ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/273348455

Participant Selection for Offline Event Marketing Leveraging Location-Based Social Networks

Article *in* IEEE Transactions on Systems, Man, and Cybernetics: Systems · June 2015

DOI: 10.1109/TSMC.2014.2383993

CITATIONS
15

READS
425

4 authors, including:

Zhiyong Yu
Fuzhou University

43 PUBLICATIONS **574** CITATIONS

SEE PROFILE

Daqing Zhang
Institut Mines-Télécom

303 PUBLICATIONS **9,913** CITATIONS

SEE PROFILE

Dingqi Yang
Université de Fribourg

45 PUBLICATIONS **946** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Mobile Crowd Sensing: Algorithms and Systems View project

Project crowdsensing View project

All content following this page was uploaded by Dingqi Yang on 03 August 2015.

The user has requested enhancement of the downloaded file.

Participant Selection for Offline Event Marketing Leveraging Location-Based Social Networks

Zhiyong Yu, *Member, IEEE*, Daqing Zhang, *Member, IEEE*, Zhiwen Yu, *Senior Member, IEEE*, and Dingqi Yang

***Abstract*—Offline event marketing invites people to participate in a sponsored gathering, thus allowing marketers to have face-to-face, direct, and close contact with their current and potential customers. This paper presents a framework that supports marketers in improving marketing effectiveness by carefully selecting invitees to such sponsored offline events by leveraging location-based social networks. In particular, we first transform the participant selection task into a combinatorial optimization problem. Second, we propose a marketing effect quantitative model that considers the distance and overlapping social influence. Third, we introduce algorithms to determine a participant team that can maximize the marketing effect while fulfilling the scale and item coverage constraints. We finally evaluate the effectiveness of the framework and validate the proposed marketing effect of the quantitative model with real-world data.**

***Index Terms*—Location-based social networks, offline event marketing, participant selection.**

I. INTRODUCTION

OFFLINE event marketing has become more popular among different brands [1], [24]. This paper typically invites people to participate in a sponsored gathering, which allows marketers to have face-to-face, direct, and close contact with their current and potential customers. Many business owners have realized that offline events can transform their current customers into loyal brand advocates and reach many potential customers. Tasting night is a typical offline event favored by many restaurants. A restaurant owner usually posts an event call in advance to the public (e.g., at the storefront, in local newspapers, on the restaurant's website, and/or on social networking sites). Social networks have become a cost-effective marketing channel with their recent surge. Take a tasting night advertisement through the social network Facebook[1] as an example. An event call usually takes the form shown in Fig. 1 with the following highlighted information.

Fig. 1. Example of a tasting night call on Facebook.

1) *Location:* The place where the event will be held, such as the address of the restaurant is shown in Fig. 1. The event is usually held at the business venue, where customers can consume products or services.
2) *Served Items:* Products that will be served in the event and their corresponding prices. Several product items are served free or at a discount to attract participants.
3) *Scale:* The number of participants allowed to join the event. The scale is sometimes not explicitly indicated, which means that the more participants, the better. However, only a limited number of participants can be entertained given the capacity of a venue. The scale and planned discount depend on the marketing budget.
4) *Others:* The event duration (i.e., when the event will start and end), a brief event description, or sometimes the policy in selecting participants.

An essential problem arises when the number of people willing to attend an event is much larger than planned. Who should be selected so that the marketing effect can be maximized under given conditions (i.e., location, served items, and scale)? Simple strategies such as random selection, first come first serve (FCFS), or "nearest people first" (NF) cannot maximize the marketing effect. They merely ensure that the

Manuscript received March 5, 2014; revised July 7, 2014 and September 11, 2014; accepted October 28, 2014. Date of publication January 13, 2015; date of current version May 13, 2015. This work was supported in part by the EU FP7 Project SOCIETIES under Grant 257493, in part by the National Basic Research Program of China under Grant 2015CB352400, in part by the National Natural Science Foundation of China under Grant 61300103, Grant 61222209, and Grant 61332005, and in part by the Natural Science Foundation of Fujian Province under Grant 2012J01264. This paper was recommended by Associate Editor F.-Y. Wang.

Z. Yu is with the Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing, College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China.

D. Zhang and D. Yang are with the CNRS UMR 5157 SAMOVAR, Institut Mines-Telecom, TELECOM SudParis, Evry 91011, France (e-mail: daqing.zhang@telecom-sudparis.eu).

Z. Yu is with the School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMC.2014.2383993

[1] http://www.facebook.com/events/185844228156829

2168-2216 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

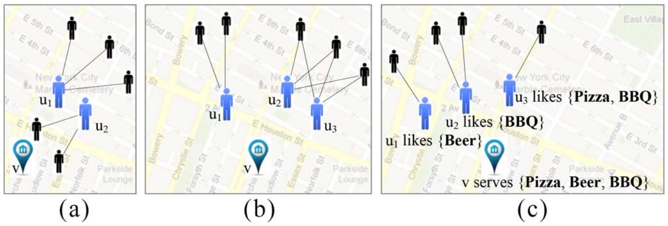


Fig. 2. Illustration of three important factors. (a) Distance. (b) Overlapping social influence. (c) Factor: item coverage.

number of participants does not exceed as planned. A better strategy is "most influential people first" (IF), which assumes that the influential participants can influence more people who visit the venue and become regular customers. The social influence of an individual can nowadays be roughly estimated by his/her number of friends and followers on social networking sites.

However, several important factors should be considered when applying social influence in participant selection for offline event marketing: 1) distance, i.e., the farther a customer from the venue, the lower probability he/she would visit the place [2]–[4]; 2) overlapping social influence, i.e., the probability that a customer influenced by multiple participants visits a venue should not be repeatedly counted; and 3) item coverage, i.e., each product item served by the venue should be favored by at least one participant. Three toy examples are shown in Fig. 2 to concretely illustrate the three factors. The locations of restaurant v and users u_j are indicated on the map. A link between any two users denotes their social relationship (e.g., friendship). Fig. 2(a) shows that u_1 has more friends than u_2, but u_1 is not necessarily prioritized over u_2 because u_2's friends are closer to v than u_1's. Fig. 2(b) shows that u_1 has two friends, whereas u_2 and u_3 have three common friends. Given the IF rule, the participant selection of $\{u_2, u_3\}$ should be better than $\{u_1, u_2\}$ because u_3 has more friends than u_1. However, u_1 influences a different user set than u_2, whereas u_3 influences the same group of users as u_2. Thus, the set $\{u_1, u_2\}$ can be a better choice than $\{u_2, u_3\}$. Fig. 2(c) shows that v mainly serves three types of food, namely, pizza, beer, and barbecue (BBQ). Given that u_1 likes beer, u_2 likes BBQ, and u_3 likes pizza and BBQ, who should be selected given the participant quota of two? Although $\{u_2, u_3\}$ can influence more users than $\{u_1, u_3\}$, they cover fewer food types. Optimizing participant selection by considering all these factors is clearly a nontrivial problem.

We need to specify quantitative information to consider these factors and uncover the effect of participants' social relationships, distances, scale, and item coverage on their marketing effect.

Recently, as smartphones, GPS, and Web 2.0 technologies have recently become ubiquitous, location-based social networking sites (LBSNs) such as Foursquare,[2] Facebook location,[3] and Google latitude[4] are becoming popular. LBSNs allow a user to "check in" at venues, which corresponds to an online record that describes his/her current physical location and shares this record with his/her friends. Many content types on LBSNs are generated, including users' social relationships, check-in history records, venue category, GPS coordinates, and tags. These digital footprints [5] contain rich quantitative information that characterize the social influence considered above. We can determine a person's social influence and its overlapping factor from the user's social relationships and check-in history records. We can model the distance factor on the basis of the location of each venue and users' check-in history records. We can estimate the covered items of selected users from the tags of each venue (some tags reveal the served items) and users' check-in history records. We can further extract more high-level knowledge such as users' "home location" and user preference on items (see Section II-B for detailed data preprocessing).

