Measuring Interactivity and Geographical Closeness of Online Social Network Users to Support Social Recommendation Systems

Guilherme Sperb Machado, Thomas Bocek, Alexander Filitz, Burkhard Stiller
Department of Informatics IFI, Communication Systems Group CSG, University of Zurich
Binzmühlestrasse 14, CH—8050 Zurich, Switzerland
[machado|bocek|stiller]@ifi.uzh.ch, alexander.filitz@uzh.ch

***Abstract*—Several applications (**e.g.**, Instagram, PiCsMu) integrate existing Online Social Networks (OSN) into the core of their solutions to explore social information. Although this integration enables more accurate social recommendation systems, the collection and monitoring of relevant OSN data by third-party applications is a challenging management task, since OSNs (a) impose rate restrictions to their Application Programming Interface (API) calls, (b) do not provide detailed information about specific OSN features, and (c) may provide incomplete or not up-to-date OSN data. Therefore, this paper covers the design, prototyping, and evaluation of JSocialLib, a new meta-API library for collecting OSN data from existing OSNs. It provides (1) an interaction- and (2) a location-based method in support of social recommendations systems.**

***Index Terms*—Online Social Networks, Social Data, Online Social Model, Monitoring, Recommendation System**

I. Introduction

Through the adoption of Online Social Networks (OSN) user-generated content is consumed with a high rate. OSN users interact twice as much than 5 years ago, also spending a large amount of hours per day in OSNs [14]. Several applications and systems integrate existing OSNs to the core of their solutions. *E.g.*, PiCsMu [10] and Instagram integrate Facebook and Twitter, which enable the spreading of content to OSNs to build a collaborative system. Advantageous for integrating OSNs into an application is the collection of OSN-specific information for a Social Recommendation System (SRS). An application can use an SRS (1) to expand the user base, *i.e.*, attract more users to use the application to build its own social graph, and (2) to enhance the user's experience, *e.g.*, recommend to share content to friends. Thus, an SRS collects and monitors a variety of information within OSNs, *e.g.,* public/private posts and messages, locality tags, user's status updates, user's preferences, and content similarity.

This paper designs and evaluates JSocialLib, a new meta-API (Application Programming Interface) library to collect OSN data from existing OSNs to provide (1) an interaction- and (2) a location-based method for social recommendation systems. Measuring the interactivity between OSN users and their geographic closeness support SRSs to recommend new OSN friends based on these two metrics. JSocialLib is a meta-API, since it relies on other libraries underneath, RestFB and Twitter4J, to communicate with multiple existing OSNs.

The collection and monitoring of relevant OSN data by third-party applications are challenging tasks, since OSNs (a) impose rate restrictions to their API calls (*e.g.*, Facebook limits 600 calls per 10 minutes per access token), (b) do not provide detailed information about specific OSN features (*e.g.*, how many messages user A exchanged with user B using a mobile phone, in the last month), and (c) may provide incomplete or not up-to-date data [6]. After collecting OSN data, SRSs must perform a reasoning process, since it must deal with the lack of data and still present accurate recommendation results: the higher the accuracy is, the higher are the chances that recommendations are accepted by users. This subjective accuracy of recommendations is important to produce attractive recommendations for users, leading to the question: how accurate can JSocialLib's interaction- and location-based methods be compared to OSN users' perception?

In order to calibrate and evaluate these new JSocialLib interaction- and location-based methods, a study was carried out. The study is composed out of three parts: first, the data set is obtained by a Web-based survey, which indicates the friends that OSN users most interacts with and which are geographically closest; second, interaction- and location-based methods are calibrated for a random half of the data set collected; third, the calibrated social recommendation methods are evaluated against the remaining half of the data set.

The remainder of this paper outlines in Section II basic terminology and related work, which is followed by the JSocialLib architecture and design in Section III. The JSocialLib implementation is discussed and applied to perform the calibration and evaluation in Section IV. Finally Section V summarizes the work and addresses future work.

II. Terminology and Related Work

A *user* is defined as an entity representing a person or individual, who is authenticated and authorized to use a certain application or service. In this paper, the term *OSN user* is used when a *user* is uniquely associated to an OSN.

A *friend* is the counterpart of a positive social relationship, possibly involving trust, intimacy, or acquaintance. When referring to the term *friend*, it is solely in the scope of an OSN. A *friend* is also a *user*. Thus, the term *OSN friend* is used to refer to an *OSN user*, who is associated to another *OSN user* through an *OSN relationship*, which can vary depending on the concepts utilized within each OSN (*e.g.,* Facebook has bidirectional friendship relations, while Twitter has the concept of unidirectional friendship relations).

