



Contents lists available at ScienceDirect

## Int. J. Human–Computer Studies

journal homepage: [www.elsevier.com/locate/ijhcs](http://www.elsevier.com/locate/ijhcs)

## Incentivizing social media users for mobile crowdsourcing

Panagiota Micholia<sup>a,\*</sup>, Merkouris Karaliopoulos<sup>a</sup>, Iordanis Koutsopoulos<sup>a</sup>, Luca Maria Aiello<sup>c</sup>, Gianmarco De Francisci Morales<sup>b</sup>, Daniele Quercia<sup>c</sup><sup>a</sup> University of Economics and Business, 76, Patission Str., GR10434 Athens, Greece<sup>b</sup> Qatar Computing Research Institute, Tornado Tower, 18th floor, Doha, Qatar<sup>c</sup> Nokia Bell Labs, Broers Building, 21 JJ Thomson Avenue, Cambridge CB30FA, UK

## ARTICLE INFO

## Keywords:

Mobile crowdsourcing  
Flickr  
Incentives

## ABSTRACT

We focus on the problem of contributor-task matching in mobile crowd-sourcing. The idea is to identify existing social media users who possess domain expertise (e.g., photography) and incentivize them to perform some tasks (e.g., take quality pictures). To this end, we propose a framework that extracts the potential contributors' expertise based on their social media activity and determines incentives for them within the constraint of a budget. This framework does so by preferentially targeting contributors who are likely to offer quality content. We evaluate our framework on Flickr data for the entire city of Barcelona and show that it ensures high levels of task quality and wide geographic coverage, all without compromising fairness.

## 1. Introduction

The computing power of mobile phones nowadays allows notifications to be sent to users on the fly, and thus labour and services can now be supplied in real time. This is made possible by the rise of mobile crowd-sourcing sites. We focus on the problem of contributor-task matching on those sites.

The main idea is that there are many social media users – especially power users – who passionately contribute to existing online communities and, as a result, become experts in specific areas but their talent remains untapped. Hence, there is a need for mechanisms to engage those experts in crowd-sourcing tasks.

To this end, we need to (1) identify experts from, e.g., their social media activity, and we do so by borrowing ideas from previous work such as Amintoosi and Kanhere (2014) and Kosinski et al. (2010); and (2) incentivize them, and we do so by identifying those who are likely to enjoy the tasks at hand, i.e., their preferences closely match the tasks.

We propose to align what social media users already do (e.g., take and post pictures on Flickr) with what a mobile crowd-sourcing marketplace demands (e.g., take pictures of local cafes and restaurants). In so doing, we make two main contributions:

- After formally describing the problem and the metrics involved (Section 2), we devise a solution for it (Section 3). We propose a framework based on a water filling algorithm in the context of mobile crowd-sourcing.

- We evaluate our proposal against real traces of Flickr users in the entire city of Barcelona (Section 4). We find that our proposal ensures high levels of quality for the completed crowdsourcing tasks and enjoys a wide geographic coverage of the city. Also, its effectiveness does not come at the expense of fairness.

## 2. Describing social media users

Our problem involves three main actors: *task suppliers*, the *service provider* ( $S_p$ ), and a set of *task contributors* ( $N$ ). The suppliers offer a pool of tasks ( $\mathcal{L}$ ), each of which comes with a budget. The service provider matches those tasks with potential contributors that are both knowledgeable and willing to carry those tasks out. To that end, the service provider should:

**Problem.** : Identify and incentivize a set of contributors in a way that the suppliers' satisfaction is maximized.

To see the practical relevance of this problem, consider a photo-sharing site such as Flickr. Flickr users are passionate about photography and share their work online. Pictures differ in quality, and those of high quality tend to be liked and further shared by other users on the site. One could imagine a situation in which Flickr users might be willing to take pictures of their city's local businesses (e.g., cafes, shops, and restaurants) under the right incentives. In this case, the goal of the service provider is to best manage the budget provided by a local business to incentivize and recruit the best Flickr users.

\* Corresponding author.

E-mail addresses: [panamixo@aubg.gr](mailto:panamixo@aubg.gr) (P. Micholia), [mkaralio@aubg.gr](mailto:mkaralio@aubg.gr) (M. Karaliopoulos), [jordan@aubg.gr](mailto:jordan@aubg.gr) (I. Koutsopoulos), [luca.aiello@nokia.com](mailto:luca.aiello@nokia.com) (L.M. Aiello), [gdfm@acm.org](mailto:gdfm@acm.org) (G.D.F. Morales), [quercia@cantab.net](mailto:quercia@cantab.net) (D. Quercia).<http://dx.doi.org/10.1016/j.ijhcs.2016.09.007>

Received 24 February 2016; Received in revised form 15 August 2016; Accepted 13 September 2016

Available online xxxx

1071-5819/ © 2016 Elsevier Ltd. All rights reserved.