This paper presents a framework that assists marketers in improving marketing effectiveness by carefully selecting invitees to an offline event. The distance, overlapping social influence, and item coverage are considered by leveraging data collected from LBSNs. Our basic assumption is as follows. An LBSN is selected as a test platform to illustrate our participant selection idea and approaches. The event's location, served items, and scale are specified by marketers. The marketing effect is measured in terms of how many customers can be expected to visit the venue in a specified period of time. Given this assumption, our framework can determine if a set of participants can meet the planned scale and item coverage requirements, as well as estimate its marketing effect. Our framework can further search a set of participants on the basis of the said ability, whose marketing effect reaches near optimal value among all possible participant combinations.

The main contributions of this paper are as follows.

1) We convert the participant selection task in offline marketing into a combinatorial optimization problem. This new strategy attempts to maximize the marketing effect and fulfill multiple constraints, including the scale and item coverage, unlike traditional participant selection strategies such as random selection, FCFS, NF, and IF.
2) We propose a marketing effect quantitative model that considers the distance and overlapped social influence. This model is built and derived from the empirical facts obtained by the users' check-in history records on LBSNs. Thus, this model can predict the marketing effect by calculating how many customers can be expected given a set of participants.
3) We develop an algorithm that addresses the participant selection problem with the marketing effect quantitative model. Given that this combinatorial optimization problem is NP-hard, we obtain a near optimal solution by applying some metaheuristics.
4) We evaluate the effectiveness of the framework and validate the proposed marketing effect quantitative model with a dataset collected from Foursquare. The results show that the framework can effectively solve the participant selection issue. The model is also better than others that do not consider distance and/or overlapped social influence factors.

[2] https://foursquare.com

[3] https://www.facebook.com/about/location

[4] http://www.google.com/latitude

TABLE I
NOTATIONS

Symbol	***Description***
USER	universal user set
u_j	a customer, $u_j \in$ USER
N_{USER}	user number
VENUE	universal set of venues
v_k	a venue, $v_k \in$ VENUE
N_{VENUE}	venue number
TAG	universal set of items
t_i	an item, $t_i \in$ TAG
N_{TAG}	item number
N_{SCALE}	scale of the offline event
T_{v_k}	items of the venue v_k
T_{u_j}	items of the customer u_j
$d(u_j, v_k)$	distance between the customer u_j and venue v_k
$f(u_j, v_k)$	number of u_j's friends that have already visited the venue v_k
$c(u_j, v_k)$	whether the customer u_j has visited the venue v_k
X	a solution (i.e., a set of participants)

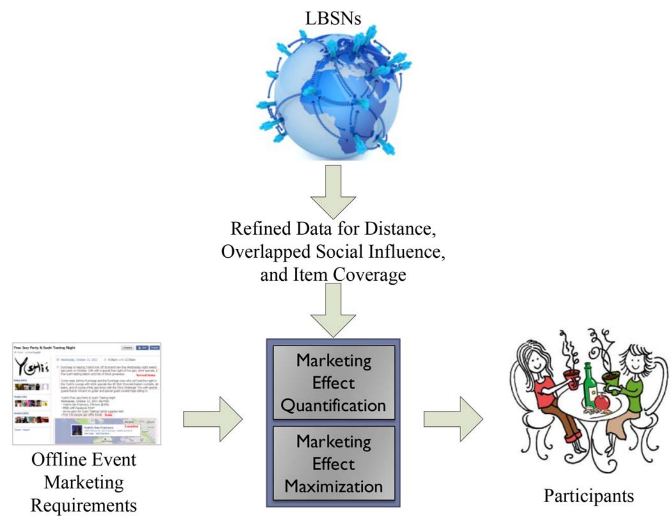


Fig. 3. Overview of the participant selection framework.

The rest of this paper is organized as follows. Section II introduces the overall framework. The marketing effect quantitative model is proposed and elaborated in Section III. Section IV presents algorithms for marketing effect maximization. The experimental results are described in Section V. Section VI discusses related work, and Section VII provides the conclusion.

II. FRAMEWORK

We first define notations used in this paper as shown in Table I. USER $= \{u_1, u_2, \ldots, u_{N_{\text{USER}}}\}$ is the customer set, and VENUE $= \{v_1, v_2, \ldots, v_{N_{\text{VENUE}}}\}$ is the venue set, where N_{USER} is the user number, and N_{VENUE} is the venue number. The scale is denoted as N_{SCALE}. A venue or a customer is tagged with a set of items. The item set is TAG $= \{t_1, t_2, \ldots, t_{N_{\text{TAG}}}\}$, where N_{TAG} is the item number. The venue items that are denoted as T_{v_k} indicate the product or service types the venue usually supplies. $T_{v_k} = <t_1(v_k), t_2(v_k), \ldots, t_{N_{\text{TAG}}}(v_k)>$, $t_i(v_k) = 1$ or 0, $k = 1, 2, \ldots, N_{\text{VENUE}}$, $i = 1, 2, \ldots, N_{\text{TAG}}$. Customer items that are represented as T_{u_j} depict the product or service types the user prefers. $T_{u_j} = < t_1(u_j), t_2(u_j), \ldots, t_{N_{\text{TAG}}}(u_j)>$, $t_i(u_j) = 1$ or 0, $j = 1, 2, \ldots, N_{\text{USER}}$. Let $d(u_j, v_k)$ denote the distance between the customer u_j and the venue v_k, whereas $f(u_j, v_k)$ denotes the number of u_j's friends who have already visited venue v_k. A visiting case can be expressed by a tuple of $<u_j, v_k, c(u_j, v_k), d(u_j, v_k), f(u_j, v_k)>$, denoting whether ($c(u_j, v_k) = \{1$ if visited, 0 otherwise$\}$) who (u_j) visited where (v_k) with their distance ($d(u_j, v_k)$ in km) and the number of visited friends ($f(u_j, v_k)$). For instance, $<u_1, v_1, 1, 1.2, 2>$ means customer u_1 has visited venue v_1, their distance is 1.2 km, and two friends of u_1 have also visited v_1. A solution (i.e., a set of participants) is $X = <x_1, x_2, \ldots, x_{N_{\text{USER}}}>$, where $x_j = 1$ if u_j should be invited; otherwise, $x_j = 0$. An available solution means it satisfies all constraints, and the optimal solution is the one that can result in the maximum marketing effect within all available solutions.

The framework is illustrated in Fig. 3. The system takes two types of inputs: 1) the user-generated contents from LBSNs and 2) the requirement specification from marketers. The system then explores the solution space through marketing effect quantification and maximization. Finally, the system outputs a near optimal solution with the number of participants who should be invited.

Marketing effect quantification is responsible for modeling and calculating the marketing effect of a given participant set. The marketing effect quantitative model is built before executing the participant selection function. We construct different models with or without considering the distance or the overlapping social influence for comparison. All models should learn their parameters by fitting the customer mobility facts recorded on LBSNs. We can calculate the marketing effect for any available solution after determining the model and parameters.

Marketing effect maximization is adopted to recommend a user set that should be invited as participants in an offline event. The determined participant set should result in a larger marketing effect than any other participant sets while it satisfies the scale and item coverage constraints. Although we can obtain the marketing effect of any available solution, we cannot rank all available solutions because of the combinatorial explosion. We use linear programming or metaheuristic algorithms to search the (near) optimal solution in the solution space according to the complexity of models.

These two modules will be described in detail in Sections III and IV, respectively.

A. Marketing Requirement Inputting

Offline event marketing requirements indicate what result is expected with the given constraints. The requirements are specified by the venue owners or marketers. Given that the requirements cannot be expressed in a formal way, our framework offers an interface for inputting requirements and representing them formally.

A common setting is as follows: a marketer should invite a participant set that can maximize the marketing effect (i.e., the expected number of customers who will visit the venue after

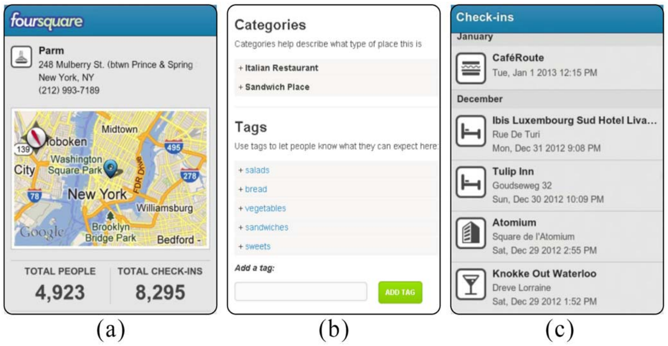


Fig. 4. Raw data examples from Foursquare. (a) Venue location. (b) Venue categories and tags. (c) User check-ins.

this marketing event) while satisfying the scale constraint and the item coverage constraint. These constraints are as follows.