The general definition of *interaction* between two human beings encapsulates individuals' action toward another, *e.g.*, to talk, to look, or to smile. However, this work focuses on *OSN interactions* (*interactions* in short), which are interactions happening within an OSN triggered by OSN users. The definition of *OSN interaction* depends on intrinsic OSN fea-

978-3-901882-66-1 © 2014 IFIP

tures or concepts. *E.g.*, Facebook has the concept to "like" an item. Other OSNs, *e.g.*, Twitter, do not offer this concept.

JSocialLib can be compared to recommendation methods, as well as to the collection and analysis of OSN data approaches. [20] summarizes traditional recommendation methods into the following categories: content-based [2], collaborative filtering [5], clustering model [3], graph model [1], and association rule graph [12]. JSocialLib applies a modified collaborative filtering approach with implicit OSN data collection [5]. The interaction- and location-based methods do not actively interact with the OSN user in order to infer recommendations. Instead, it uses an implicit OSN data collection to measure OSN interactivity and geographical closeness.

Traditional recommendation methods only have an information input type, *e.g.*, "items bought by user A on Web site X" is the input for a recommendation method to provide further recommendations to user A. Social recommendations are a new approach in the field of recommendation methods [15], since they aim to specifically recommend OSN items and OSN friends to a given OSN user. In contrast to traditional recommendations, social recommendations have to deal with heterogeneous information of one or more sources. The heterogeneity is due to different elements that OSNs introduce, *e.g.*, users, pages, likes, posts, private messages, tags, locations, or complex relationships between all of these items [4].

TABLE I: COMPARISON OF RELATED WORK.

	Recommendation Categories	Interaction	Location	Evaluation OSN Data Set
[4]	genetic	-	-	public
[18]	modified collaborative filtering	-	-	public
[12]	content-based, collaborative filtering	✓	-	public
[13]	ranking	✓	-	public
[20]	random-walk, pair-wise learning	-	-	public
[4]	affinity matrices, interest familiarity	-	✓	public
[19]	random-walk, ranking	-	✓	public
JSocialLib	modified collaborative filtering with implicit data collection, social recommendation, ranking	✓	✓	public and private OSN user's data

[7] developed a method to crawl representative and unbiased samples of Facebook users and is used in most of the above mentioned related work. JSocialLib does not use [7], since the interaction- and location-based methods rely on OSN data inherent to OSN users's account and, therefore, do not rely on publicly accessible data that can be constantly crawled. Moreover, JSocialLib uses a combination of public and private OSN data information to conduct the calibration and evaluation of its methods. Therefore, the collection of private OSN user's data differs from other research works.

Different approaches to measure OSN interactivity and geographical closeness were discussed. [17] proposed the use of interaction graphs to quantify OSN user interactions. It analyzed interaction graphs derived from Facebook and showed that it has significantly lower levels of the "small-world" properties. [16] conducted a measurement study of the Flickr OSN and showed that only a small fraction of users in the main component of the friendship graph is responsible for the vast majority of user interactions. [9] analyzed the geographical location of LiveJournal users and showed a strong correlation between friendship and geographic proximity. [8] observed users on Twitter and analyzed geographical spread of Twitter usage in relation to user's behavior. Differently to what have already been developed, JSocialLib focuses on providing results that should be as close as possible to what OSN users perceive as the OSN friends that it most interact, and the OSN friends that are geographically closest.

III. SYSTEM AND METHOD DESIGN

JSocialLib relies on public APIs and can be integrated to existing applications' SRSs (cf. Section III-A). Based on a common model (see [11]), the two interaction- and location-based methods are described in Section III-B and C.

A. System Architecture

JSocialLib fits into an application architecture being integrated into one or more existing OSNs. The architecture figure is presented on [11]. The application (APP) has users interacting through a front end which, *e.g.*, can be a Graphical User Interface (GUI). Application users (APP users) might also be OSN users, since they have an account in the supported OSN. Due to the integration of the application and existing OSNs, APP users might associate their APP identity to their existing OSN identities. Therefore, JSocialLib can be used by applications' SRSs to measure OSN interactivity and geographical closeness of APP users that associated their OSN identities to this application.

