Data & Knowledge Engineering 92 (2014) 20–38

Contents lists available at ScienceDirect

Data & Knowledge Engineering

journal homepage: www.elsevier.com/locate/datak

Editorial

Spatial-aware interest group queries in location-based social networks

Yafei Li [a,b], Dingming Wu [c], Jianliang Xu [a,*], Byron Choi [a], Weifeng Su [b]

[a] Hong Kong Baptist University, Hong Kong
[b] BNU-HKBU United International College, China
[c] Hong Kong University, Hong Kong

ARTICLE INFO

Article history:
Received 25 November 2013
Received in revised form 28 May 2014
Accepted 4 June 2014
Available online 12 June 2014

Keywords:
Spatial database
Query processing
Group queries
Location-based service

ABSTRACT

With the recent advances in positioning and smartphone technologies, a number of social networks such as Twitter, Foursquare and Facebook are acquiring the dimension of location, thus bridging the gap between the physical world and online social networking services. Most of the location-based social networks released check-in services that allow users to share their visiting locations with their friends. In this paper, users' interests are modeled by check-in actions. We propose a new type of *Spatial-aware Interest Group* (SIG) query that retrieves a user group of size k where each user is interested in the query keywords and they are close to each other in the Euclidean space. We prove that the SIG query problem is NP-complete. A family of efficient algorithms based on the IR-tree is thus proposed for the processing of SIG queries. Experiments on two real datasets show that our proposed algorithms achieve orders of magnitude improvement over the baseline algorithm.

1. Introduction

Smartphone and social networking have been identified as the two most important innovations that have emerged in the past ten years [1]. The convergence of these two technologies has given rise to a new line of applications, namely *location-based social networks* that bridge the gap between the physical world and online social networking services. Existing works in the literature have considered group queries in location-based social networks. [2,3] aim at finding a group of attendees close to a rally point and ensure that the selected attendees have a good social relationship to create a good atmosphere in the activity. [4] aims to find the activity time and attendees with the minimum total social distance to the initiator. [5,6] explore a group of experts whose skills can cover all the requirements and the communication cost among group members is low.

Most of the location-based social networks released check-in services that allow users to share their visiting locations with their friends. These locations, considered as spatial objects, are usually associated with a few tags that describe the features of those locations, e.g., spatial object 'starbucks' with tags 'food', 'beverage', and 'coffee'. If a user checks in the spatial object 'starbucks', the user may be interested in 'food', 'beverage', or 'coffee'. These voluntary check-in actions reflecting users' interests can benefit many applications. Utilizing such information, this paper proposes a new type of *Spatial-aware Interest Group* (SIG) query that retrieves a user group of size k where each user is interested in the query

* Corresponding author.
E-mail address: xujl@comp.hkbu.edu.hk (J. Xu).
URL: http://comp.hkbu.edu.hk/~xujl (J. Xu).

http://dx.doi.org/10.1016/j.datak.2014.06.001
0169-023X/

keywords and the users are close to each other in the Euclidean space. Different from existing work, the SIG query retrieves a user group of size k that maximizes a ranking function combining the diameter of the group (i.e., the distance between the farthest pair of users) and group's interest in the query keywords. Such queries are useful in many scenarios. For example, consider that a company wants to hold promotion campaigns in some regions. The company is interested in identifying the regions containing potential customers who are interested in the features (query keywords) of the product promoted. Another example is for interest-based group gathering. Query keyword 'movie' may find a group of nearby people who are movie lovers, while query keyword 'NBA' could retrieve a group of nearby people who like playing basketballs. Note that the group size in these queries is usually constrained due to limited venue capacity and/or financial budget.

Fig. 1 shows an example 3-size SIG query over several users shown as black dots. Consider three user groups G_1, G_2 and G_3, where $I(G_i, t)$ means group G_i's interest in query keyword t. Bigger value indicates higher interest. Comparing group G_1 with group G_2, group G_1 is more preferable, since the interest of G_1 on query keyword t is bigger than that of G_2, although G_1 and G_2 have the same layout in geo-space. Considering group G_1 with group G_3, group G_3 is more preferable, since the inter-user distances in G_3 is smaller than that of G_1, although G_1 and G_3 have the same interest value on query keyword t. Overall, group G_3 is the most preferable among these three groups.

To process SIG queries, a naive method is to enumerate all possible groups of size k, and then find the group with the highest ranking score. Obviously, this method is inefficient given a large number of possible groups. To avoid enumerating all possible groups, in this paper, we propose two efficient algorithms, namely Interest Oriented Algorithm (IOAIR) and Diameter Oriented Algorithm (DOAIR), based on the IR-tree [7] for the processing of SIG queries. The basic idea is to derive an upper bound on the ranking score for each possible user group and construct user groups in an iterative fashion; if the ranking score of the current found group is higher than the ranking upper bounds of the unseen groups, the query processing can be stopped by returning the current found group as the result. The two proposed algorithms adopt different orders in the construction of user groups: Algorithm IOAIR explores the search space by group interest while algorithm DOAIR proceeds by group diameter. We conduct extensive experiments on two real datasets to evaluate the performance of the proposed algorithms. The experiment results show that our algorithms achieve orders of magnitude improvement over the baseline algorithm.

Our contributions made in this paper can be summarized as follows:

1. Proposing a new type of query SIG that finds a k-size maximum interest group in location-based social networks and proving that the SIG query problem is NP-complete.
2. Developing two efficient algorithms IOAIR and DOAIR based on the IR-tree for the processing of SIG queries.
3. Validating the performance efficiency of the proposed query processing algorithms by empirical evaluation.

The rest of the paper is organized as follows. Section 2 presents the problem definition. Section 3 presents two efficient algorithms based on IR-tree for the processing of SIG queries. Section 4 shows the empirical study of our proposed algorithms on two real datasets. Section 5 reviews recent work on related queries in location-based social networks. Section 6 concludes the paper and outlines possible future work.

2. Problem definition

Let $\mathcal{D}$ be a set of spatial objects. Each spatial object p is associated with a set of tags $p.\Gamma$. Let $\mathcal{U}$ be a set of users. Each user $u \in \mathcal{U}$ is a triple (id, λ, v), where id is user's identifier, λ is user's location, and v is a vector of user's interests for the tags that are associated with

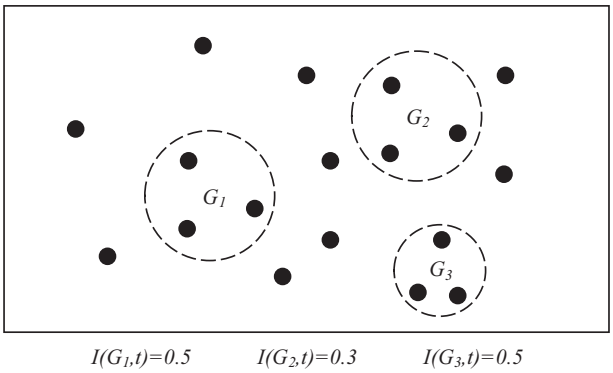

Fig. 1. A general case of 3-size SIG query.