1) The scale constraint is the participant number that equals the planned scale

$$\sum_{j=1}^{N_{\text{USER}}} x_j = N_{\text{SCALE}}. \tag{1}$$

2) The item coverage constraint is the tags of the participants who can cover all venue tags

$$\forall_{i=1}^{N_{\text{TAG}}} t_i, \sum_{j=1}^{N_{\text{USER}}} x_j \times t_i\left(u_j\right) \geq t_i(v_k). \tag{2}$$

3) Let $E(X)$ be the marketing effect of a solution X. Our target then is to maximize this value

$$\arg\max_X E(X). \tag{3}$$

B. LBSN Data Preprocessing

Raw data collected from LBSNs are usually unstructured and cannot be used by a particular application directly. Data preprocessing cleans and transforms raw data to fit the application requirements. The refined data should reflect the factors of distance, overlapping social influence, and item coverage in this paper.

Besides social relationships, raw data on LBSNs include the following (examples from Foursquare are shown in Fig. 4).

1) The categories and location of each venue are added by the user who created it. Fig. 4(a) shows the location of a venue in New York called Parm. The top part of Fig. 4(b) depicts its categories, namely, Italian Restaurant and Sandwich Place.
2) Venue tags added by users inform people of what items are served there. The bottom part of Fig. 4(b) shows that Parm serves salads, bread, vegetables, sandwiches, and sweets.
3) Users' check-in history records. Fig. 4(c) presents a partial record of a user who has visited Knokke Out Waterloo, Atomium, Tulip Inn, Ibis Luxembourg Sud Hotel, and CafeRoute.

Fig. 5 shows the sources and results of data preprocessing. Each venue belongs to a category (e.g., food, nightlife, shop, entertainment). $(long, lat)_{v_k}$ refers to v_k's longitude and latitude. A check-in record, $(u_j \rightarrow v_k)$, indicates that u_j checked v_k at a certain time. $(u_j \rightarrow u_{j'})$ means that u_j follows $u_{j'}$.

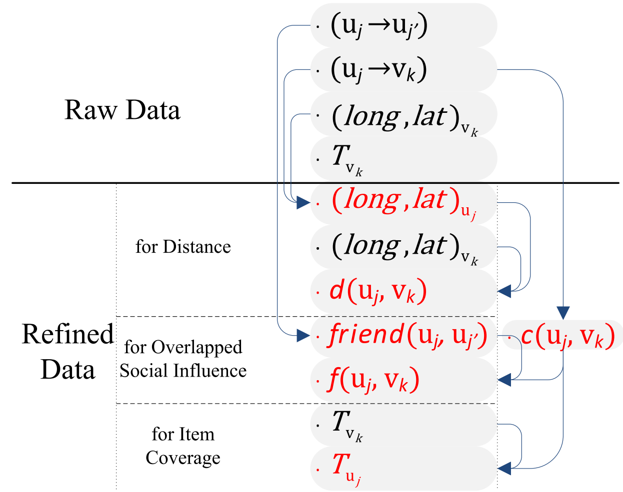


Fig. 5. Sources and results of data preprocessing.

Refined data are in the form of a complete list of visiting cases $\{<u_j, v_k, \ c(u_j, v_k), \ d(u_j, v_k), \ f(u_j, \ v_k)>\}$, where $c(u_j, v_k)$ can be easily obtained from the check-in history records $\{(u_j \rightarrow v_k)\}$. We describe the refined data for distance, overlapping social influence, and item coverage in detail in the following section.

The refined data for the distance factor includes $\{d(u_j, v_k)\}$. We should obtain a user's home location $(long, \ lat)_{u_j}$ first to calculate $d(u_j, v_k)$. A user's home location is the geographical center of his/her most frequent check-ins [6]. This information can be obtained by improving the recursive grid search [7]. Particularly, first set each venue that the user visited as the center of a circle, and set radius to 50 km. From these circles, select the circle that contains the most check-ins. Then take each visited venue within the selected circle as a center, set smaller circles (i.e., set radius to 5 km), and select the circle with the most check-ins. And so on set radius to 0.5 km. Finally, the geographical center of check-ins in this area is set as the home location. The distance between the user's home location and venue location, $d(u_j, \ v_k)$, is calculated by the length of the large circle arc on the surface of the Earth.

The refined data for the overlapping social influence includes $\{f(u_j, \ v_k)\}$. First, we define that two users are connected if they follow each other in social networks. We then set a friendship if two users are connected directly or they connect to the same user, i.e., friend $(u_j, u_{j'}) = 1$; otherwise, friend $(u_j, u_j) = 0$. Finally, let $f(u_j, \ v_k)$ be the number of u_j's friends who have already visited the venue v_k. $f(u_j, v_k) > 1$ means more than one friends of u_j have visited v_k.

The refined data for item coverage includes $\{T_{u_j}\}$ (given that $\{T_{v_k}\}$ is already known). Venues serve different items. If a user visited no less than n (a threshold) venues with the same item, we assume that this user likes this item. For instance, a user u_1 has visited three venues, v_1, v_2, and v_3, with $T_{v_1} = <0, 1, 0, 1>$, $T_{v_2} = <0, 0, 1, 1>$, and $T_{v_3} = <1, 0, 1, 1>$, respectively. If the threshold n is set as 3, we can obtain $T_{u_1} = <0, 0, 0, 1>$ on the basis of the above reasoning rule.

TABLE II
MARKETING EFFECT QUANTIFICATION MODELS

Distance	**Overlapped Influence**	**Marketing Effect**
No	No	$\boldsymbol{E}_{\mathbf{SI}}$
Yes	No	$\boldsymbol{E}_{\mathbf{SLI}}$
No	Yes	$\boldsymbol{E}_{\mathbf{TI}}$
Yes	Yes	$\boldsymbol{E}_{\mathbf{TLI}}$

III. MARKETING EFFECT QUANTIFICATION

Marketing effectiveness is the quality of how marketers perform while optimizing their spending to achieve good results [8]. Quantifying the effect achieved by marketing activities is a necessary step in the process of connecting marketing to finance. Marketing effect quantification attempts to model and calculate the marketing effect for a given offline event in this paper. The marketing effect of an event is clearly relevant with its location (better in a busier street), served items (better with more and cheaper items), scale (better in a larger scale), and participants. Once the location, served items, and scale are determined, the most dominant factor is the participants. We count how many users are influenced to visit the venue to predict the marketing effect caused by the invited participants before the real event occurs. Note that, we calculate only the number of customers influenced by the participants, excluding the customers who visit the venue by themselves. "visiting" also refers only to when a customer checks-in through LBSNs, whereas customers who physically visit the venue but do not check-in through LBSNs are excluded. Although the marketing effect defined in this paper does not cover all visiting customers in the real world, it can be used as an indicator to reflect the real-world situation. Unlike the independent cascade model usually adopted by viral marketing analysis for social influence, we limit the influence within one step because the marketing effectiveness depends on the spread of the action "visiting a venue" and not the spread of an advertisement message. This essential difference makes the spread in our setting much slower and less infectious.

Considering the distance or overlapping social influence, four different models are used for marketing effect quantification, as shown in Table II. Note that these models are all statistical models, so their parameters can be learned from empirical data.

A. E_{SI}: Sum of Individual Influences

Without considering the distance, customer u_1 is influenced (i.e., to visit the venue in a future period) with a probability of p_{SI} when u_1 has a friend u_j who is a participant of an event held at a venue (i.e., $x_j = 1$). If also we do not consider the overlapping social influence, u_1's visiting probability is $N \times p_{\mathrm{SI}}$ when u_1 is a friend of N participants. Conversely, the individual influence of the participant u_j is the sum of all his/her friends' p_{SI}. Apparently, the IF strategy results in a participant set with the maximum "sum of individual influences of all participants." In particular, the marketing effect is quantified as follows:

$$E_{\mathrm{SI}}(X) = \sum_{j=1}^{N_{\mathrm{USER}}} x_j \times \sum_{i=1}^{N_F(u_j)} p_{\mathrm{SI}} \tag{4}$$

where $N_F(u_j)$ is the number of u_j's friends. Other notations are shown in Table I.