**Table 1.**  
Table of symbols.

$i \in \mathcal{N}$	Contributor
$j \in \mathcal{L}$	Task
$q_i$	Quality of contributor $i$
$Q_j$	Quality score of task $j$
$\mathcal{N}_j$	Contributors to task $j$
$e_i$	Activity of contributor $i$
$f_i$	Number of positive feedbacks received by contributor $i$
$c_i$	Number of contributions by $i$
$\mathcal{A}$	Set of tasks' thematic categories
$c_i^a$	Number of contributions by $i$ in thematic area $a$
$g_i^a$	Interest of contributor $i$ to thematic area $a$
$S_p$	Service provider
$B_j$	Budget for task $j$
$\alpha_{ij}$	Attractiveness of task $j$ for contributor $i$
$p_{ij}$	Payment offered to contributor $i$ for task $j$
$w_{ij}$	Willingness of contributor $i$ to perform task $j$

### 2.1. Identifying and characterizing contributors

The best contributors for a task are identified depending on their *skills* and *willingness* to carry out the task. In what follows, we detail how we quantify these two user attributes. The ideal mechanism should combine the two main types of incentive in the literature: intrinsic motivation (you do something because you find it fun) and extrinsic one (you do something because you get paid for it). Our incentive scheme preferentially targets contributors who would not necessarily visit a place but might well offer quality content for it. This scheme is simple yet powerful, in that it mixes intrinsic motivation (pleasure for the task) with extrinsic one (monetary gain). Explanations for all the symbols that follow can be found in (Table 1).

*Skilled contributor:* A contributor's skill depends on how likely her/his contributions will meet the expertise required by the task. In Flickr, this expertise consists in taking high quality pictures. More generally, in social media, we assume that one's expertise is reflected by the number of likes one's content has received. In previous work (Schifanella et al., 2015), it has been shown that there is a correlation between quality and popularity in Flickr pictures (rank correlation 0.4). So liked pictures are generally of quality. However, quality and likes do not go always together. That has been shown to be true for pictures that are not liked at all: those pictures can be either of questionable quality (more likely) or of decent quality (less likely). By combining those two considerations, one sees that the set of quality pictures consists of pictures with many likes and pictures with a few likes or none. In a conservative fashion, we assume that only pictures with many likes are of quality. Thus contributor  $i$ 's quality becomes:

$$q_i = \frac{f_i}{c_i} \frac{1}{\max_{k \in \mathcal{N}} \frac{f_k}{c_k}}$$

where  $f_i$  is the number of positive feedback that contributor  $i$  has collected overall (e.g.,  $i$ 's number of likes),  $c_i$  is the total number of  $i$ 's contributions (e.g.,  $i$ 's number of pictures), and  $\mathcal{N}$  is the set of all contributors. The ratio at the denominator normalizes  $q_i$  in the interval  $[0, 1]$ .

The aggregate quality of a task  $j$  depends on the qualities of all  $i$ 's contributors. Based on the context, this dependency could be:

- Either linear, leading to a *quality score* for task  $j$

$$Q_j = \sum_{i \in \mathcal{N}_j} q_i \quad (1)$$

where  $\mathcal{N}_j$  is the set of task  $j$ 's contributors;

- or of diminishing returns:

$$Q_j = 1 - \prod_{i \in \mathcal{N}_j} (1 - q_i) \quad (2)$$

*Willing contributor:* In addition to being able to carry out a task, a contributor should also be willing to do it. This depends not only on whether the contributor is active on the site at all but also on whether the task is of interest.

- *Contributor's activity:*  $i$ 's activity is the  $i$ 's total number of contributions  $c_i$  normalized by the most active contributor

$$e_i = \frac{c_i}{\max_{k \in \mathcal{N}} c_k}$$

- *Contributor's interests:*  $i$ 's interest in area  $a$  is

$$g_i^a = \frac{c_i^a}{c_i}$$

where  $c_i^a$  is the number of  $i$ 's contributions that belong to category  $a \in \mathcal{A}$ . For instance, in Flickr, these categories are reflected by photo tags such as "food" and "arts". We made the choice of representing user interests as a vector of categories because previous work showed that, from very simple information such as Facebook likes, one can even predict personality traits (Kosinski et al., 2013). That might be worrying because of privacy concerns. However, our way of gathering user interests is far more coarse-grained than the information that is already publicly available on social media profiles. Consequently, our approach ends up being more privacy-conscious than what existing social media sites tend to be.

Upon activity and interest, we could compute *willingness*. One way of doing so is out of the activity and interest measures:

$$\alpha_{ij} = \frac{e_i + g_i^a}{2}$$

However, this formula assumes that a contributor's willingness scales linearly with the type and quantity of contributions. In reality, willingness is subject to diminishing returns: the willingness of a contributor who has taken one picture and that of a contributor who has taken ten pictures are likely to differ less than ten times. As such, we could define the willingness of contributor  $i$  to perform task  $j$  as:

$$w_{ij}(a_{ij}) = 1 - e^{-\alpha_{ij}} \quad (3)$$

This definition has three main desirable properties: (i) it ranges in  $[0, 1]$  so that it can be treated as a probability; (ii) it is monotonic with respect to  $\alpha_{ij}$ , i.e., the higher the attractiveness of a task to a contributor, the higher their willingness to carry it out; and (iii) it shows *diminishing returns*.

