more than 5 POIs. Since every user rated the profile POIs differently, the resulting aggregated POIs that contain the combined recommendations of all the weighting methods are different for each rater. The combined recommendations of a rater are then evaluated against another survey that is forwarded to that rater. This new questionnaire has the same structure as the previous survey in terms of the Likert scale levels and is personalized for each user as the recommendations are different for them. The recommendations are performed for the same context as the profile so that the same user that receives the additional survey is more likely to have prior experience about the recommended POIs. The recommendation evaluation scheme was only done in Thessaloniki as we did not manage to get enough responses from users to complete the experiment in Kassandra. Specifically, for Thessaloniki, we received an extra 8 responses, of which 1 had very little information filled in and, thus, we consider 7 of these replies.

The recommendation system is evaluated using the following measures:

1. Precision at rank 5 ($P@5$)
2. Normalized discounted cumulative gain at rank 5 ($NDCG@5$)
3. Mean reciprocal rank (MRR)

which were also part of the evaluation of the submissions of the TREC 2016 competition. $P@5$ shows the fraction of the personalized suggestions that the user liked, where "liked" refers to the evaluation as "Interesting" or "Very Interesting"; the other 3 Likert options were considered as non relevant. Since the suggestions were 5 for each weighting method, the values of $P@5$ can be 0, 1/5, 2/5, 3/5, 4/5 or 1, with 1 meaning that the user liked all 5 of the algorithm recommendations and is considered the perfect score. However, because there were responses marked as "No Opinion", in these cases we use $P@4$ or even $P@3$, depending on how many answers were missing. Moreover, the $NDCG@5$ measure is an indication of the ranking quality of the 5 personalized recommendations and, like DCG, does not take into account whether the recommendation was relevant or not but only the ranking of the top POIs. For this, we are using 5 ordered, arithmetic classes, one for each response option. A value of 1 corresponds to the scenario where the user evaluated the 5 suggestions in the same ranking as the algorithm did. Finally, the MRR measure shows the average of the ranks of the first relevant ("Interesting" or "Very Interesting") suggestions. A value of 1 means that on all cases the first recommended POI was liked by the evaluator. A value of 0.5 means that, on average, the first recommended POI was irrelevant but the second was relevant. Lastly, a value of 0 means that there

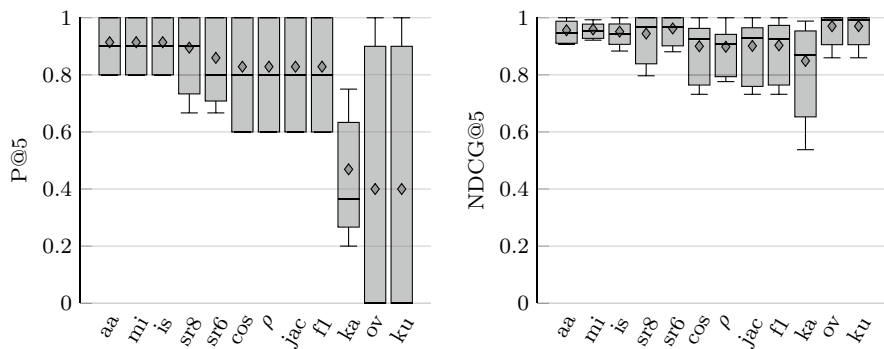


Fig. 10 Recommendation evaluation scheme results for Thessaloniki (7 data points). Each box displays the 4 quartiles of the user distribution while the average is displayed as a diamond symbol. The measures are ordered by the average P@5

Table 4 Recommendation evaluation using the MRR measure

	aa	mi	is	sr8	sr6	cos	ρ	jac	fl	ka	ov	ku
MRR	1.00	1.00	1.00	0.88	0.80	0.93	0.79	0.83	0.83	0.72	0.33	0.67

The ordering is the same as Fig. 10 for consistency

was no relevant POI in the recommendation list. The nature of MRR dictates special considerations because of the presence of “No Opinion” values in the surveys: where the MRR cannot be accurately calculated, we do not consider it at all in the average calculation. One example would be a ranking where the first relevant POI was below one or more POIs marked with “No Opinion”, in which case it is unclear what the value of MRR should be. As a result, there were some weighting methods with just 4 user responses considered; this is the also the minimum amount.