The Social Core component is responsible to interact directly to OSNs for, *e.g.*, share application-specific content, and for authentication and authorization, and it provides for the association of APP and OSN users' identities. First, APP users specify on which OSN they have an OSN user account with. Second, the Social Core asks permission of APP users' to allow the APP to directly interact with the OSN. Most OSNs use OAuth for authentication and authorization of third-party applications [11]. The application calls the SRS, which analyzes OSN data from the OSN user to infer what and to whom recommend to. The JSocialLib supports SRSs in order to reduce the implementation complexity by providing information about OSN user's interaction and geographical closeness. The SRS component calls methods provided by the JSocialLib specifying the OSN user, the authorization and authentication token, and the time frame that interactivity and geographic closeness should be observed in.

B. Interaction-based Method

The interaction-based method measures the interaction between an OSN user and his/her OSN friends, in both directions and independently: from the OSN user to each of his/her OSN friends and from each of his/her OSN friends to the OSN user. Input parameters include: an OSN user (`u`), the OSN user's authentication token (`at`), and the observation time frame (`time`). The output is a list of OSN friends (`IM`, interaction-based method) which includes the list of most interacted OSN friends that the given OSN user has interacted with, during the specified time frame. The interaction-based method output `IM` is represented in a descending order, compiled by each interactivity score *IS* between `u` and his/her OSN friends (`f`). This method considers Public Post (PP) and Private Message (PM) as the OSN interactions. Each PP or PM can be one of the following types:

Public Post Sent (PPS): A Public Post (PP) sent is a public OSN interaction of an OSN user toward another OSN user, *e.g.*, OSN user A writing on OSN user B's profile. **Public Post Replied (PPR):** It is an extension of the PPS interaction, but in the opposite direction: *e.g.*, OSN user B publicly posting something to OSN user A, but relating to a previous PPS from OSN user A to B. **Private Message Sent (PMS):** A Private Message (PM) sent is a direct message sent from an OSN user to another OSN user, which can only be accessed by the two, privately. **Private Message Replied (PMR):** It is an extension of the PMS interaction, but in the opposite direction: *e.g.*, OSN user B sending a private message to OSN user A, but relating to a previous PMS from OSN user A to B. [11] presents why these OSN interactions were chosen.

The interaction-based method does not take the content of each OSN interaction into consideration, but only the quantity of interactions. *E.g.*, all PPS from OSN user A to OSN user B are accounted as well as PPR from B to A. Such approach preserves the OSN users' privacy. In order to compose the result (`IM`), it is required to define the Interactivity Score IS in both directions. Since PPs and PMs are considered as OSN interactions, the interactivity score is defined as the sum of PPs and PMs:

$$IS(u,f) = I_{pp} + I_{pm}$$

As mentioned, this method emphasizes a balance of OSN interactions. Thus, the ratio of PPS and PPR (R_{pp}) as well as PMS and PMR (R_{pm}) is calculated as follows, respectively:

$$R_{pp} = \frac{PPR}{PPS} \qquad R_{pm} = \frac{PMR}{PMS}$$

In addition, lower and upper bound ratio thresholds (TL and TU) are defined, respectively, where t is a threshold value that can be adjusted depending on how strict or tolerant the method should measure the balance of interactivity:

$$TL = \frac{1}{t} \qquad TU = t$$

Therefore, if R_{pp} and R_{pm} are within the lower and upper bound ratio threshold, OSN interactions have to be considered, otherwise they are discarded. The lower and upper bound ratio threshold allows the method to identify that, *e.g.*, an OSN user is sending messages, but the OSN friend is not answering or ignoring them. In some cases this may be abuse or spam, which is not interactivity acknowledged.

Let S_α define the sum of α, where α represents either PPS, PPR, PMS, or PMR. ω_α defines a weight for each S_α, which determines the degree of importance for each of these OSN interactions considered, *e.g.*, the higher the weight value on PPS, the higher importance of PPS to measure the final score of interactivity. Based on these definitions I_{pp} and I_{pm}, are defined, respectively, while the calibration of the values ω_α and t is performed in Section IV-C:

$$I_{pp} = \begin{cases} \omega_{pps}S_{pps} + \omega_{ppr}S_{ppr}, \text{ if } (TL \le (R_{pps} \text{ and } R_{ppr}) < TU) \\ \omega_{pps}S_{pps}, \text{ if } (TL \le R_{pps} < TU) \\ \omega_{ppr}S_{ppr}, \text{ if } (TL \le R_{ppr} < TU) \\ 0 \end{cases}$$

$$I_{pm} = \begin{cases} \omega_{pms}S_{pms} + \omega_{pmr}S_{pmr}, \text{ if } (TL \le (R_{pms} \text{ and } R_{pmr}) < TU) \\ \omega_{pms}S_{pms}, \text{ if } (TL \le R_{pms} < TU) \\ \omega_{pmr}S_{pmr}, \text{ if } (TL \le R_{pmr} < TU) \\ 0 \end{cases}$$

C. Location-based Method

This method measures OSN friends that are geographically closest to a given OSN user, being based on OSN location data found either in OSN user attributes, *i.e.*, hometown and current location, or embedded to OSN content, *i.e.*, pictures, public posts, and private messages with location tags. Location tags represent additional information in an OSN content to express locations through GPS coordinates. Although there are other ways to obtain location data, such as read content of private messages and public posts to parse for location keywords, this method considers location data in OSN user attributes without parsing and interpreting content.