Yet, again, this new formula assumes that we do not pay contributors. To model the effect of payment on a contributor's willingness function, we incorporate  $p_{ij}$  into (3) and have:

$$w_{ij}(\alpha_{ij}, p_{ij}) = 1 - e^{-(\gamma_i^\alpha \alpha_{ij} + \gamma_i^p p_{ij})} \quad (4)$$

The two non-negative parameters  $\gamma_i^\alpha$  and  $\gamma_i^p$  weigh the effect of payment  $p_{ij}$  against  $\alpha_{ij}$ . They should reflect the social media community at hand: it should reflect how sensitive to payments the community's members generally tend to be. Payments may be either monetary, vouchers, or coupons to be redeemed from the task supplier (e.g., free coffee and discounts on the admission fee).

### 3. Task assignment and budget allocation framework

A task can be assigned to one or more contributors. Given that the

budget is limited, a payment allocation strategy is needed to select the set of contributors that maximizes the overall task quality. This joint task and budget allocation problem is presented here under two scenarios. In the *oracle (offline) scenario*, the service provider  $S_p$  knows exactly when and where the potentials contributors are available at any point in time. In [Section 3.1](#), we formalize the oracle scenario as an optimization problem and describe a numerical method for its solution. In the *live (online) scenario*,  $S_p$  knows only those who are currently available. In [Section 3.2](#), we present a heuristic approach to solve the live scenario.

### 3.1. The oracle scenario

In the oracle scenario, since we know where contributors are and when they are available, we set out to solve a global optimization problem. The objective is to allocate the budget  $B_j$  to a subset of contributors  $\mathcal{N}_j \subseteq \mathcal{N}$ . When a payment for a task is offered, the contributor's willingness to do that task increases (as per Formula (4)). The expected return of giving payment  $p_{ij}$  to contributor  $i$  can thus be expressed as  $i$ 's quality multiplied by  $i$ 's willingness to complete the task.

$$\begin{aligned} & \text{maximize} && \sum_{j \in \mathcal{L}} \sum_{i \in \mathcal{N}} q_i^{w_{ij}}(\alpha_{ij}, p_{ij}) x_{ij} \\ & \text{subject to} && \sum_{i \in \mathcal{N}} p_{ij} x_{ij} = B_j, \quad \forall j \in \mathcal{L} \\ & && p_{ij} \geq 0, \quad \forall j \in \mathcal{L}, \quad \forall i \in \mathcal{N} \\ & && x_{ij} \in \{0, 1\}, \quad \forall j \in \mathcal{L}, \quad \forall i \in \mathcal{N} \end{aligned} \quad (\text{P0})$$

where  $x_{ij}$  is a binary variable that is 1, if task  $j$  is assigned to  $i$ ; otherwise, it is 0. To make the maximization problem tractable, we do not maximize over all the contributors  $\mathcal{N}$  but only over a selected number. This selection follows either of two *assignment rules*. After selecting all contributors within geographical reach, task  $j$  is assigned to contributor  $i$ : (1) if  $j$  is the best matching task based on  $i$ 's interests<sup>1</sup> (*interest-based*); or (2) if  $j$  is the geographically closest task (*proximity-based*).

Once the restricted set of contributors  $\mathcal{N}_j$  is determined for each task  $j$ , the optimal budget distribution is computed, as we shall detail next.

#### 3.1.1. Task assignment and budget allocation problem

The next goal is to allocate payments. We do so through the following optimization:

$$\begin{aligned} & \text{maximize} && \sum_{j \in \mathcal{L}} \sum_{i \in \mathcal{N}_j} q_i^{w_{ij}}(\alpha_{ij}, p_{ij}) \\ & \text{subject to} && \sum_{i \in \mathcal{N}_j} p_{ij} = B_j, \quad \forall j \in \mathcal{L}, \\ & && p_{ij} \geq 0, \quad \forall j \in \mathcal{L}, \quad \forall i \in \mathcal{N}_j \end{aligned} \quad (\text{P1})$$

This optimization maximizes the expected returns from the payments and can be viewed as an instance of the broader family of Generalized Assignment Problems (GAP) (see, for example, [Martello and Toth, 1990, Chapter 7](#)). The additive objective function in (P1) is separable and can be written as  $|\mathcal{L}|$  different summands to be maximized independently, one for each task. Hence, the original optimization problem can be rewritten as  $|\mathcal{L}|$  independent discrete problems of the form (P2):

$$\begin{aligned} & \text{minimize} && - \sum_{i \in \mathcal{N}_j} q_i^{w_{ij}}(\alpha_{ij}, p_{ij}) \\ & \text{subject to} && \sum_{i \in \mathcal{N}_j} p_{ij} - B_j = 0, \quad j \in \mathcal{L} \\ & && -p_{ij} \leq 0, \quad \forall j \in \mathcal{L}, \quad \forall i \in \mathcal{N}_j \end{aligned} \quad (\text{P2})$$

**Algorithm 1.** Water filling algorithm for task  $j$ .

<sup>1</sup>  $i$ 's interests are reflected by  $w_{ij}(\alpha_{ij}, 0)$ ,  $\gamma_i^\alpha = 1$  in Formula (4). The 0 expresses the fact that in the selection of contributors there is no payment.