Figure 10 presents a box plot of the recommendation evaluation with P@5 and NDCG@5. The Adamic/Adar index (aa), MI (mi), intersection (is) and SimRank (sr8, sr6) appear to perform almost flawlessly, with average precision over 0.9; in practice more than 9 out of 10 POIs on average were correctly recommended. However, because our sample size is relatively small, it is unclear which of these is the best. Interestingly, most of the similarity functions performed very well at NDCG@5, even the weaker measures and a possible explanation is due to the granularity of the method, where only 5 discrete values are present, of which the two highest were used almost exclusively. Table 4 shows the MRR values of the experiments, which seem to be in agreement with the P@5 results. Deviations from this behavior is because MRR only considers the first relevant POI and, as a result, is biased towards mostly the top result, instead of P@5 that assigns equal weight to all 5 of the suggestions.

Further examination of the results reveal an interesting observations about the varied performance of SimRank. The only POIs that both versions (0.6 and 0.8) of this weighting method failed with respect to the P@5 measure appear to be 3 distinct

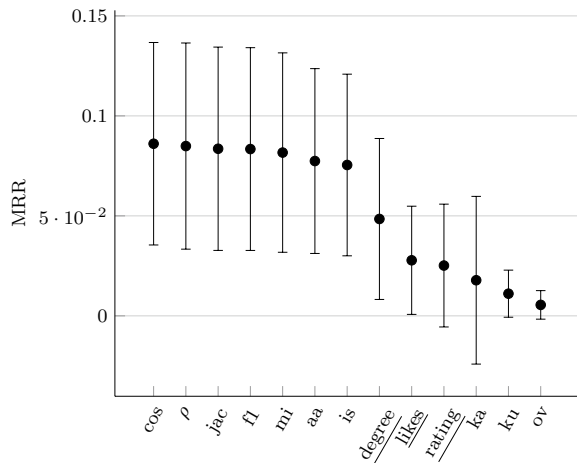
recommended POIs within a single list each and no other metadata, such as rating, reviews or other information. It is possible that these POIs are wrong submissions that happened by some Foursquare users in an effort to identify a check-in or another spatially identified action. In fact, 2 of these POIs, 7 months after the acquisition of the dataset, are now not part of any list in Foursquare, possibly because it was an error that was corrected by the users themselves. We believe that these POIs should not have been included in the experiment in the first place because of the trivialized information contained in their records and an extension to this experiment would confirm the increased accuracy in this scenario. Even though we did not perform any filtering in the dataset besides the necessary POIs that are contained within at least one list, we still consider the existing results very strong in terms of recommendation effectiveness.

5.3 Offline evaluation

Finally, the offline evaluation aims to complement the limited survey responses of the online experiments by utilizing user generated information that is already present in the LBSN. In this section, we describe the offline experiment via which we evaluate our recommendation method by considering the POI lists as virtual user profiles. Despite the lists themselves being part of the dataset, the assumption that lists can be used as user profiles is reasonable since they are being created by users to represent one dimension of their own interests. For this study, we have adopted evaluation techniques that are common in the domains of session-based recommendation or next-track (playlist) recommendation that is further explored in the related work section. In particular, a list is being considered as a user, and the recommender is used to predict held-out elements of the list. This evaluation process is explained in more detail below. Lastly, in this experiment, we have included 3 popularity baselines about the POIs: the number of likes as reported by Foursquare, the rating of the POI as obtained by the Foursquare API (decimal number in the scale [0, 10]) and the degree of the POI (number of lists containing that POI).