B. E_{SLI}: Sum of Individual Local Influences

If we consider the distance, the visiting probability decreases with the increasing distance from the user home location to the venue; thus, local friends have more probabilities. This relation between the probabilities and distances follows the power law proven in [2]–[4]. We can model the visiting probability of a customer who is a friend of one (and only one) participant in the form of $p_{\mathrm{SLI}} = \alpha_{\mathrm{SLI}} \times d(u_j, v_k)^{\beta_{\mathrm{SLI}}}$, where α_{SLI} and β_{SLI} are parameters of the power law distribution. At this point, we call the sum of all friends' probabilities of a participant as the individual local influence. If we do not consider the overlapping social factor, the overall influence of the participants is simply the sum of individual local influences of all participants. In particular, the marketing effect is quantified as follows:

$$E_{\mathrm{SLI}}(X) = \sum_{j=1}^{N_{\mathrm{USER}}} x_j \times \sum_{i=1}^{N_F(u_j)} \alpha_{\mathrm{SLI}} \times d(u_i, v_k)^{\beta_{\mathrm{SLI}}}. \tag{5}$$

C. E_{TI}: Team Influence

A customer will probably be influenced by the participant set to visit the venue if he/she is a friend of at least one participant. Although the community structure [25] of users is the source of overlapping social influence, we do not characterize it from this perspective; instead, we characterize it directly by calculating how many users can be influenced by a team [26]. If we do not consider the distance, a user is influenced by his/her friends to visit the venue with a probability of p_{TI} no matter how many of his/her friends are participants. If we consider the overlapping social influence, the overall influence of the participants can be called "team influence," which should be the sum of p_{TI} of those who are friends of at least one participant. In particular, the marketing effect is quantified as follows:

$$E_{\mathrm{TI}}(X) = \sum_{j=1}^{N_{\mathrm{USER}}} p_{\mathrm{TI}} \times \mathrm{sign}\left(f\left(u_j, v_k\right)\right) \tag{6}$$

where sign $(f(u_j, v_k))$ equals 1 if $f(u_j, v_k) > 0$; otherwise, it equals 0. Remember that $f(u_j, v_k)$ is the number of u_j's friends who have already visited the venue v_k.

D. E_{TLI}: Team Local Influence

If we consider both the distance and overlapping social influence, the probability of a customer who is a friend of at least one participant can be noted as $p_{\mathrm{TLI}} = \alpha_{\mathrm{TLI}} \times d(\mathrm{u}_j, v_k)^{\beta_{\mathrm{TLI}}}$ no matter how many of his/her friends are participants. The overall influence of the participants (or the team local influence) should be the sum of p_{TLI} of those who are friends of at least one participant. In particular, the marketing effect is quantified as follows:

$$E_{\text{TLI}}(X) = \sum_{j=1}^{N_{\text{USER}}} \alpha_{\text{TLI}} \times d\big(u_j, v_k\big)^{\beta_{\text{TLI}}} \times \text{sign}\left(f\left(u_j, v_k\right)\right). \tag{7}$$

Distance and social influence are two factors that can affect the visiting possibilities, which have been proven in existing studies [19]–[23]. Based on the observations described in Section I, we can denote that considering the distance and overlapped social influence can accurately quantify the marketing effect. More knowledge clearly leads to better prediction. E_{TLI} includes knowledge of not only a user's number of friends but also who and where they are. By contrast, E_{SI} knows only a user's number of friends, E_{SLI} knows the user's number of friends and where they are, and E_{TI} knows the user's number of friends and who they are. However, this condition does not mean that the more knowledge the better because a more complex model suffers the risk of over-fitting.

IV. Marketing Effect Maximization

Regardless of the quantitative model adopted, the application requirements can be transferred to a combinatorial optimization problem as follows:

$$\begin{cases} \arg\max\limits_{X} E(X) \\ \sum_{j=1}^{N_{\text{USER}}} x_j = N_{\text{SCALE}} \\ \forall_{i=1}^{N_{\text{TAG}}} t_i,\ \sum_{j=1}^{N_{\text{USER}}} x_j \times t_i\left(u_j\right) \geq t_i(v_k). \end{cases} \tag{8}$$

For case with E_{SI} or E_{SLI}, we can solve it by linear programming easily. For E_{TI} and E_{TLI}, because they are NP-hard problems [9], we can obtain only near optimal solution by using greedy algorithms or metaheuristics such as simulated annealing (SA), genetic algorithm, tabu search, and particle swarm optimization. Given that maximizing E_{SI} and E_{SLI} are easy, they can be regarded as greedy to maximize E_{TI} and E_{TLI}. We designed an algorithm for marketing effect maximization based on SA [10]. SA is adopted because it is easy to implement. We have also attempted to use the genetic algorithm. However, we failed to find an easy way to generate several available solutions (i.e., population) in the initialization, crossover, or mutation step because the available solutions are very sparse in the corresponding vector space. Nevertheless, the efficiency of the maximizing algorithm is not a key issue in this paper. We focus only on demonstrating that this problem can be approximately solved in an acceptable running time.

We formulate Algorithm 1 that employs a random search that accepts not only changes that increase the objective function but also several changes that decrease it. The acceptance probability is controlled by a declining temperature. We specify the following parameters.

1) *Solution Space:* All available solutions $\{X\}$ that satisfy (1) and (2).

Algorithm 1 Marketing Effect Maximization Based on SA

```
1)  set t_max; t_min; down; iterateTimes;
2)  X = initialSolution();
3)  E = E(X);
4)  t = t_max;
5)  while t > t_min
6)   for iterate = 1:iterateTimes
7)      [newX, deltaE] = neighbor(X);
8)      if deltaE > 0 ||rand < exp(deltaE / t)
9)           X = newX;
10)          E = E + deltaE;
11)     end
12)  end
13)  t = t * down;
14) end
```

2) *Objective Function:* (6) or (7).
3) *Initial Solution Generator:* Initialsolution() that will be described later.
4) *Candidate Generator:* Neighbor() that will be described later.
5) *Acceptance Probability Function:* Follows the metropolis criterion (Line 8 in Algorithm 1).
6) Annealing schedule, including initial temperature ($t_{\max}$), stopping temperature ($t_{\min}$), cooling rate (*down*), and inner iterate times (*iterateTimes*). They are all empirically adjusted for this specific problem because no general way to determine the best choices exists.

The functionality of the initial solution generator initialSolution() (Line 2 in Algorithm 1) is selecting a starting point in the solution space. Given the multiple constraints, the solution space shapes are irregular. Thus, randomly selecting N_{SCALE} users may not result in an available solution. At this point, we supply a better implementation. Let "the fittest user" be a user who can make the intersection of his/her items and the restaurant's uncovered items the largest. For N_{SCALE} steps, we select the fittest user for each step until we already have N_{SCALE} users as the initial solution.

The functionality of the candidate generator neighbor() (Line 7 in Algorithm 1) is to create a new solution from a given solution. A good candidate generator follows two principles: it can navigate the whole space quickly, and nearby solutions have similar objective values (i.e., *deltaE* is relatively small). Thus, we design the candidate generator as follows: a user is replaced inside the old solution with a random outside user; the new solution is accepted if it satisfies the two constraints; otherwise, try again. This candidate generator follows the said principles because: 1) reaching its farthest solution from a given solution requires only up to N_{SCALE} steps and 2) one user is replaced each time, which results in a limited change to the objective value.