---

```

1: Input:  $B_j, \{\gamma_i^\alpha\}, \{\gamma_i^p\}, \{\alpha_{ij}\}, \{q_i\}, i \in \mathcal{N}_j$ 
2: Output:  $\{p_{ij}\}$ 
   Initialization
3:  $wl(i) = (\gamma_i^\alpha \alpha_{ij} - \ln(\gamma_i^p) \cdot q_i)$ ;  $p_{ij} = 0$ ;  $c = 1$ ;
   Main Body
4: Sort  $\mathcal{N}_j$  in order of increasing  $wl$ 
5: while  $c < |\mathcal{N}_j|$  and  $\sum_{i \in \mathcal{N}_j} p_{ij} < B_j$  do
6:    $lag = wl(c + 1) - wl(c)$ 
7:   if  $\sum_{j=1}^c \frac{lag}{\gamma_i^p} \leq B_j - \sum_{i \in \mathcal{N}_j} p_{ij}$  then
8:     increase payments of each user  $i$  among bottom- $c$  contributors by  $lag/\gamma_i^p$ 
9:   else
10:    increase payments of each user  $i$  among bottom- $c$  contributors by  $\frac{\gamma_i^p}{\gamma_i^p} p_b$ , where  $p_b = \frac{1}{(1 + \sum_{j=1}^c \frac{1}{\gamma_i^p})/\gamma_j^p} \cdot (B_j - \sum_{i \in \mathcal{N}_j} p_{ij})$ 
11:   end if
12:    $c = c + 1$ 
13: end while

```

---

one for each task  $j \in \mathcal{L}$ . From (P2) we compute  $p_{ij}$  considering that the convex minimization problem with linear constraints can be solved using the Karush–Kuhn–Tucker (KKT) first-order necessary conditions for the existence of a local optimum ([Avriel, 2003](#)). Since the objective function is convex, the local optimum is also a global one. The application of KKT (see [Appendix A](#) for the detailed derivation) yields the following expression for  $p_{ij}$ :

$$p_{ij} = \begin{cases} \frac{1}{\gamma_i^p} \left[ \ln \frac{1}{\lambda_j} - (\gamma_i^\alpha \alpha_{ij} - \ln(q_i \gamma_i^p)) \right] & \text{if } \lambda_j < q_i \gamma_i^p e^{-\gamma_i^\alpha \alpha_{ij}} \\ 0 & \text{if } \lambda_j \geq q_i \gamma_i^p e^{-\gamma_i^\alpha \alpha_{ij}} \end{cases} \quad (5)$$

where  $\lambda_j$  is the Lagrange multiplier for the single equality constraint in (P2). The direct closed-form computation of  $p_{ij}$  and  $\lambda_j$  is not possible since it is not known which instances of  $p_{ij}$  are zeros and which ones are positive. However, their values can be computed numerically by applying the so-called *water filling* algorithm ([Palomar and Feron, 2005](#)). Initially, the algorithm sorts all contributors in increasing order of *water level* defined as:

$$wl(i) = \gamma_i^p p_{ij} = \gamma_i^\alpha \alpha_{ij} - \ln(\gamma_i^p q_i) \quad (6)$$

The payments, which are initialized to zero, are then increased in successive rounds. In the first round, the payment is offered to the contributor with the minimum level. In the  $n$ 'th round, payments are offered to  $n$  contributors such that their levels will reach the  $(n + 1)$ 'th minimum level. The algorithm stops when the budget is exhausted. After termination, the value of water level  $wl$  for all contributors who receive an offer is equal to  $-\ln \lambda$ . The procedure is described in detail in [Algorithm 1](#) and schematically represented in [Fig. 1](#).

### 3.2. The live scenario

As opposed to the oracle scenario, in the live scenario, we do not know where contributors are and when they are available; hence, therefore a global optimization problem cannot be formulated. Every time a contributor becomes available, we make  $K$  offers to the contributor and (s)he needs to decide whether to take one of them. Since an offer consists of a task and a corresponding payment, we need to assign the task and then determine the payments.

#### 3.2.1. Task-to-contributor assignment

We offer the  $K$  tasks in three different ways. In addition to the interest-based and proximity-based strategies defined in [Section 3.1](#), we add a *help-the-weakest* strategy. This rule accounts for each task's current status, that is, its residual budget, residual time, and its current task quality score  $Q_j(t)$ . It works by making offers, at a given time  $t$ , for

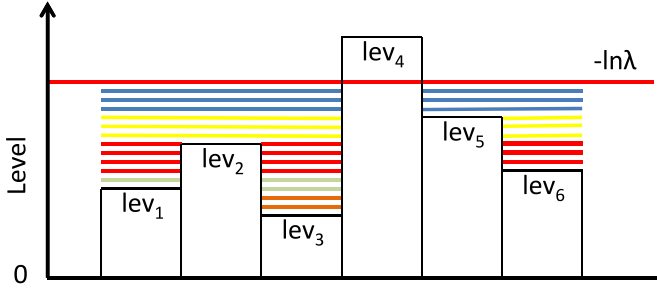


Fig. 1. Schematic description of the water filling algorithm for a task  $j$ . The bars' heights reflect the initial water level,  $wl(i)$ , of each contributor  $i$ . Lines of different colors represent the successive increases of weighted payments  $\gamma_i^p p_{ij}$  over different rounds of the algorithm. Eventually, offers are made to all contributors, except for contributor 4.

the  $K$  tasks with the lowest quality scores.