Input parameters include an OSN user (`u`), the OSN user's authentication token (`at`), and the observation time frame (`time`). The output is a list of OSN friends (`LM`, location-based method) geographically closest to the given OSN user during the time frame specified. The output `LM` is represented in a descending order and compiles each location score LS between `u` and his/her OSN friends (`f`). The formula to calculate LS is presented in [11].

Location tags associated to OSN content can considerably vary depending on where OSN users are located, *e.g.*, OSN users can post pictures or send messages while travelling. This is not the case for OSN user attributes since, *e.g.*, current location or hometown tend to not be changed very often. In order to take such behaviour into account, D represents the time difference in seconds of when the OSN content pair was created. Thus, the longer the difference, the less the specific close match counts for the final location score. As an example to illustrate how LS is calculated for OSN user A, assume that OSN user A has friends B and C. Assume that locations *X* and *Y* are already identified as a close match. OSN user A was at location *X* on day *d*. OSN user B was at location *X* on *d*-10 and again on d-3, as well as at location *Y* at *d*-2. OSN user C was at location *X* at *d*+1. In this example, the method accounts a higher location score between A and C. The reason is that OSN user A and C have a close match with only one day of difference, while OSN user A and B have a close match with 10, 3, and 2 days of difference.

IV. Implementation and Evaluation

A Java-based prototypical implementation of JSocialLib was developed. It was built as a JAR (Java Archive) package exposing a set of Java methods to application's SRSs and its two methods have been evaluated.

JSocialLib defines a *OSNProvider* class, which is the interface to specific OSN provider's implementation. JSocialLib has three concrete OSN provider implementations: *FacebookOSNProvider*, *TwitterOSNProvider*, and *GooglePlusOSN Provider* classes. Each of these concrete OSN providers uses third-party libraries to communicate to the respective OSN: RestFB, Twiter4J, and Google+ Java API client. The *OSNProvider* implementation exposes a set of methods that are used by the JSocialLib interaction- and location-based methods for a given OSN user. *E.g.*, *OSNProvider* methods are *countPrivateMsgs(Date d1, Date d2)* that counts private messages between a time interval, *getContentBetweenDate(Date d1, Date d2)* that fetches all OSN content found between a time interval, *getCurrentLocationAttribute()* that returns the current location OSN attribute, *getRelations()* which retrieves all the OSN relationships, amongst others. Moreover, JSocialLib defines the *JSocialLibAPI* class, which is the concrete implementation exposing Java methods for the interaction-based and ocation-based method.

A challenge faced by JSocialLib is to find when to fetch OSN information for analysis. Thus, the limitations of API requests as well as other application restrictions imposed by OSNs [11] had to be taken into consideration. *E.g.*, Facebook and Twitter limit API calls to a certain rate per OSN user, and also per OSN application. *E.g.*, Facebook has the limit of 600 calls per 10 minutes per OSN user authentication token. To solve this problem, JSocialLib applies a waiting time that increases by 5 minutes every time that an error is returned.

Another challenge faced is the amount and accuracy of data returned by OSN API calls. Facebook, *e.g.*, limits to obtain some of its OSN users data (*e.g.*, stream public messages) only retrospectively to 30 days or maximally 50 posts, whichever is greater. Moreover, the data returned by Facebook may be not consistent [6]. To overcome this issue, JSocialLib applies weights within the developed methods.

A. Evaluation Method

The evaluation procedure consists of 4 main steps: (1) data collection, (2) hypothesis compilation, (3) methods' calibration, and (4) evaluation results. Step 1 collects the data of real OSN users from different nationalities, which was achieved through a Web survey. This OSN data was randomly divided into two equal portions: while the first portion is used in the calibration (Step 3), the second portion is applied to execute the calibrated methods. Step 2 compiles the hypothesis based on experimental and exploratory procedures (see [11]). The first portion of the collected data was manually explored to gain insights of what average results for JSocialLib methods' evaluations could be. JSocialLib methods are calibrated in Step 3 to determine the best set of parameters (weights and thresholds). Large parameter sets are generated and statistics computed for each of the two JSocialLib methods. The calibration's goal is to perform as close as possible with OSN users' perceptions. Step 4 uses calibrated JSocialLib methods and executes them on the second portion of the OSN data collected.