When considering a certain list as a virtual user profile, that list must be removed from the dataset before the recommendation process. Since a removal from the dataset can interfere with the pairwise similarity factors, the complete bipartite projection must also be computed every time we use a list as a profile. As a result, the lists that contain POIs with singular degree cannot be used as profiles as those POIs would be missing from the dataset and, despite existing in the profile, would not be eligible recommendation targets. For consistency with the online experiments we also exclude the lists that contain POIs with degree 2 as their removal from the dataset would leave these POIs existing in a single list. Furthermore, we did not consider lists with less than 6 POIs as profiles as an attempt to increase the list degrees; for comparison, the online surveys consisted of 11 and 19 POIs for Chalkidiki and Thessaloniki respectively. Lastly, we only utilized the lists with the *collaborative* field set to *false* as profiles. This restriction constitutes an additional check to further establish that only the lists that were made by a single person are being used as profiles, since collaborative lists might correspond to inconsistent interests. The

Fig. 11 Results of the offline evaluation. The error bars display the average MRR and the standard deviation. The baseline measures degree, likes and rating are underlined. Each projection corresponds to 747 virtual profiles



lists with this property were less than 2% of the total lists. Given the aforementioned constraints, there are 747 lists that do not fall into one of the filtering categories and are used as profiles in this experiment. Despite using only these 747 lists as profiles for practical reasons, all of the lists are being utilized as part of the dataset.

For each list being used as a virtual profile, we perform a leave-one-out cross-validation and attempt to predict the missing item from the list in a way similar to the preference evaluation scheme. In particular, for every missing item, the rest of the profile is used to create a ranked recommendation list for all POIs in the dataset based on their relative preference weights (Eq. 1) and note the position at which the correct (missing) one appeared. Because lists do not explicitly contain preferences as they were not created to represent profiles, we consider all POIs in a list to have to best preference (Very Interesting). The MRR measure is used to capture the average rank at which the missing item was predicted. The MRR value of 1 corresponds to the situation where all POIs of the virtual profile were correctly predicted in the first position of the ranked recommendation list, a value of 0.5 means that on average they were on the second rank.

Figure 11 shows the results of the offline evaluation grouped by similarity measure. Each projection is identified by the 747 lists that we use as virtual profiles and display the average MRR and the standard deviation of the MRR as an error bar. In this scenario, the average MRR corresponds to the average of the average reciprocal rank. The plot includes the underlined baselines of degree, likes and rating as mentioned previously. For these baselines, the recommendation ignored the profile and returned the POIs sorted based exclusively on the respective measure. Essentially, the average of the average rank of the missing POI in a list was 17.75 for Ochiai (cos), measured under 3,397 POIs, which corresponds to 0.5 percentile average position. Another observation is the performance of the ka, ku and ov measures which are below baseline, a property that is in alignment with Fig. 10. The performance of the other measures (cos, ρ , jac, f1, mi, aa, is) is also in alignment with the online experiments and their effectiveness

is almost identical. As a result, there does not appear to be a similarity measure with clear advantages over the others in this group, an observation that is consistent with the online recommendation experiment. Moreover, these 7 similarity measures are considerably more effective than the baselines and produce better recommendations than simply recommending the most popular. Finally, among the baseline recommenders, degree appears to be the most prominent while likes and rating have similar effectiveness.

Due to technical restrictions and its computational complexity restrictions SimRank is not included in this experiment. Specifically, every projection had to be recreated for each of the 747 virtual profiles because it had to be removed from the dataset in advance. As a result, this process was prohibitive for SimRank as the similarities had to be recomputed for each virtual profile. However, based on the results of the online study and the consistencies with the offline experiments, it is possible that SimRank would also perform equally well with the rest of the 7 top measures.

5.4 Results discussion

Our twofold evaluation analysis yields interesting combined results about the algorithm and the similarity measures that we use, and allows us to juxtapose the results and compare our approach on two settings. Initially, we discuss the observation that our approach appears to be better in the online and offline recommendation evaluation but only moderate in the preference evaluation. This can be due to a combination of reasons, for example the algorithm may simply be more effective at identifying the top quantities, or the top POIs contain more information (such as more lists) in the LBSN and, thus, easier to extract. Additionally, it may also be a result of the difficulty of the problem itself, since positioning an element in a ranking is generally more difficult than assigning a weight to it. Nevertheless, neither of these approaches is trivial as it has to scale for a large set of POIs in the recommendation context, possibly in the magnitude of thousands or more.