### 3.2.2. Payment determination

The payment is a function of the contributor's willingness and intrinsic ability. The intrinsic ability is fixed, while the payment and willingness can be changed. To change them, we need to estimate them. We cannot estimate the willingness, so we set it to a desired target value and determine the corresponding payment each time. Such a value should match the provider's ability of spending the residual budget within the time left. That is, the more the residual budget and the less the time left (the higher the ratio budget over time), the higher the payment should be. Now, to see how this plays out over time, consider the time between the campaign's start and its end. In this time window, if one were to pay at a constant rate, this rate would be  $B_j/T_j$ . However, since payments are not precisely constant, we need to make adjustments over time depending on:

$$adj_i = \frac{B_j(t)/T_j(t)}{B_j/T_j} \quad (7)$$

If  $adj_i > 0$ , then we have been paying up to time  $t$  more than the constant rate would suggest. If it is lower, we have been paying less. To converge to the constant rate, we need to set the target willingness as:

$$w_{reg}(t) = \min((1 - c) \sqrt{q_i(t)} + c \cdot adj_i, 1) \quad (8)$$

where  $c$  balances the relative weight of the adjustment parameter. By having those quantities into Eq. (4), we obtain the payment as:

$$p_{ij}(t) = \max(0, \frac{-\ln(1 - w_{reg}(t)) - \gamma_i^\alpha \alpha_{ij}}{\gamma_i^p}) \quad (9)$$

## 4. Experiments

The goal of our framework is to complete the tasks with the highest quality whilst ensuring the fairest payment. That is, the goal is to maximize both the suppliers' satisfaction and the contributors' payments. To ensure that our framework meets this goal, our evaluation ought to answer four main questions:

- Which of the two ways of allocating the budget (waterfilling in the oracle scenario vs. heuristic in the live scenario) results in a high number of completed tasks with a high quality? (Section 4.4.2)
- If the contributors have the possibility of choosing between multiple offers, does it impact the cumulative quality and the way budget is spent? (Section 4.4.3)
- Which task-to-contributor assignment rule (proximity, interest, help the weakest) is more effective? (Section 4.4.4)
- Does the quality depend on geographic density and task types? (Section 4.4.5)

We test the performance of both the oracle and live methods with three main indicators: *quality score*  $Q_j$  (cumulative quality contributed to the task), *budget spent*, and *task coverage* (fraction of tasks with  $Q_j > 0$ ). Next, we introduce some algorithmic baselines as comparison references, the dataset we use for the evaluation, and the simulation set-up. In Section 4.4, we discuss the simulation results and we answer the four questions above.

### 4.1. Baselines for the payment allocation

In the oracle scenario, we compare our waterfilling algorithm against two simple baselines.

*No payment baseline*: This reflects a situation in which there is no incentive mechanism: contributors are not paid, and, when they contribute, they do so purely because of their interests.

*Fixed payment baseline*: All contributors are paid the same fixed amount, computed by equally dividing the task budget  $B_j$  to the set of contributors  $N_j$ , i.e.,  $p_{ij} = B_j/|N_j|$ ,  $\forall i \in N_j$ .

### 4.2. Datasets

In our scenario, Flickr users are incentivized to take pictures of various Foursquare venues in Barcelona.

*Foursquare*: We collect information from 14k+ Foursquare venues in Barcelona, and categorize them into nine top-level categories (the set  $\mathcal{A}$  in Section 2.1): *Arts & Entertainment*, *College & Education*, *Food*, *Nightlife*, *Outdoors & Recreation*, *Shops*, *Travel & Transport*, *Professional & Other Places*, and *Residence*.

*Flickr*: Out of the full set of the public geo-referenced Flickr pictures, we select a random sample of 3.2M photos taken by 77k distinct users within the bounding box of Barcelona in 2014. For each picture, we collect the anonymized *owner* identifier, the free-text *tags* attached to the photo by the owner, the *timestamp* noting when the photo was taken, and its *geographic coordinates*. The 15M *tags* in our set are then matched to the nine top-level Foursquare categories.

### 4.3. Simulation set-up

To evaluate our framework we resort to a data-driven agent-based simulation. Agents are Flickr users who have taken geo-tagged pictures. Each geo-tagged photo is mapped to a simulation event represented by a tuple  $(t, u, l)$ , where  $t$  is the event's timestamp,  $u$  is the user, and  $l$  is the location (i.e., geographic coordinates) of the photo. We coalesce multiple time- and space-contiguous photos into a single event to avoid counting a single upload of multiple shots multiple times. Starting from our original datasets, after matching and filtering, we obtain a set of 1800 Flickr users who are involved in approximately 20k events. We also do one-to-one mappings of 1000 (a representative number for our user set) Foursquare venues to tasks around the city.