TABLE II: SURVEY QUESTIONS.

Question Group	Question Number	Question
Interaction-based	1.1	How often do you interact with Facebook friends?
	1.2	Who are the top 5 Facebook friends you most interacted with during the past 6 months?
	1.3	How do you rank those selected Facebook friends based on the amount of interaction in a descending order?
Location-based	2.1	How often do you specify or attach your current location when posting on Facebook?
	2.2	Who are the top 5 Facebook friends you were the geographically closest to during the past 6 months?
	2.3	How do you rank those selected Facebook friends based on geographical closeness in a descending order?

B. Data Collection

The OSN data collection was performed via a survey for Facebook users (http://social.csg.uzh.ch), since Facebook has the largest amount of monthly active users for an evaluation.

Within the first step of this survey the OSN authentication and authorization was performed, where each participant authenticates its Facebook identity and accepts permissions for the survey Facebook application. The result of this step is the authentication and authorization token (OAuth token) stored in the MySQL database. The second step followed by answering questions posed. The two main questions with three sub questions for both of them are shown in Table II.

The first three questions (1.1 to 1.3) are related to the interaction-based method and the last three questions (2.1 to 2.3) are related to the location-based method. Question 1.1 and 2.1 measure the frequency which the participant interacts with or uses locations' tags on Facebook. The fixed answer set allows for either *frequently*, *occasionally*, or *never* as possible answers. Answers for questions 1.2 and 2.2 are given by selecting Facebook friends, with an interface to enable fast searches. Question 1.2 captures the OSN user perception regarding the user's interactivity, while question 2.2 captures the user perception regarding the geographical closeness of friends. In question 1.3 and 2.3 the participant has to rank those friends selected in a descending order; the first position defines the friend that the participant most interacted with or was geographically closest to, respectively. The outcome of questions 1.2, 1.3, 2.2, and 2.3, is a list of 5 OSN friends for each participant of the survey, ranked in a descending order.

In total 290 participants completed the survey, collected during 30 days. The demographics related to the gender and age of all survey participants is found in [11]. Moreover, the survey Web site had visits from 20 countries and survey answers are from 9 countries.

C. Methods' Calibration

Each JSocialLib method depends on a set of parameters, such as weights (ω_n), distance thresholds (c_{attr} and c_{cont}), and interactivity ratio threshold (t). The calibration of the methods has the goal to find the best values for these parameters. The idea is to test as many parameter sets for each method as possible. [11] presents, in detail, how the parameter sets for each method were generated.

In order to compare the JSocialLib methods' results applying the values of each generated parameter set, two metrics are defined: *rank difference* and *match count*. The metrics compare the OSN user perception (answered in the survey) to the JSocialLib method results. The metric values are calculated based on the parameter sets generated.

Rank Difference calculates the difference in ranking positions between the ranked list of 5 friends answered in the survey, to the JSocialLib methods' result for each generated parameter set. This is done for all parameters sets and the mean of total rank differences is computed (*rank difference mean*). [11] shows an example of such computation.

Match Count counts the matches of rank positions. If the rank position is the same for the OSN user answer and in the JSocialLib result, it is counted as an *exact match count*. If an OSN friend is ranked within any of the first 5 rank positions in the OSN user survey answer and in the JSocialLib method result, it is counted as an *top5 match count* (see [11]).

Based on these values, a set of filtering methods is used to exclude biased and totally inaccurate results, in support of an accurate calibration. Filtering methods detect imbalances introduced by OSN errors or OSN user's lack of perception (due to a widely applied survey). Two different filtering methods have been determined: *failed completely* and *zero score*. *Failed completely* filters out parameter set results, when the JSocialLib cannot retrieve data related to that specific OSN user due to authentication or authorization restrictions. This is the case when the survey participant removed permissions for the Facebook application right after the survey was submitted and, consequently, the JSocialLib methods were not able to fetch any OSN data. *Zero score filtering* filters out OSN friends, who have a score value of exactly zero. If all OSN friends present a score value equaling to zero, the entire OSN user is filtered out. It is important to highlight that both *failed completely* and *zero score filtering* only filters out true negatives, not impacting evaluation results. More information about the employed filtering methods can be found in [11].