Regarding the use of the similarity measures, it can be observed that some of them are effective when measured via one evaluation method but mediocre via the other scheme. This may indicate that not all similarity functions are appropriate for every circumstance and each one may be used more effectively depending on the individual problem. In addition, the results suggest that not all similarity functions are appropriate for the study of this paper. For example, the overlap function (ov) appears to deliver weaker results in comparison with the rest of the similarity measures.

The combination of the online evaluation schemes, however, seems to converge into the conclusion that SimRank is marginally more consistent and effective. Specifically, SimRank with parameter 0.8 (sr8) is better than the 0.6 version (sr6), but this difference is negligible. Interestingly, SimRank is the only global similarity index that takes the properties of the whole network into consideration, as opposed to the local similarities that only consider local information (Schall 2015, Section 2.2). While SimRank leverages the whole bipartite graph to compute the similarity among x and y , the rest of the measures only take into consideration the lists

of x and y . An exception to this is the Adamic/Adar index which is a unique measure that considers local information (the intersection of the lists between two POIs) as well as the degrees of the intersected lists as an additional step. Thus, if the set theoretic measures are considered as 1-step operations (the adjacent nodes of the POIs), the Adamic/Adar index is a 2-step operation (the adjacent of the adjacent of the POIs) and the global indices are a multi-step operation (values that propagate through the graph until convergence). The same concept is also described in Goyal and Ferrara (2018) using the terms first-order, second-order and higher-order proximity.

Based on these observations, the pattern that emerges is that global similarity measures perform better than local similarity functions in the context of this paper. Specifically, while SimRank appears to be on average a well balanced projection in terms of effectiveness, the Adamic/Adar index, which is a semi-global index, also seems to be on par, particularly on the recommendation evaluation scheme. Whether this hypothesis is true, and to what extent, remains to be seen in future work experiments and in a larger scale.

The instances that our approach failed can be attributed to the difficulty of determining when a POI is similar to another. Each individual may have a different judgment when asked to decide if two POIs are similar as they are subjective to their own preferences and, hence, failures will naturally occur in such recommendation systems. For example, when comparing two cafes, some may conclude that they are similar based on the kind of music that they play while others based on the art style or atmosphere of the two POIs. As a result, similarity, in a general sense, may be difficult to express using a single measure for every occasion. In fact, since lists are user-generated content, they might contain a form of average of the criteria of what people think constitutes similarity, and, thus, one may have to selectively pick lists that express a particular form of similarity to satisfy perspectives of different users.

6 Conclusions

The purpose of this paper was to study the potential of user generated Point-of-Interest lists in recommendation systems. We proposed a methodology that only takes into consideration the bipartite graph structure of the POIs-lists graph, that is present in the Foursquare lists feature, to develop a POI recommendation system based on user profiles. This was made possible via similarity criteria that operate under this bipartite graph and can formulate the similarity between pairs of POIs. Our assumption was that information about the POIs is encoded in the lists and, in particular, that similar POIs will be submitted under the same list. Our evaluation confirmed, up to a certain extent that this assumption is reasonable. Specifically, we performed a user survey using local volunteers and used the responses to drive our algorithms and the evaluation process. The results indicated a significant correlation between the output of our method and the responses of the locals, particularly during the evaluation of the system as a personalized POI recommendation system. The

offline evaluation strongly confirmed our approach and was in agreement with the findings of the online experiments.

We conclude this paper by arguing that social network analysis can be an effective way of obtaining information about how users perceive POIs. Our methods took advantage of large amounts of publicly available user generated content to extract useful conclusions about the relationships among POIs, are very easy to reproduce and can be performed in real time. Through this work we hope to inspire the use of lists in further research as we believe the amount of useful information in this structure is profound and its potential is not limited to a recommendation or a points-of-interest perspective.