Upon each event, the simulator selects a subset of up to  $K$  pending tasks within distance  $D$  from the contributor's  $i$  current location and ranks them according to the task-to-contributor assignment rule. The contributor parses the ordered list and considers sequentially the offers, which include information about the venue that requests the photo-shooting task, as well as the personalized payment that is meant to serve as incentive. With probability  $w_{ij}$ , the contributor  $i$  executes task  $j$ , the task satisfaction index  $Q_j$  is increased by  $q_i$  according to Eq. (1), the task budget  $B_j$  is decreased by  $p_{ij}$ , and the contributor stops parsing the list. Otherwise, with probability  $1 - w_{ij}$ , the contributor parses the next task in the ranked list. If the whole list is scanned without selecting a task, the contributor does not contribute to any task.

Unless otherwise stated,  $B_j = B$  for all  $j \in \mathcal{L}$  and  $D = 1.5$  km (which conservatively represents ten times the typical block walking distance). We run experiments with several combinations of the parameters  $\gamma_i^p$  and  $\gamma_i^\alpha$  at the exponent of the willingness function. We plot results for

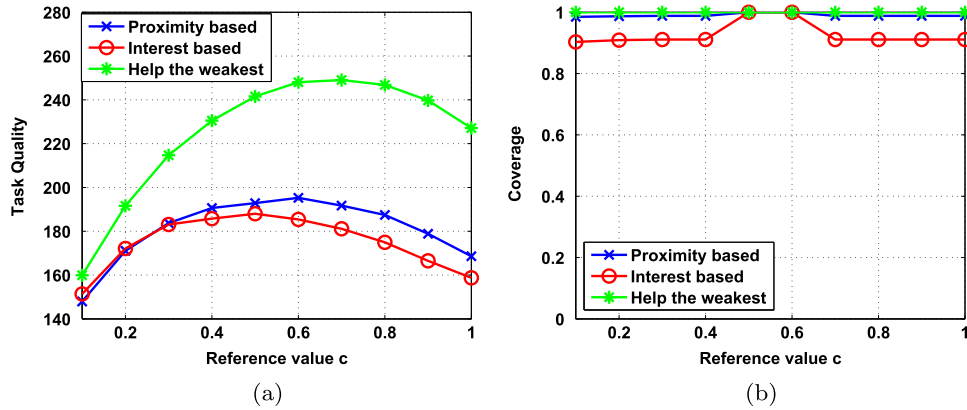


Fig. 2. Task (a) quality and (b) coverage obtained by the heuristic as a function of the reference value  $c$  for different assignment rules.

$\gamma_i^p = 0.3$  and  $\gamma_i^a = 1$ , which ensure the exploration of adequate ranges for  $w_{ij}$  and  $a_{ij}$  in our experiments.

#### 4.4. Results

##### 4.4.1. Parameter tuning for heuristic

We first find the value of parameter  $c$  in Eq. (8) that optimizes the performance of the heuristic. In Fig. 2a, the achieved aggregate task quality under all three assignment rules is maximized when  $c$  is in the interval  $[0.5, 0.7]$ . At low  $c$  values, the initial offers tend to be too small, contributors most often reject them and the offers grow only towards the very end of the campaign, given the low budget consumption rate. Conversely, for high  $c$  values, the difference in payments with the contributor quality index  $q_i$  is minimal, and the payments are no longer directed towards the more skilled contributors. The task coverage (Fig. 2b) is not very sensitive to  $c$ . Under the interest-based and help-the-weakest assignment rules, the heuristic achieves full coverage irrespective of the value of  $c$ .

##### 4.4.2. Task quality across payment methods

Fig. 3a plots the overall achieved quality as a function of the available task budget  $B$  in the oracle scenario (solved with waterfilling or with the fixed-payment baseline) and in the live scenario (solved with our heuristic approach, for different values of  $K$ ).

In the absence of payments ( $B = 0$ ), the quality results into a minimum value that sets the common reference point for all the single-offer (i.e.,  $K = 1$ ) curves. The introduction of incentives ( $B > 0$ ) always increases the aggregate quality but to different extents depending on the budget allocation schemes. The single-offer heuristic positions itself between the fixed payment baseline and the water filling scheme across all budget values. The distribution of the task quality values across all

tasks in Fig. 3b (for  $B = 200$ ) points to *stochastic dominance* relationships between the three schemes. The attracted quality of contributions under the waterfilling scheme stochastically dominates the heuristic, which in turn stochastically dominates the fixed payment scheme.

In summary, when only a single offer is given, the oracle scenario solved with waterfilling yields better overall quality than the live scenario solved with the heuristic approach.

In Fig. 4, we compare all budget allocation schemes with respect to the budget amount they spend for their achieved task quality. For small budget values, which do not suffice to pay everybody, the water filling algorithm focuses its payments to the most skilled contributors who are not intrinsically motivated to contribute to a task. The algorithm makes maximal use of budget  $B$ , hence its efficiency decreases for higher values of  $B$ . Large payments are offered to a greater number of contributors and the willingness of the contributors is subject to diminishing returns with respect to payments. In this way, payment diversity does not reflect the diversity in quality contributions and all contributors will be highly rewarded. On the contrary, the heuristic sets a strict upper limit to the offers made ( $w_{max}$  parameter in Eq. (8)), which does not change as the budget  $B$  increases. Therefore, it avoids the unnecessary high payments computed by the fixed payment baseline. The same level of task quality is achieved with much lower budget.