In our implementation, $t_{\max} = 1, t_{\min} = 0.0001$, $down = 0.95$, and $iterateTimes = 500$. Thus, $iterateTimes * \log(t_{\min}/t_{\max})/\log(down) = 89\,500$ iterations. Fig. 6 shows the convergence process of an SA running case. In the beginning, the objective value E largely fluctuates, which means

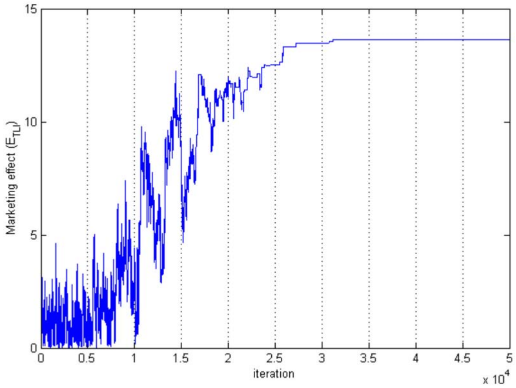


Fig. 6. Convergence of SA.

that accepting neighbors who decrease E is easy when the temperature is high. E is an increasing monotone after approximately 25 000 iterations, which means only neighbors that can increase it can be accepted when the temperature becomes low. E then converges gradually. The SA-based algorithm is more efficient in solving this problem compared with exhaustive search, which needs $C_{N_{\text{USER}}}^{N_{\text{SCALE}}}$ iterations.

V. Experimental Evaluation

This section provides experimental evaluation of the proposed framework. We use the real-world data from LBSNs to test the effectiveness of the framework. In particular, these data can be used to obtain the optimal participant set to maximize the market effect.

A. Data Set

All the experiments are conducted based on a dataset obtained from Foursquare. We use the publicly available APIs to run a crawler that collected check-in records from October 2011 to February 2012 (approximately four months). We select a tasting night hosted by a restaurant as the offline marketing event. We then select venues under the "food" category in Foursquare. Given that most customers come from the same city as the restaurant, we retained only users and venues from New York City because of its relatively higher density. The city area is a circle with a center coordinate of (40.764, −73.979) and radius of 10 km. A user is considered only if his/her home location is in New York and has checked at least one restaurant in the city. Users' social relationships are extracted from their Twitter profiles. Section II-B discusses the data refining process.

Finally, we obtain $N_{\text{USER}} = 1927$, $N_{\text{VENUE}} = 1797$, and a visiting case list with a size of $N_{\text{USER}} \times N_{\text{VENUE}} = 3\,462\,819$. We select items (tags) that are owned by at least five restaurants, whereas the items that are too special are discarded. This scenario results in TAG = {coffee, tea, alcohol, bakery, burger, pizza, dessert, bbq, vegan, fries, seafood, chicken, steak, soup, noodles, cheese, tapas, Japanese, Mexican, and Indian}.

The data statistics are shown below (Table III). 10 372 cases have $c(u_j, v_k) = 1$, so the average visiting probability is $10\,372/3\,462\,819 = 0.003$. Within these cases, the average $d(u_j, v_k)$ is 2.45 km and the average $f(u_j, v_k)$ is 0.035.

TABLE III
Data Statistics

N_{USER}	1,927
N_{VENUE}	1,797
N_{TAG}	20
# total cases	3,462,819
# cases with $c(u_j, v_k) = 1$	10,372
average $d(u_j, v_k)$	2.45 km
average $f(u_j, v_k)$	0.035
# cases with $c(u_j, v_k) = 0$	3,452,447
average $d(u_j, v_k)$	4.03 km
average $f(u_j, v_k)$	0.006
average # customers of a restaurant	5.53
average # restaurants of a customer	5.50
average # items of a restaurant	1.90
average # items of a customer	2.57
average # friends of a customer	21.22

Within the cases with $c(u_j, v_k) = 0$, the average $d(u_j, v_k)$ is 4.03 km and the average $f(u_j, v_k)$ is 0.006. A restaurant is visited by 5.53 users, serves 1.90 items, whereas a user visits 5.50 restaurants, likes 2.57 items, and has 21.22 friends on average.

B. Parameter Learning

We adopt the maximum likelihood method to estimate the parameters using simple counts. A case = <u_j, v_k, $c(u_j, v_k)$, $d(u_j, v_k)$, $f(u_j, v_k)$> (Section II). The probabilities of an influenced customer visiting the venue in the four models can be calculated as follows:

$$p_{\text{SI}} = \frac{\#\text{case with } c\left(u_j, v_k\right) = 1, f\left(u_j, v_k\right) = 1}{\#\text{case with } f\left(u_j, v_k\right) = 1} \tag{9}$$

$$\begin{aligned} p_{\text{SLI}} &= \alpha_{\text{SLI}} \times d\left(u_j, v_k\right)^{\beta_{\text{SLI}}} \\ &= \frac{\#\text{case with } c\left(u_j, v_k\right) = 1,\ d\left(u_j, v_k\right),\ f\left(u_j, v_k\right) = 1}{\#\text{case with } d\left(u_j, v_k\right),\ f\left(u_j, v_k\right) = 1} \end{aligned} \tag{10}$$

$$p_{\text{TI}} = \frac{\#\text{case with } c\left(u_j, v_k\right) = 1, f\left(u_j, v_k\right) > 0}{\#\text{case with } f\left(u_j, v_k\right) > 0} \tag{11}$$

$$\begin{aligned} p_{\text{TLI}} &= \alpha_{\text{TLI}} \times d\left(u_j, v_k\right)^{\beta_{\text{TLI}}} \\ &= \frac{\#\text{case with } c\left(u_j, v_k\right) = 1,\ d\left(u_j, v_k\right),\ f\left(u_j, v_k\right) > 0}{\#\text{case with } d\left(u_j, v_k\right),\ f\left(u_j, v_k\right) > 0} \end{aligned} \tag{12}$$

Note that, we need to discretize $d(u_j, v_k)$; otherwise, the number of cases with an exact $d(u_j, v_k)$ will be only one. We select 1 km as the interval, where $d(u_j, v_k) \in [0, 1]$ is reset to $d(u_j, v_k) = 1$, $d(u_j, v_k) \in (1, 2]$ is reset to $d(u_j, v_k) = 2$, and so on. The obtained empirical probabilities are shown in Table IV.

Table IV shows that the probability p_{SLI} and p_{TLI} increase with the decreasing distance $d(u_j, v_k)$. The probability also follows the power law distribution. This finding confirms the existence of the distance factor. The differences in p_{SI} versus p_{TI} or p_{SLI} versus p_{TLI} also show that individual influence and team influence are different.

TABLE IV
EMPIRICAL PROBABILITIES AND FITTED PARAMETERS

p / d	p_{SI}	p_{SLI}	p_{TI}	p_{TLI}
1	**0.0056** (709/125897)	**0.0161** (279/17311)	**0.0068** (1085/160088)	**0.0186** (431/23198)
2		**0.0049** (129/26105)		**0.0062** (215/34451)
3		**0.0046** (94/20371)		**0.0055** (144/26250)
4		**0.0041** (73/17885)		**0.0047** (105/22438)
5		**0.0033** (50/15222)		**0.0036** (67/18822)
6		**0.0034** (34/9986)		**0.0037** (45/12198)
7		**0.0024** (18/7408)		**0.0034** (30/8942)
…		…		…
		α_{SLI}=0.0160 β_{SLI}=-0.9671		α_{TLI}=0.0183 β_{TLI}=-0.9239

TABLE V
AVERAGE MARKETING EFFECT ($N_{\mathrm{SCALE}} = 10$)

Average		**Quantification models** E_{SI}	E_{SLI}	E_{TI}	E_{TLI}
Maximization approaches	SI	**16.11**	16.83	4.28	4.35
	SLI	16.06	**16.88**	4.26	4.34
	TI	13.10	13.68	**4.51**	4.58
	TLI	12.90	13.54	4.48	**4.60**

C. Performance Comparison

Four approaches are used to maximize four different $E(X)$ values. We use SI to denote the approach that maximizes $E_{\mathrm{SI}}(X)$ for convenience. Each approach results in a solution for a given restaurant, and we can quantify four $E(X)$ values of a solution. The first objective of performance comparison is to compare these $E(X)$ values. The second objective, which is also an important concern of any optimization problem, is to evaluate the runtime performance of the four approaches. The third objective is to check that the solutions of all four approaches satisfy the item coverage constraint. At the same time, the necessity to keep this constraint should be proved; otherwise, the solutions do not necessarily satisfy it. We use the coverage rate to measure the degree to which the restaurant items are covered by the participants

$$\text{Coverage Rate} = \frac{\left|\left(\bigcup_{x_j=1}\{\mathrm{t}_i|t_i(u_i)=1\}\right)\bigcap\{t_i|t_i(v_k)=1\}\right|}{|\{\mathrm{t}_i|t_i(v_k)=1\}|}. \tag{13}$$

All N_{VENUE} restaurants are considered to avoid prejudice against special restaurants. We set $N_{\mathrm{SCALE}} = 10$ as the default, which is typical for free trial advertising by a restaurant.