Future research on POI lists should focus on the extension of the graph structure as a way to leverage more available user generated data. A possible expansion is the incorporation of users in the graph, where they can be related to other entities as list creators or contributors. This scheme can be modeled as a tripartite graph of users, lists and POIs and can be processed using specific graph theoretic notions. The incorporation of users can expand the approach into possibly a user-centric method, for example based on user similarities or even in combination with a user relationship method such as collaborative filtering. Furthermore, while in this work we made use of the lists in the Foursquare LBSN, we believe our methods can be applied to any portal that utilizes user generated lists, which might not necessarily be POI-related. For example, Amazon users can specify list of products (Listmania), and IMDb users can define their own lists of entities like movies and actors. In addition, Twitter has a lists feature via which users can group other profiles under the same combined feed. Similarly to our lists, these platforms share characteristics with respect to the physical interpretation of lists and, often, it is reasonable to consider them as groups of related entities. It is a question of future research to investigate the applicability of our approach to these or other platforms.

Acknowledgements This work has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH CREATE INNOVATE (project code: T1EDK-02474, Grant no.: MIS 5030446). We would also like to thank the anonymous reviewers who helped us improve our work, and especially for suggesting the offline experiment of Section 5.3.

References

- Abdollahpouri H, Burke R, Mobasher B (2017) Controlling popularity bias in learning-to-rank recommendation. In: Proceedings of the Eleventh ACM Conference on Recommender Systems, ACM, New York, NY, USA, RecSys '17, p 42–46. <https://doi.org/10.1145/3109859.3109912>
- Agresti A (2010) Analysis of ordinal categorical data, 2nd edn. Wiley, New York
- Ahmedi L, Rrmoku K, Sylejmani K, Shabani D (2017) A bimodal social network analysis to recommend points of interest to tourists. Soc Netw Anal Min 7(1):14. <https://doi.org/10.1007/s13278-017-0431-8>
- Akoglu H (2018) User's guide to correlation coefficients. Turk J Emerg Med 18(3):91–93. <https://doi.org/10.1016/j.tjem.2018.08.001>