##### 4.4.3. Increasing the number of offers $K$

More offers per contributor increase the chances that the contributor will eventually pick up one of them. Figs. 3a and b show that the heuristic outperforms the waterfilling approach for  $K = 3$ . Hence, the heuristic effectively overcomes all the main limitations of the oracle waterfilling (need of complete information and intractability of the problem for  $K > 1$ ) and exhibits superior performance when combined with multiple offers.

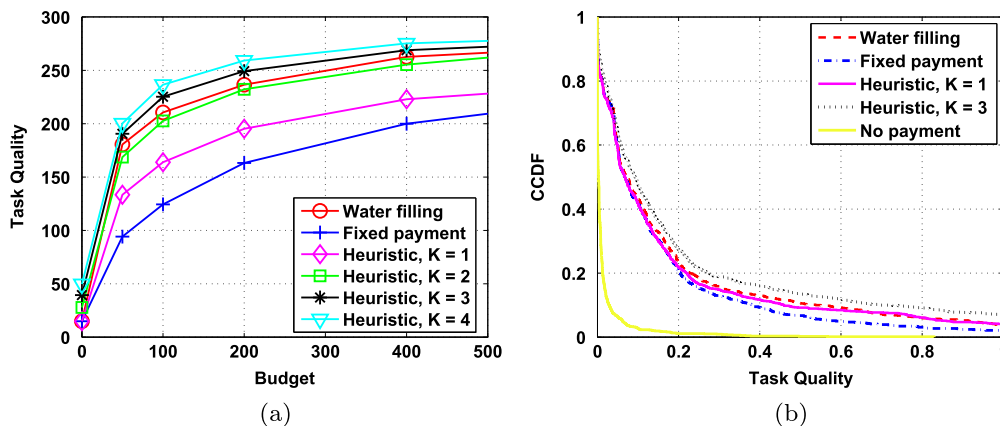


Fig. 3. (a) Task quality and (b) its distribution for different allocation algorithms, all following the proximity-based rule with  $B = 200$ .

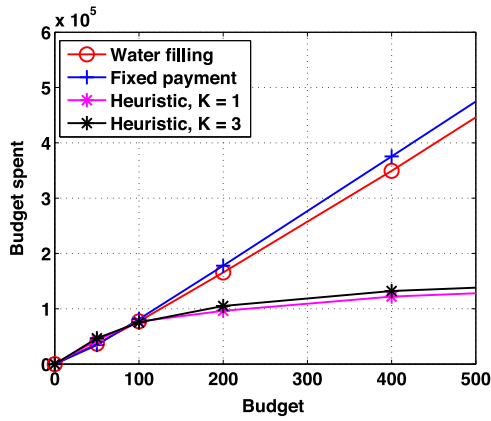


Fig. 4. Budget spent as a function of the available budget per task.

4.4.4. Task quality across assignment rules

Fig. 5a plots the achieved task quality as a function of  $K$  for all three task assignment rules. The overall trend shows a diminishing return property, as expected. The help-the-weakest rule achieves the best score and saturates faster than the other two rules. It does so by tapping into otherwise-underused task budgets and generating larger offers on average, as can be seen from the markedly higher value of spent budget in Fig. 5b. To assess the quality/budget tradeoff, we measure the average budget spent per quality unit (Fig. 6). The interest-based rule is the most efficient, since it aligns tasks to user interests, and thus needs smaller payments to achieve the same level of willingness. The help-the-weakest rule is instead the most expensive, bearing the cost of the imposed fairness, that is reached by incentivizing contributors to tackle unpopular tasks.

4.4.5. Impact of geographic density and task types on task quality

We investigate how the task depends on the geographical density of venues and the type of tasks.

*Area density:* The completion of task in low-density areas might be challenging. To test that, we compute the distribution of contributor density across all venues and classify them into four quartiles. Each quartile denotes the 25% of contributor visits, with quartile Q1 including the bottom 25% of venues and Q4 being the top 25%. Fig. 7a compares the three budget allocation algorithms and the no payment baseline under the proximity-based assignment rule. Fig. 7b compares all three task assignment rules when combined with the heuristic. In both figures, we report the *actual achieved* task quality in each quartile normalized by the maximum *achievable* one had all offers made to the contributors been accepted.