1) Average Marketing Effect: Table V shows that different solutions are quantified using the same $E(X)$ in each column, and a solution is quantified using different $E(X)$ in each row

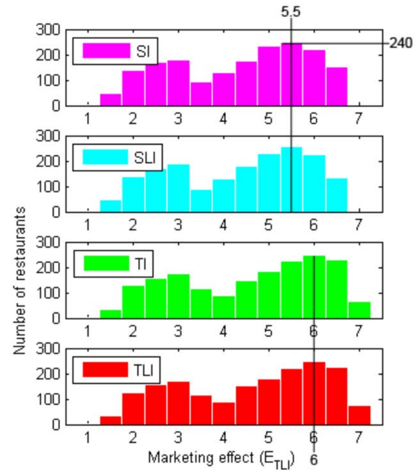


Fig. 7. Marketing effect distribution.

[e.g., row 1 column 1 (16.11) is the E_{SI} of SIs solution]. Diagonal values are always the best as a matter of course: each of them is exactly the corresponding maximization approach for the quantification models. For example, SI reaches the largest $E_{\mathrm{SI}}(X)$ at 16.11.

Given that $E_{\mathrm{TLI}}(X)$ is the most suitable quantification, we take it as a metric to compare the four approaches. The last column (i.e., 4.35, 4.34, and 4.58 versus 4.60) shows that TI behaves similarly to TLI, and both results are better than SI and SLI. Note that TI and TLI use the same algorithm based on SA, but only differ in their optimization objectives. This scenario indicates that their objectives E_{TLI} and E_{TI} are close (i.e., a solution that maximizes one nearly maximizes the other one). We can conclude that overlapping social influence plays more roles than distance in our model. SLI is the worst, even worse than SI. The reason for this result is that if we consider only the distance without considering the overlapping social influence, the approach favors the users who have more friends near the restaurant. This condition aggravates the overlapping factor. Fig. 7 shows the distribution of the $E_{\mathrm{TLI}}(X)$ of the four approaches. SI and SLI allow most restaurants (approximately 240 restaurants) to reach $E_{\mathrm{TLI}}(X) = 5.5$, whereas TI and TLI allow most restaurants to reach $E_{\mathrm{TLI}}(X) = 6$ and several to reach $E_{\mathrm{TLI}}(X) = 7$. The gap is not tiny because it is an indicator rather than an exhaustive customer number. Note that each restaurant is visited by 5.53 fresh customers on average in the past four months, whereas a restaurant can reach the other 4.60 customers on average by inviting ten participants to a tasting night by using our approach.

2) Average Time Cost: All implementations ran on a Core2 Duo 2.4 GHz PC with Windows 7 and 4 GB memory. N_{SCALE} is set to 3, 5, 10, 20, and 40. Fig. 8 shows that SI and SLI require a short period of time because of the efficiency of linear programming methods, whereas TI and TLI require hundreds of seconds to generate a solution for each restaurant. All approaches require more time to obtain solutions with the increasing N_{SCALE}. We can conclude that when more factors are generally considered, more time is required. If the problem becomes an NP-hard combinational optimization problem, the

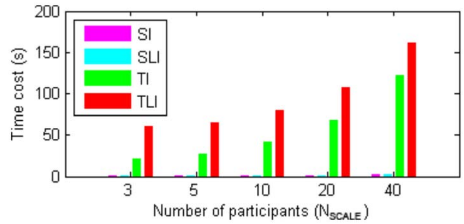


Fig. 8. Average time cost for a restaurant.

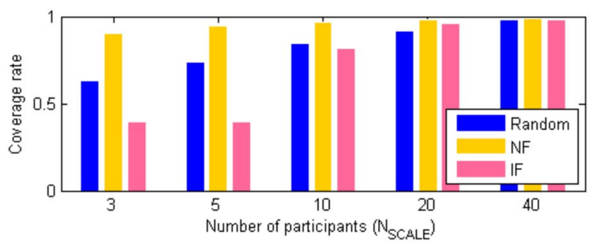


Fig. 9. Average coverage rate for a restaurant.

time cost becomes very high and the exact cost depends on the optimization algorithm that we adopted. However, obtaining a high-quality recommendation from participants within hundreds of seconds for a single restaurant is acceptable.

3) Average Coverage Rate: We take random, NF, and IF as the baseline approaches that do not consider the item coverage constraint. N_{SCALE} is set to 3, 5, 10, 20, and 40. Fig. 9 shows that all three approaches can reach a higher coverage rate with the increasing N_{SCALE}. This condition is reasonable because more participants are likely to cover more items. NF $\succ$ Random $\succ$ IF ("$\succ$" means "better than"), which indicates that local users are more likely to taste all food types, whereas the item coverage among the most influential people is limited. However, they cannot guarantee that all items owned by the restaurant are liked, even for an N_{SCALE} larger than 40. This result confirms that the item coverage constraint is necessary.

D. Case Study

We select a particular restaurant denoted as v_{494} in our data set for this case study. v_{494} is located in the center of the Manhattan district (indicated by a yellow pin in Fig. 10). It is visited by six users during the data collection period. N_{SCALE} is set to ten. Four approaches result in the following solutions.

SI: $\{u_{36}, u_{50}, u_{104}, u_{116}, u_{276}, u_{406}, u_{418}, u_{425}, u_{889}, u_{1251}\}$.

SLI: $\{u_{36}, u_{50}, u_{104}, u_{116}, u_{245}, u_{406}, u_{418}, u_{425}, u_{889}, u_{1251}\}$.

TI: $\{u_{50}, u_{58}, u_{116}, u_{316}, u_{425}, u_{524}, u_{623}, u_{997}, u_{1311}, u_{1544}\}$.

TLI: $\{u_{50}, u_{104}, u_{116}, u_{134}, u_{168}, u_{406}, u_{425}, u_{532}, u_{623}, u_{1544}\}$.

We will analyze their marketing effects and the reason for the gaps between TLI and other approaches in the following section.

1) Marketing Effect: The marketing effects of the approaches are shown in Table VI, which has the same structure as Table V, but only for v_{494} and not the average of all restaurants. These findings still exist in this particular case. For example, we have TLI $\succ$ TI $\succ$ SI $\succ$ SLI in terms of E_{TLI}.

2) Participants' Friends: Fig. 10 shows where the friends of the participants are. Fig. 10(d) contains all users that at least one friend invited as a participant (indicated with red hexagons) according to TLI. Fig. 10(a)–(c) illustrates the differences between SI (purple triangles) versus TLI, SLI (blue diamonds) versus TLI, and TI (green stars) versus TLI, as well as neglect all common users.

We find that participants recommended by SI reach fewer users, participants recommended by SLI reach even fewer but nearer users, and participants recommended by TI reach almost the same number but farther users compared with TLI. We will show the quantitative differences in detail with regard to the distance and overlapping social influence in the following sections.

3) Distance of Participants' Friends: Fig. 11 compares the distances of the participants' friends from the restaurant v_{494}. Fig. 11(d) shows the distance distribution of the participants' friends by TLI (i.e., how many friends are a distance of d km from the restaurant). Fig. 11(a)–(c) shows the differences between SI, SLI, and TI with TLI. These figures confirm that SI and SLI reach fewer near and far users, whereas TI reaches more far users but fewer near users.

4) Invited Friends of Participants' Friends: Fig. 12 shows the number of invited friends who are participants' friends. Fig. 12(d) shows the $f(u_j, v_{494})$ distribution of the participants' friends by TLI (i.e., how many friends are invited to the restaurant v_{494}). Fig. 12(a)–(c) shows the differences between SI, SLI, and TI with TLI. These results confirm that participants recommended by SI or SLI have many common friends [i.e., the overlapping factor is remarkable because more users have a larger $f(u_j, v_{494})$ compared with TLI], whereas TI and TLI allow every participant to bring a different user, thus providing a better marketing effect.

