Pervasive and Mobile Computing 11 (2014) 88–105

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

Fast track article

Leveraging Bluetooth co-location traces in group discovery algorithms

Daniel Boston [a,*], Steve Mardenfeld [a], Juan (Susan) Pan [a], Quentin Jones [b], Adriana Iamnitchi [c], Cristian Borcea [a]

[a] Computer Science Department, New Jersey Institute of Technology, Newark, NJ, USA
[b] Information Systems Department, New Jersey Institute of Technology, Newark, NJ, USA
[c] Department of Computer Science and Engineering, University of South Florida, Tampa, FL, USA

ARTICLE INFO

Article history:
Available online 6 November 2012

Keywords:
Mobile social computing
Group discovery
Co-location traces
Smart phones
Distributed computing

ABSTRACT

Smart phones can collect and share Bluetooth co-location traces to identify ad hoc or semi-permanent social groups, which can enhance recommender systems or allow detection of epidemic events. Group discovery using Bluetooth co-location is practical due to low power consumption, short range, and applicability to decentralization. This paper presents the Group Discovery using Co-location traces (GDC) and Decentralized GDC (DGDC) algorithms, which leverage user meeting frequency and duration to accurately detect groups. GDC and DGDC are validated on one month of data collected from 141 smart phones carried by students on our campus, and by comparison against ground truth groups.

1. Introduction

The incorporation of social information into online and mobile applications or services has become mainstream in the past several years. However, collecting social networking information is generally application dependent and limited to online activities. As such, a large amount of social information that results from face-to-face interactions is not captured. This is especially true for informal groups, which do not advertise their meetings online. Examples of such groups include a student study group, faculty who routinely have lunch together, or coworkers who sometimes play golf.

Capturing information about social groups formed through face-to-face interactions could be used in several types of services. Recommender services can benefit from a significant amount of additional social information to provide geo-social group recommendations such as meeting places or events of interest based on common user preferences [1]. Systems such as Mobius [2] can enforce socially-aware access control: multimedia content generated at a group meeting is only shared with group members, including those who were not present. The information about groups could also be leveraged in systems such as RoamWare [3] for mobile computer supported cooperative work. Forwarding in delay-tolerant ad hoc networks can be made socially aware by selecting a next hop device that belongs to a person who shares a group with the destination; in this way, the next hop has a higher probability to meet the destination [4]. Finally, disease epidemic events can be monitored [5] and potentially thwarted using social information, by alerting at-risk individuals before they come into contact with a source of infection. As well, the same information can be used to accelerate the delivery of treatment to those already suffering. Beyond biological epidemics, the spread of mobile computer viruses could be monitored, contained, and reversed with the effective use of social group data.

One way to identify this type of social group is to leverage information collected automatically from smart phones carried by mobile users, such as location or co-location. Using this information, a group can be defined as a relatively small set of

* Corresponding author. Tel.: +1 8566307796.
E-mail addresses: djb38@njit.edu (D. Boston), sm424@njit.edu (S. Mardenfeld), jp238@njit.edu (J. Pan), qjones@njit.edu (Q. Jones), anda@cse.usf.edu (A. Iamnitchi), borcea@cs.njit.edu (C. Borcea).

1574-1192/$ – see front matter
doi:10.1016/j.pmcj.2012.10.003

users who spend a significant amount of time together and meet for a significant number of times [6].[1] Discovering such groups, however, is difficult: (i) group members do not necessarily attend all group meetings, (ii) guests or people who pass by the meeting location can appear to be part of groups, (iii) group members spend different amounts of time at meetings, (iv) the collected data is incomplete due to sampling frequency and mobility, and (v) users may collect different data for the same group meeting.

Our GPI algorithm [7] used location to identify, with high accuracy and few false positives, groups and their associated places. However, there are several issues which prompted the need for a different algorithm. First, GPI requires a localization system on every mobile device, which is not always available (especially indoors). Second, many users are often reluctant to share location traces for any length of time due to privacy concerns. Even anonymous location traces are vulnerable to user identification through data mining techniques, which can lead to user tracking and home identification [8]. Third, localization systems running on phones (e.g., GPS, Place Lab [9]) consume a significant amount of battery power. Finally, the accuracy of localization systems can vary significantly across different places.