- Arentze T, Kemperman A, Aksenov P (2018) Estimating a latent-class user model for travel recommender systems. *Inf Technol Tourism* 19(1–4):61–82. <https://doi.org/10.1007/s40558-018-0105-z>
- Bao J, Zheng Y, Wilkie D, Mokbel M (2015) Recommendations in location-based social networks: a survey. *GeoInformatica* 19(3):525–565. <https://doi.org/10.1007/s10707-014-0220-8>
- Belletti F, Lakshmanan K, Krichene W, Chen YF, Anderson J (2019) Scalable realistic recommendation datasets through fractal expansions. [arXiv:1901.08910](https://arxiv.org/abs/1901.08910)
- Celik M, Dokuz AS (2018) Discovering socially similar users in social media datasets based on their socially important locations. *Inf Process Manag* 54(6):1154–1168. <https://doi.org/10.1016/j.ipm.2018.08.004>
- Cenamor I, de la Rosa T, Núñez S, Borrajo D (2017) Planning for tourism routes using social networks. *Expert Syst Appl* 69:1–9. <https://doi.org/10.1016/j.eswa.2016.10.030>
- Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. *SIAM Rev* 51(4):661–703. <https://doi.org/10.1137/070710111>
- David-Negre T, Almedida-Santana A, Hernández JM, Moreno-Gil S (2018) Understanding European tourists' use of e-tourism platforms. Analysis of networks. *Inf Technol Tourism* 20(1–4):131–152. <https://doi.org/10.1007/s40558-018-0113-z>
- Eirinaki M, Gao J, Varlamis I, Tserpes K (2018) Recommender systems for large-scale social networks: a review of challenges and solutions. *Future Gener Comput Syst* 78:413–418. <https://doi.org/10.1016/j.future.2017.09.015>
- Goel S, Broder A, Gabrilovich E, Pang B (2010) Anatomy of the long tail: Ordinary people with extraordinary tastes. In: Proceedings of the third ACM international conference on web search and data mining, ACM, New York, NY, USA, WSDM '10, p 201–210. <https://doi.org/10.1145/1718487.1718513>
- Goyal P, Ferrara E (2018) Graph embedding techniques, applications, and performance: a survey. *Knowl-Based Syst* 151:78–94. <https://doi.org/10.1016/j.knosys.2018.03.022>
- Hariri N, Mobasher B, Burke R (2012) Context-aware music recommendation based on latent topic sequential patterns. In: Proceedings of the sixth ACM conference on recommender systems—RecSys '12, ACM, Dublin, Ireland, p 131. <https://doi.org/10.1145/2365952.2365979>
- Hashemi SH, Clarke CL, Kamps J, Kiseleva J, Voorhees EM (2016) Overview of the TREC 2016 contextual suggestion track. In: 25th text retrieval conference—TREC '16. NIST, Gaithersburg, Maryland USA, pp 1–10
- Hidasi B, Quadrana M, Karatzoglou A, Tikk D (2016) Parallel recurrent neural network architectures for feature-rich session-based recommendations. In: Proceedings of the 10th ACM conference on recommender systems, ACM, Boston Massachusetts USA, pp 241–248. <https://doi.org/10.1145/2959100.2959167>
- Jackson LS, Forster PM (2010) An empirical study of geographic and seasonal variations in diurnal temperature range. *J Clim* 23(12):3205–3221. <https://doi.org/10.1175/2010JCLI3215.1>
- Jeh G, Widom J (2002) Simrank: a measure of structural-context similarity. In: 8th ACM SIGKDD international conference on knowledge discovery and data mining (KDD '02), ACM, New York, NY, USA, pp 538–543. <https://doi.org/10.1145/775047.775126>
- Karagiannis I, Arampatzis A, Efraimidis PS, Stamatiatos G (2015) Social network analysis of public lists of POIs. In: 19th Panhellenic Conference on Informatics - PCI '15, ACM, Athens, Greece, pp 61–62. <https://doi.org/10.1145/2801948.2802031>
- Kefalas P, Symeonidis P, Manolopoulos Y (2018) Recommendations based on a heterogeneous spatio-temporal social network. *World Wide Web* 21(2):345–371. <https://doi.org/10.1007/s11280-017-0454-0>
- Li C, Han J, He G, Jin X, Sun Y, Yu Y, Wu T (2010) Fast computation of SimRank for static and dynamic information networks. In: 13th international conference on extending database technology - EDBT '10, ACM, Lausanne, Switzerland, p 465. <https://doi.org/10.1145/1739041.1739098>
- Li L, Lee KY, Yang SB (2019) Exploring the effect of heuristic factors on the popularity of user-curated 'best places to visit' recommendations in an online travel community. *Inf Process Manag* 56(4):1391–1408. <https://doi.org/10.1016/j.ipm.2018.03.009>
- Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. *J Am Soc Inform Sci Technol* 58(7):1019–1031. <https://doi.org/10.1002/asi.20591>
- Lu W, Stepchenkova S (2015) User-generated content as a research mode in tourism and hospitality applications: topics, methods, and software. *J Hosp Mark Manag* 24(2):119–154. <https://doi.org/10.1080/19368623.2014.907758>