5) Item Coverage: The restaurant v_{494} serves coffee, bakery, burger, pizza, vegan, and soup. We confirm that all solutions of the four approaches can cover these items completely. However, participants recommended by IF (with items he/she liked in brackets) are as follows: u_{36} (alcohol, Mexican), u_{50} (alcohol), u_{104} (none), u_{116} (none), u_{276} (none), u_{406} (coffee, tea, alcohol, bakery, burger, pizza, dessert, vegan, chicken, noodles, Japanese), u_{418} (alcohol), u_{425} (coffee, bakery, burger, noodles), u_{281} (Mexican), and u_{1251} (none). No user likes soup in this solution.

VI. Related Work

A. Influence Maximization

Social influence has long been used to develop the market. Given that social influence can be estimated and conducted easily on social networks, using social networks as a word-of-mouth tool through self-replicating has recently become popular (viral marketing). Influence maximization selects seed customers (who cost the marketing budget) who can maximize the potential customers (who are unpaid). Domingos [11] proposed the idea that online social networks can be leveraged

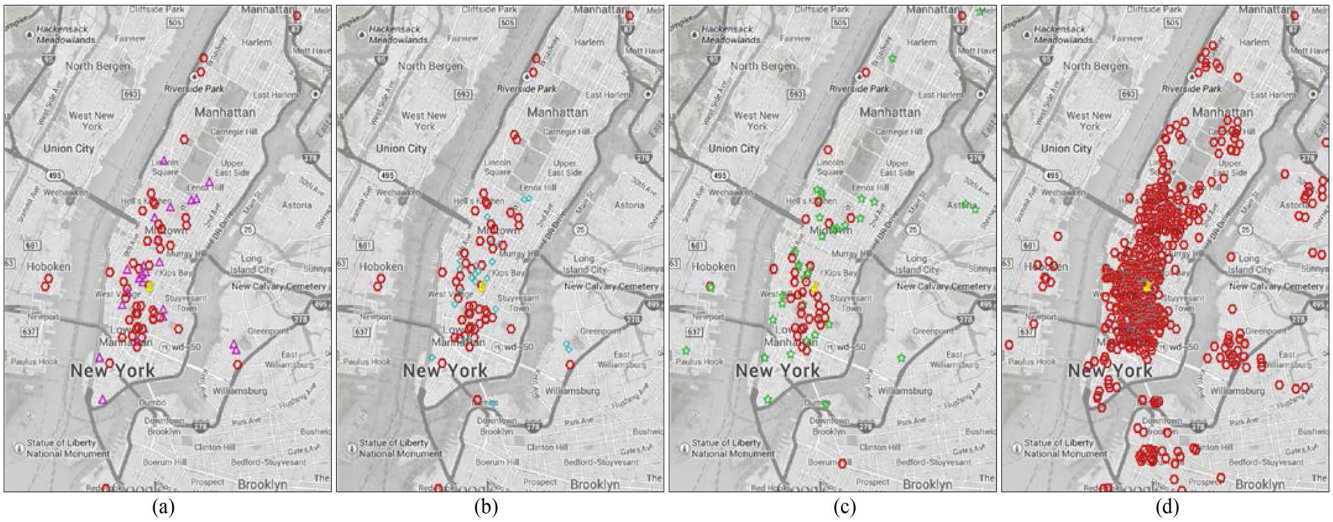


Fig. 10. Locations of participants' friends. (a) SI versus TLI. (b) SLI versus TLI. (c) TI versus TLI. (d) TLI.

TABLE VI
MARKETING EFFECT OF v_{494} ($N_{\mathrm{SCALE}} = 10$)

For v_{494}		Quantification models			
		E_{SI}	E_{SLI}	E_{TI}	E_{TLI}
Maximization approaches	SI	**16.18**	26.01	4.28	6.65
	SLI	16.13	**26.05**	4.24	6.57
	TI	11.14	17.61	**4.46**	6.75
	TLI	12.10	19.40	4.44	**6.88**

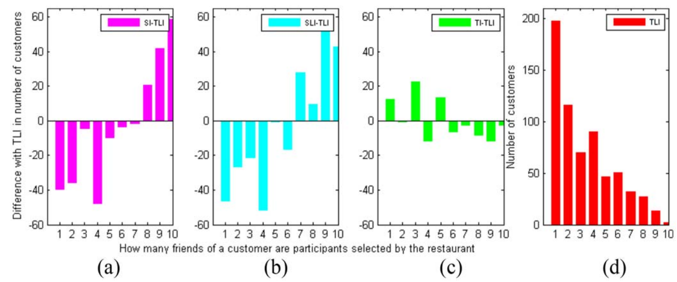


Fig. 12. $f(u_j, v_{494})$ distribution of participants' friends. (a) SI versus TLI. (b) SLI versus TLI. (c) TI versus TLI. (d) TLI.

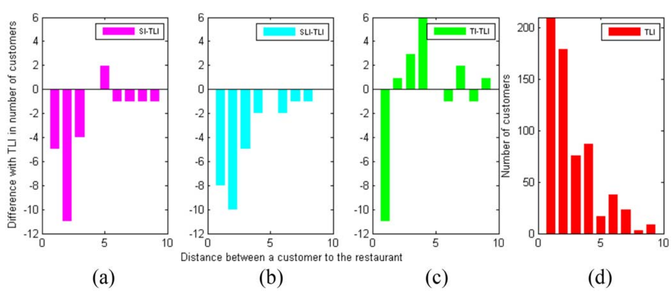


Fig. 11. $d(u_j, v_{494})$ distribution of participants' friends. (a) SI versus TLI. (b) SLI versus TLI. (c) TI versus TLI. (d) TLI.

to calculate the customer value, which provides a more precise marketing decision. Bhagat *et al.* [12] also leveraged social influence to increase profit and maximize product adoption in online social networks. However, the independent cascade model of influence maximization does not fit the offline event marketing. Viral marketing is efficient for nonspatial products (e.g., movies and books), but not for location-sensitive products (e.g., restaurants and museums). We also consider the distance and item coverage factors besides social relationships.

B. Team Formation

The team formation problem was first described by Lappas *et al.* [9]. Given a task, a user set with different skills, and a social network that captures the compatibility among these users, this problem attempts to find a subset of users to perform the task. Seminal works consider only the skill coverage [13], which is similar to the item coverage in this paper. Social relationship constraints [9], [14], online tasks [15], and load balance [16] are later included. Anagnostopoulos *et al.* [17] focused on how to satisfy these constraints at the same time. Nevertheless, all of these researchers did not consider the geographical factors of the task and users. We consider these factors in this paper in terms of decaying visiting probability with distance. Socio-spatial group query [18] is similar to team formation, which attempts to minimize total distance from group members to the rally point; however, it does not consider the social relationship and item coverage. In addition, we focus on an application-oriented framework unlike team formation, which focuses on theory issues of operational research.

C. Mobility Prediction

Mobility prediction predicts where a user will go under certain conditions (e.g., certain time and place or after a period of trajectory) and is referred to by our marketing effect quantification. Calabrese *et al.* [19] discovered that people near an event are more likely to be attracted than people who are far away; the same event type attracts people with a similar spatial origin distribution. Song *et al.* [20] and Gonzalez *et al.* [21] analyzed the mobility phenomenon of 100 000 mobile phone users and claimed that a high regularity in temporal and spatial dimension exits in human trajectory. The predictability is as high as 93%. Many mobility prediction works have emerged given the prevalence of LBSNs [22], [23]. However, the "next place" prediction or place recommendation problem is different from the one we are concerned with. We focus on the sum of the visiting probability, which is related to a type of collective mobility pattern without considering the user's individual preference.

VII. Conclusion

This paper presents a framework that assists marketers in improving marketing effectiveness by carefully selecting invitees to a sponsored offline event. The distance, overlapping social influence, and item coverage are considered in leveraging LBSNs. Given the location, served items, and scale, as well as assuming candidates are all users in the market and the marketing effect is in terms of how many customers are influenced to visit the venue, our framework can check the validity and predict the marketing effect of an arbitrary participant set. Thus, a participant set whose marketing effect (nearly) reaches the maximum in all possible participant combinations can be determined.