TABLE III: CALIBRATION RESULTS.

Method	Parameter and Value
Interaction-based	$\omega_{pps} = 3$, $\omega_{ppr} = 9$, $\omega_{pms} = 1$, $\omega_{pmr} = 1$, $t = 3$
Location-based	$\omega_{attr} = 3$, $\omega_{cont} = 7$, $c_{attr} = 9$, $c_{cont} = 7$

In order to determine the best parameter set for each method, all results are ordered by the *rank difference* value mean in a descending order. Therefore, iterating through each ordered parameter set, the one with the highest *top5 match count* mean value and with the lowest standard deviation is selected. With this approach, the best parameter set is selected, prioritizing the *top5 match count* mean followed by the standard deviation and *rank difference* mean. Table III shows the calibration outcome for both JSocialLib methods.

D. Results

The JSocialLib methods are properly calibrated with the best found parameters sets, and applied to the remaining half of the OSN data obtained through the Web survey. All metrics are calculated again for the second half of the OSN data and compared to the metric values from the calibration step.

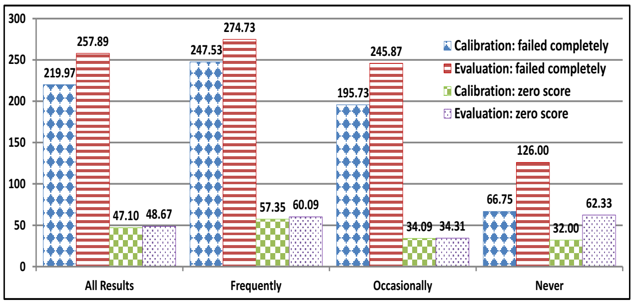

Fig. 1. Interaction-based method: Comparison of the *rank difference* mean for the filtering methods *failed completely* and *zero score*.

Interaction-based Method: Results based on *zero score filtering* considers 137 OSN users in the calibration and 139 OSN users in the evaluation (out of 145 OSN users each). Five OSN users were filtered out in both calibration and evaluation due to *failed completely filtering*. Due to *zero score filtering*, 1 and 3 OSN users were discarded, respectively, for calibration and evaluation. Moreover, on average, 1.66 and 1.88 OSN friends were removed due to *zero score filtering* from the calibration and evaluation results of each OSN user that answered the Web survey.

Fig. 1 shows the comparison of the interaction-based method for the *rank difference* metric with the *failed completely* and *zero score filtering*. These values represented by bars indicate the *rank difference* mean for all OSN users ("All Results") or for each frequency group. The lower the *rank difference* mean, the better the method performed. Best results were achieved for OSN users who answered "occasionally" as their interaction frequency. It is shown that the *rank difference* mean for the *zero score filtering* is basically the same for both calibration and evaluation data sets. For OSN users who selected "never" as an interaction frequency, evaluation results performed worse than for those calibration results. More result details about Fig. 1 are presented in [11]. Fig. 2 shows the comparison of the interaction-based method for the *top5 match count* metric. The *top5 match count* metric shows the same values for all filtering methods, because filtering does not affect this metric. Therefore, for simplicity reasons, it only shows results comparing the calibration with the evaluation. These values indicate the *top5 match count* mean and error bars show the standard deviation. The higher the *top5 match count* mean and the lower the standard deviation, the better the method performed. Best results were achieved for OSN users that answered "occasionally" as their interaction frequency. Also, for the *top5 match count* metric, the interaction-based method achieves very close results for the calibration and evaluation. These measurement results for OSN users that answered "never" considerably vary due to a low confidence in the mean.

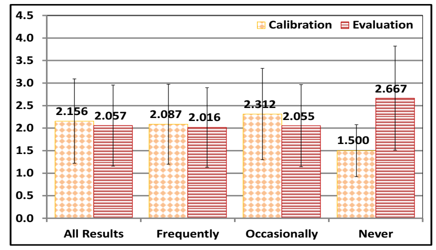

Fig. 2. Interaction-based method: comparison of the *top5 match count*.

Location-based Method: The results based on *zero score filtering* considers 92 OSN users in the calibration and 99 OSN users in the evaluation (out of 145 OSN users each). Due to *failed completely filtering*, 6 calibration and 5 evaluation OSN users were discarded. Respectively for the calibration and evaluation, 40 and 39 OSN users were filtered, since they did not have any kind of location data at all and, therefore, the location-based method is not able to produce meaningful results. Moreover, 7 and 2 OSN users, respectively for the calibration and evaluation, were filtered due to *zero score filtering*. In average, 1.99 and 2.27 OSN friends were removed, respectively, from the calibration and evaluation results, due to *zero score filtering*.