This paper presents the Group Discovery using Co-location traces (GDC) and Decentralized GDC (DGDC) algorithms, which use Bluetooth co-location traces and policy-based data sharing to identify groups. A co-location trace for a user is a set of records of other users who are within a certain proximity at the same time. GDC leverages the Bluetooth discovery protocol to collect these traces. GDC is practical and achieves good efficiency for several reasons. Unlike localization systems, Bluetooth is available on nearly all mobile phones and can work indoors. Although some mobile development platforms restrict the use of Bluetooth either by limiting the discovery mode or by requiring regular user feedback, solutions do exist such as in the Android 4.0 API [10], where developers can request unlimited Bluetooth discovery and educated users could consent. Sharing co-location information, instead of absolute location, may be a participation incentive for those users with concerns about their location being tracked. DGDC builds on this incentive by further reducing sharing to those other users with whom you already interact, while still providing reasonable group detection. Bluetooth consumes much less power than GPS and WiFi, while helping with the accuracy of the algorithm due to its short transmission range (i.e., 10 m). Finally, while not guaranteeing face-to-face interactions, proximity-based interactions captured by Bluetooth combined with GDC's rejection of social tie transitivity leads to social interaction data similar to face-to-face encounters.

DGDC has some important benefits over GDC. As it is entirely decentralized, privacy concerns are specifically addressed. Users can control the extent of their sharing through simple (or complex) policies reflecting their willingness to share information. DGDC data collection could also be selectively disabled to prevent undesirable sharing, such as on a daily commute or while riding on public transportation. There are some potential drawbacks, as communication costs can be greater than with GDC, and some discovered groups may be incomplete due to perspective isolation. As our evaluation will show, DGDC achieves reasonable results over a variety of policy choices.

Well-known graph algorithms, such as K-Clique [11] and WNA [12], could be employed to detect groups based on co-location. They work by inserting an edge in the graph between any two users who have been around each other for a certain amount of time. However, since these algorithms use only pair-wise information, there is no guarantee that their detected groups spent any time together. Furthermore, important parameters that can be used to classify groups, such as group meeting frequency and total group meeting time, are lost. Finally, K-Clique does not allow for weighted edges, which leads to lower group detection accuracy, and WNA does not work for overlapping groups, thus missing many groups.

This paper makes the following contributions:

- It presents GDC, a practical algorithm that leverages Bluetooth co-location traces to accurately identify social groups. GDC preserves the group meeting times and aggregate meeting frequency with each group, which can be used to compare, categorize, and rank groups.
- It describes DGDC, a distributed version of GDC involving data sharing only among other people who have actually been met, while still achieving reasonably similar group detection characteristics.
- It validates GDC with one month of data collected from 141 smart phones carried by students on our campus. Additionally, GDC was tested on the Reality Mining dataset [13].
- It compares GDC against K-Clique on our dataset by asking the users to participate in a survey deployed on Facebook. Using a Likert scale, we observed that GDC performed well: GDC-discovered groups received 30% higher ratings than K-Clique's groups.
- It evaluates GDC and DGDC against a careful reconstruction of "true groups" created using a novel tool (TIE [14]) and deep investigation of the source data, a reconstruction named the ground truth data set (GTS). We show that GDC and DGDC outperform K-Clique in both accuracy and terseness.

The rest of this paper is organized as follows. Section 2 describes the GDC algorithm. Section 3 presents DGDC in detail, together with its policies and benefits. Section 4 gives a three-part analysis of GDC, DGDC, and similar algorithms through a user study, a relative comparison, and comparison against ground truth. Section 5 discusses the related work, and Section 6 concludes the paper.

[1] These solutions discover co-located people. Additional sensing (e.g. voice recognition) or user confirmation is required to determine if the co-located users did indeed interact. An alternative solution for detecting face-to-face interaction is the use of RFIDs [5], but this solution requires people to carry an artifact (e.g., badge) with an embedded RFID; furthermore, it does not scale well as it requires RFID readers everywhere.