We can exploit methods to evaluate these models objectively in the future. The data can be split into two parts: one part for learning and the other part for testing. The real-world marketing effect that selected participants can bring in can be determined by conducting a survey. Given that the overlapping social influence plays more roles than distance, more information can be considered to better model this factor (e.g., intimacy between a user and his friend, and the user's novelty seeking characteristics). If these factors were leveraged in the future, we would require less historical data and can predict better.

References

[1] B. Schmitt, *Experience Marketing: Concepts, Frameworks and Consumer Insights*. Boston, MA, USA: Now, 2011.

[2] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, San Diego, CA, USA, 2011, pp. 1082–1090.

[3] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, Beijing, China, 2011, pp. 325–334.

[4] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, "Mining user mobility features for next place prediction in location-based services," in *Proc. IEEE 12th Int. Conf. Data Min. (ICDM)*, Brussels, Belgium, 2012, pp. 1038–1043.

[5] D. Zhang, B. Guo, and Z. Yu, "The emergence of social and community intelligence," *Computer*, vol. 44, no. 7, pp. 21–28, Feb. 2011.

[6] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo, "Socio-spatial properties of online location-based social networks," in *Proc. 5th Int. AAAI Conf. Weblogs Soc. Media (ICWSM)*, Barcelona, Spain, 2011, pp. 329–336.

[7] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring millions of footprints in location sharing services," in *Proc. 5th Int. AAAI Conf. Weblogs Soc. Media (ICWSM)*, Barcelona, Spain, Jul. 2011, pp. 81–88.

[8] Wikipedia. (Sep. 2014). *Marketing Effectiveness*. [Online]. Available: http://en.wikipedia.org/wiki/Marketing_effectiveness

[9] T. Lappas, K. Liu, and E. Terzi, "Finding a team of experts in social networks," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, Paris, France, 2009, pp. 467–476.

[10] S. Kirkpatrick, "Optimization by simulated annealing: Quantitative studies," *J. Stat. Phys.*, vol. 34, nos. 5–6, pp. 975–986, 1984.

[11] P. Domingos, "Mining social networks for viral marketing," *IEEE Intell. Syst.*, vol. 20, no. 1, pp. 80–82, Jan./Feb. 2005.

[12] S. Bhagat, A. Goyal, and L. V. S. Lakshmanan, "Maximizing product adoption in social networks," in *Proc. 5th ACM Int. Conf. Web Search Data Min.*, Seattle, WA, USA, 2012, pp. 603–612.

[13] S.-J. Chen and L. Lin, "Modeling team member characteristics for the formation of a multifunctional team in concurrent engineering," *IEEE Trans. Eng. Manage.*, vol. 51, no. 2, pp. 111–124, May 2004.

[14] M. Kargar and A. An, "Discovering top-k teams of experts with/without a leader in social networks," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, Glasgow, U.K., 2011, pp. 985–994.

[15] A. Anagnostopoulos, L. Becchetti, C. Castillo, and S. Leonardi, "Power in unity: Forming teams in large-scale community systems," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage.*, Toronto, ON, Canada, 2010, pp. 599–608.

[16] A. Majumder, S. Datta, and K. Naidu, "Capacitated team formation problem on social networks," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, Beijing, China, 2012, pp. 1005–1013.

[17] A. Anagnostopoulos, L. Becchetti, C. Castillo, A. Gionis, and S. Leonardi, "Online team formation in social networks," in *Proc. 21st Int. Conf. World Wide Web*, Lyon, France, 2012, pp. 839–848.

[18] D.-N. Yang, C.-Y. Shen, W.-C. Lee, and M.-S. Chen, "On socio-spatial group query for location-based social networks," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, Beijing, China, 2012, pp. 949–957.

[19] F. Calabrese, F. C. Pereira, G. D. Lorenzo, L. Liu, and C. Ratti, "The geography of taste: Analyzing cell-phone mobility and social events," in *Proc. 8th Int. Conf. Pervasive Comput.*, Helsinki, Finland, 2010, pp. 22–37.

[20] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[21] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

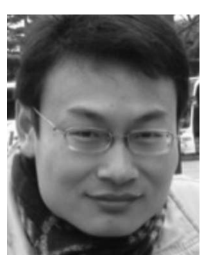
[22] W. Mathew, R. Raposo, and B. Martins, "Predicting future locations with hidden Markov models," in *Proc. ACM Conf. Ubiq. Comput.*, Pittsburgh, PA, USA, 2012, pp. 911–918.

[23] L. Backstrom, E. Sun, and C. Marlow, "Find me if you can: Improving geographical prediction with social and spatial proximity," in *Proc. 19th Int. Conf. World Wide Web*, Raleigh, NC, USA, 2010, pp. 61–70.

[24] X. Liu *et al.*, "Event-based social networks: Linking the online and offline social worlds," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min. (KDD)*, Beijing, China, 2012, pp. 1032–1040.

[25] Z. Wang *et al.*, "Discovering and profiling overlapping communities in location based social networks," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans,* vol. 44, no. 4, pp. 499–509, Apr. 2014.

[26] R. Crowder, M. Robinson, H. Hughes, and Y.-W. Sim, "The development of an agent-based modeling framework for simulating engineering team work," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 42, no. 6, pp. 1425–1439, Nov. 2012.



Zhiyong Yu (M'13) received the M.E. and Ph.D. degrees in computer science and technology from Northwestern Polytechnic University, Xi'an, China, in 2007 and 2011, respectively.

He is currently a Lecturer with the College of Mathematics and Computer Science, Fuzhou University, Fuzhou, China. He was also a Visiting Student at Kyoto University, Kyoto, Japan, from 2007 to 2009 and a Visiting Researcher at the Institut Mines-Telecom, TELECOM SudParis, Evry, France, from 2012 to 2013. His current research interests include pervasive computing, social computing, mobile social networks, and participatory sensing.

Dr. Yu is the Publicity Chair of IEEE International Conference on Ubiquitous Intelligence and Computing (UIC) 2014, a Poster Chair of SocialCom 2014, and the Workshop Chair of CloudCom-Asia 2013.

Daqing Zhang (M'11) received the Ph.D. degree from the University of Rome "La Sapienza," Rome, and the University of L'Aquila, L'Aquila, Italy, in 1996.

He is currently a Professor with the Institut Mines-Telecom, TELECOM SudParis, Evry, France. His current research interests include large-scale data mining, urban computing, context-aware computing, and ambient-assistive living. He has published over 180 referred journal and conference papers.

Prof. Zhang was the recipient of the Ten Years CoMoRea Impact Paper Award at PerCom 2013, the Best Paper Award at UIC 2012, and the Best Paper Runner-Up Award at Mobiquitous 2011. He is currently an Associate Editor of four journals, including the *ACM Transactions on Intelligent Systems and Technology*. He is a frequently invited speaker of different international events on ubiquitous computing.

Zhiwen Yu (SM'11) received the B.E., M.E., and Ph.D. degrees in engineering in computer science and technology from Northwestern Polytechnic University, Xi'an, China, in 2000, 2003, and 2005, respectively.

He is currently a Professor with the School of Computer Science, Northwestern Polytechnic University. He was a Research Fellow at Kyoto University, Kyoto, Japan, from 2007 to 2009, a Post-Doctoral Researcher at Nagoya University, Nagoya, Japan, from 2006 to 2007, and a Visiting Researcher at the Institute for Infocomm Research, Singapore, from 2004 to 2005. His current research interests include pervasive computing, context-aware systems, human-computer interaction, and personalization.

Prof. Yu was an Alexander Von Humboldt Fellow at Mannheim University, Mannheim, Germany, from 2009 to 2010.

Dingqi Yang received the B.S. degree in instrumentation engineering from Xidian University, Xi'an, China, in 2009, and the M.E. degree in telecommunication networks from the Institut Mines-Telecom, TELECOM SudParis, Evry, France, in 2011. He is currently pursuing the Ph.D. degree with the Pierre and Marie Curie University and the Institut Mines-Telecom, TELECOM SudParis.

His current research interests include ubiquitous computing, social network analysis, location-based services, and personalized information retrieval.

View publication stats