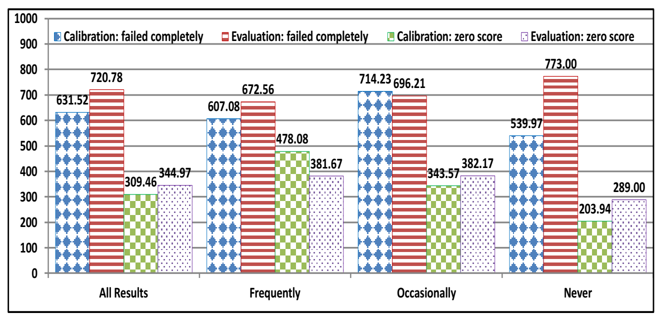

Fig. 3. Location-based method: comparison of the *rank difference* mean for the filtering methods *failed completely* and *zero score*.

Fig. 3 shows the comparison of the location-based method for the *rank difference* metric with the filtering methods *failed completely* and *zero score*. These values indicate the *rank difference* mean. For "all results", the evaluation's *rank difference* mean is higher than the calibration's *rank difference* mean. Taking into account that the *rank difference* scale is also higher compared to the interaction-based method (309.46/344.97 of location-based compared to 47.1/48.67 of interaction-based method), the evaluation of both location- and interaction-based methods had almost the same performance related to the results produced. Nevertheless, in terms of the *rank difference* metric, the location-based method performs worse when compared to the interaction-based method. An interesting observed aspect is the fact that for OSN users who answered that they "never" attach locations to their posts on Facebook the lowest *rank difference* mean is achieved — meaning that the method performs best for this particular case. In contrast to the interaction-based method, where only 4 OSN users answered with "never" as interaction frequency, 71 and 65 OSN users, respectively for the calibration and evaluation, answered that they never attach location data to their Facebook posts or messages. This result shows indicatives that the majority of OSN users and OSN friends could had been filtered out, *e.g.*, due to the lack of attached location data (*zero score filtering*). However, the location-based method found one or more attached location data for 35 and 37 OSN users and his/her friends, respectively for calibration and evaluation. These numbers show indications that the location-based method is aligned to the OSN user perception even with a minimum set of OSN data.

Fig. 4 shows the comparison of the location-based method for the *top5 match count* metric. These values indicate the *top5 match count* mean, and the error bars represent the corresponding standard deviation. Also for the *top5 match count* metric, the interaction-based method achieves similar results for both OSN data set portions (calibration and evaluation).

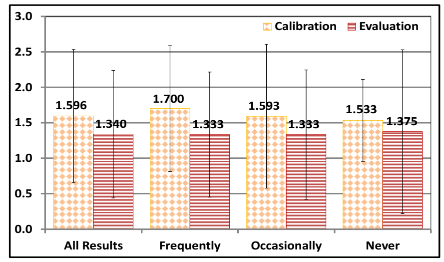

Fig. 4. Location-based method: comparison of *top5 match count* mean.

V. Summary and Conclusions

This work designed, implemented, and evaluated JSocialLib, a meta-API library providing interaction- and location-based methods in order to support SRSs and the management of social recommendation. OSN data was collected through a Web survey with real OSN users.

The calibrated interaction-based method estimates, on average, 2 out of 5 OSN friends that an OSN user also perceives as he/she interacts most with. The calibration presents that Facebook users which answered the survey perceive public interaction as more important than private interaction, *i.e.*, PPS (Public Posts Sent) have a weight value three times higher than PMS (Private Messages Sent), while PPR (Public Posts Received) have a weight value nine times higher than PMR (Private Messages Received). The calibrated location-based method estimates, on average, at least 1.3 out 5 OSN friends that an OSN user also perceives as the geographically closest to. Facebook users have more interaction OSN data than location OSN data, since OSN users tend to avoid the attachment of locations to OSN content mainly due to privacy issues. Therefore, the location-based method performs worse in both metrics (*rank difference*, *top5 match count*) due to the imprecision and lack of OSN data from OSN users. The calibration results show that OSN users consider a radius of 7 kilometers as where the geographically closest OSN friends are located, regarding users' current location.

This work concludes that OSN users either interact more with their OSN friends with public posts than those with private messages, or they likely perceive more OSN friends which interacted publicly than privately. Moreover, this work concludes that even with the OSN heterogeneity and OSN data collection challenges imposed, applications' SRS can benefit by exploring the social graph of third-party OSNs to perform recommendations aligned to OSN user's perception.