- Ludewig M, Jannach D (2018) Evaluation of session-based recommendation algorithms. *User Model User-Adap Inter* 28(4–5):331–390. <https://doi.org/10.1007/s11257-018-9209-6>
- Manning CD, Raghavan P, Schütze H (2008) *Introduction to information retrieval*. Cambridge University Press, Cambridge
- Noulas A, Scellato S, Lathia N, Mascolo C (2012) A random walk around the city: New venue recommendation in location-based social networks. In: 2012 international conference on privacy, security, risk and trust and 2012 international conference on social computing, IEEE, pp 144–153. <https://doi.org/10.1109/SocialCom-PASSAT.2012.70>
- Preoțiuc-Pietro D, Cranshaw J, Yano T (2013) Exploring venue-based city-to-city similarity measures. In: 2nd ACM SIGKDD international workshop on urban computing—UrbComp '13, ACM, Chicago, Illinois, p 1. <https://doi.org/10.1145/2505821.2505832>
- Quadran M, Cremonesi P, Jannach D (2018) Sequence-aware recommender systems. *ACM Comput Surv* 51:4. <https://doi.org/10.1145/3190616>
- Rakesh V, Jadhav N, Kotov A, Reddy CK (2017) Probabilistic social sequential model for tour recommendation. In: 10th ACM international conference on web search and data mining—WSDM '17, ACM, Cambridge, United Kingdom, pp 631–640. <https://doi.org/10.1145/3018661.3018711>
- Ravi L, Vairavasundaram S (2016) A collaborative location based travel recommendation system through enhanced rating prediction for the group of users. *Comput Intell Neurosci* 2016:1–28. <https://doi.org/10.1155/2016/1291358>
- Sarwar B, Karypis G, Konstan J, Riedl J (2001) Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th international conference on World Wide Web, ACM, New York, NY, USA, WWW '01, p 285–295. <https://doi.org/10.1145/371920.372071>
- Schall D (2015) Link prediction for directed graphs, vol 2. Springer, Cham, pp 7–31. https://doi.org/10.1007/978-3-319-22735-1_2
- Sertkan M, Neidhardt J, Werthner H (2019) What is the “Personality” of a tourism destination? *Inf Technol Tourism* 21(1):105–133. <https://doi.org/10.1007/s40558-018-0135-6>
- Song Q, Cheng J, Yuan T, Lu H (2015) Personalized recommendation meets your next favorite. In: Proceedings of the 24th ACM international conference on information and knowledge management, ACM, New York, NY, USA, CIKM '15, pp 1775–1778. <https://doi.org/10.1145/2806416.2806598>
- Wang H, Terrovitis M, Mamoulis N (2013) Location recommendation in location-based social networks using user check-in data. In: 21st ACM SIGSPATIAL international conference on advances in geographic information systems—SIGSPATIAL'13, ACM, Orlando, Florida, pp 374–383. <https://doi.org/10.1145/2525314.2525357>
- Wörndl W, Hefe A, Herzog D (2017) Recommending a sequence of interesting places for tourist trips. *Inf Technol Tourism* 17(1):31–54. <https://doi.org/10.1007/s40558-017-0076-5>
- Wu D, Mamoulis N, Shi J (2015) Clustering in geo-social networks. *Bull IEEE Comput Soc Tech Committee Data Eng* 38:47–57
- Van der Zee E, Bertocchi D (2018) Finding patterns in urban tourist behaviour: a social network analysis approach based on tripadvisor reviews. *Inf Technol Tourism* 20(1–4):153–180. <https://doi.org/10.1007/s40558-018-0128-5>
- Zheng Y, Zhou X (2011) *Computing with spatial trajectories*. Springer, New York. <https://doi.org/10.1007/978-1-4614-1629-6>
- Zhou T, Ren J, Mcv Medo, Zhang YC (2007) Bipartite network projection and personal recommendation. *Phys Rev E* 76:046115. <https://doi.org/10.1103/PhysRevE.76.046115>
- Zhou T, Lü L, Zhang YC (2009) Predicting missing links via local information. *Eur Phys J B* 71(4):623–630. <https://doi.org/10.1140/epjb/e2009-00335-8>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Giorgos Stamatelatos^{1,2}  · George Drosatos¹ · Sotirios Gyftopoulos^{1,2} · Helen Briola^{1,2} · Pavlos S. Efraimidis^{1,2}

George Drosatos
gdrosato@athenarc.gr

Sotirios Gyftopoulos
sgyftopo@athenarc.gr; sgyftopo@ee.duth.gr

Helen Briola
elenbrio@athenarc.gr; ebriola@ee.duth.gr

Pavlos S. Efraimidis
pefraimi@athenarc.gr; pefraimi@ee.duth.gr

¹ Institute for Language and Speech Processing, Athena Research Center, Xanthi, Greece

² Department of Electrical and Computer Engineering, Democritus University of Thrace, Xanthi, Greece