VI. Acknowledgements

This work was partially supported by the SmartenIT and the FLAMINGO projects, funded by the EU FP7 Program under Contract No. FP7-2012-ICT-317846 and No. FP7-2012-ICT-318488, respectively.

References

[1] C. C. Aggarwal, J. Wolf, K. Wu, P. Yu: *Horting Hatches an Egg: A New Graph-theoretic Approach to Collaborative Filtering*. 5th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD99), New York, USA, 1999, pp 201-212.

[2] M. Balabanovic, Y. Shoham: *Fab: Content-based, Collaborative Recommendation*. Communications of the ACM, 1997, pp 66-72.

[3] J. S. Breese, D. Heckerman, C. Kadie: *Empirical Analysis of Predictive Algorithms for Collaborative Filtering*. 14th Conference on Uncertainty in Artificial Intelligence (UAI98), San Francisco, USA, 1998, pp 43-52.

[4] C. Chu, W. Wu, C. Wang, T. Chen, J. Chen: *Friend Recommendation for Location-based Mobile Social Networks*. 7th Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2013, pp 365-370.

[5] M. Deshpande, G. Karypis: *Item-based Top-n Recommendation Algorithms*. ACM Transactions on Information Systems (TOIS), Vol. 22, No. 1, January 2004, pp 143-177.

[6] Facebook Developer Website: *Bug Report November 25, 2013*. https://developers.facebook.com/x/bugs/694034083948939, April 2014.

[7] M. Gjoka, M. Kurant, C. T. Butts, A. Markopoulou: *Walking in Facebook: A Case Study of Unbiased Sampling of OSNs*. IEEE INFOCOM, 2010, pp 1-9.

[8] B. Krishnamurthy, P. Gill, M. Arlitt: *A Few Chirps about Twitter.* ACM SIGCOMM WOSN, 2008.

[9] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, A. Tomkins: *Geographic Routing in Social Network*. PNAS, 2005.

[10] G. S. Machado, T. Bocek, M. Ammann, B. Stiller: *Cloud Storage Overlay to Aggregate Heterogeneous Cloud Services*. 38th IEEE Conference on Local Computer Networks (LCN), Sydney, Australia, Oct 2013, pp 597-605.

[11] G. S. Machado, T. Bocek, A. Filitz, B. Stiller: *Design, Prototyping, and Evaluation of Methods to Measure Interactivity and Geographical Closeness of Online Social Network Users in Support of Social Recommendation Systems*. IFI Technical Report, IFI 2014.3, May 2014.

[12] J. Naruchitparames, M. H. Gunes, S. J. Louis: *Friend Recommendations in Social Networks Using Genetic Algorithms and Network Topology.* IEEE Evolutionary Computation (CEC), 2011, pp 2207-2214.

[13] R. Nia, F. Erlandsson, H. Johnson, S. F. Wu: *Leveraging Social Interactions to Suggest Friends*. 33rd IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW), 2013.

[14] Pew Research Internet Project: *Social Network Fact Sheet, 2013*. Available at: pewinternet.org/fact-sheets/social-networking-fact-sheet/.

[15] V. Podobnik, D. Striga, A. Jandras, I. Lovrek: *How to Calculate Trust between Social Network Users?* 20th International Conference on Software, Telecommunications, Computer Networks (SoftCOM), Split, Croatia, 2012, pp 1-6.

[16] M. Valafar, R. Rejaie, W. Willinger: *Beyond Friendship Graphs: A Study of User Interactions in Flickr.* ACM SIGCOMM WOSN, 2009.

[17] C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy, B. Y. Zhao: *User Interactions in Social Networks and Their Implications*. ACM EuroSys, Nürnberg, Germany, 2009.

[18] C. Yang, J. Sun, Z. Zhao: *Personalized Recommendation Based on Collaborative Filtering in Social Networks*. IEEE Progress in Informatics and Computing (PIC), Vol. 1, 2010, pp 670-673.

[19] X. Yu, A. Pan, L. Tang, Z. Li, J. Han: *Geo-friends Recommendation in GPS-based Cyber-physical Social Networks*. Advances in Social Networks Analysis and Mining (ASONAM), 2011, pp 361-368.

[20] J. Zhang, J. Tang, B. Liang, Z. Yang, S. Wang, J. Zuo, J. Li: *Recommendation over a Heterogeneous Social Network*. 9th International Conference on Web-Age Information Management (WAIM08), Zhangjiajie, China, 2008, pp. 309-316.