All three budget allocation schemes appear to achieve higher scores

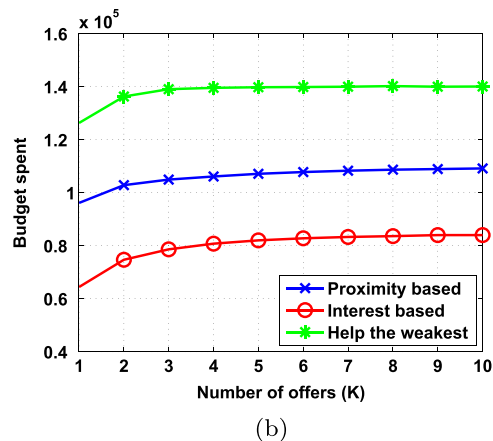
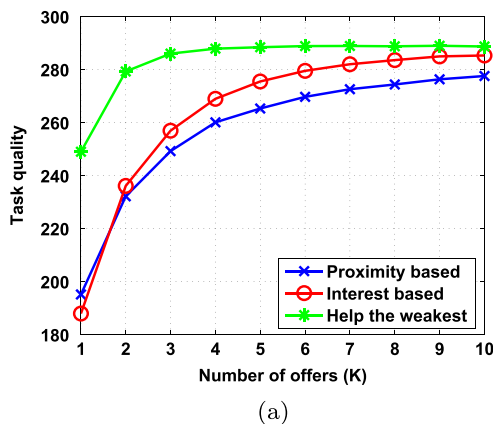


Fig. 5. (a) Task quality and (b) budget spent, for the heuristic as a function of the number of offers  $K$  under the three task assignment rules with  $B = 200$ .

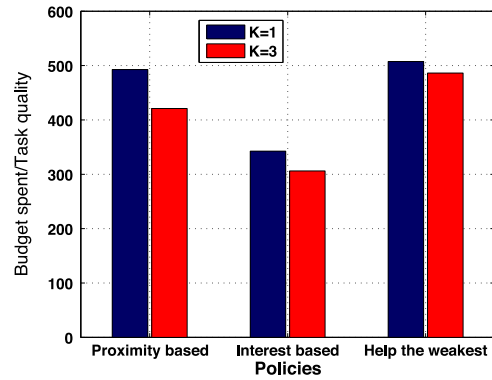


Fig. 6. Budget spent per quality unit achieved for the heuristic under all task assignment rules with  $B = 200$ .

of normalized quality in the lower three quartiles than in the top one (Q4). That is, all schemes tend to take advantage of almost every single contribution opportunity for tasks where opportunities are scarce, while they are less effective for the top-25% of venues. In the most visited venues, the task budget is split over more contributors, payments tend to be lower and result greater number of rejected offers. Fig. 7a also implies that the performance differentiation of the three budget allocation schemes, as earlier reported in Fig. 3a, is due to the way they treat the top-quartile venues. Note that in the absence of incentives, there is not much difference among the quartiles, so the differences highlighted above can be entirely attributed to the budget allocation algorithms.

We observe a similar pattern for the heuristic in Fig. 7b. The proximity- and interest-based task assignment rules yield lower scores for tasks at the top quartile. The help-the-weakest rule, on the contrary, distributes the quality values more uniformly across the four quartiles.

*Task category:* Certain types of task might not attract contributors. In our dataset, we observe a mismatch between the interests of the contributors and the available tasks (see Table 2). Whereas Flickr users are mostly keen on taking pictures outdoors in parks, the Foursquare venues fall predominantly under the categories *Shops*, *Professional & Other Places*, and *Food*. We explore how much the alternative schemes for task assignment and budget distribution resolve this mismatch.

In Fig. 8a, we plot the task quality attracted for each category under the proximity-based assignment rule. In line with intuition, the three categories with the top scores are *Shops*, *Professional & Other Places*, and *Food*. The contributors, hence, are directed to venues of the most popular Foursquare categories. All budget allocation schemes only amplify a difference that is already present even without any incentives. The trend is reversed under the interest-based rule, as shown in Fig. 8b. Irrespective of the payment allocation algorithm, contributors

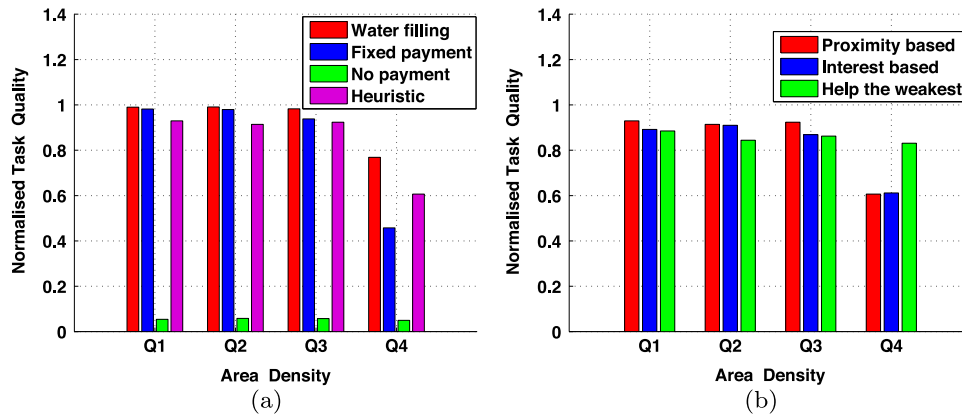


Fig. 7. Normalized task quality in areas of different contributor density (a) for different payment allocation algorithms under the proximity-based rule, (b) for the heuristic under all task assignment rules for  $K=1$  and  $B=200$ .

Table 2.

Distribution of Flickr user interests over Foursquare categories and percentage of tasks in each thematic category.

Categories	User interests (%)	Tasks (%)
Arts & Entertainment	8.64	4.80
College & Education	0.17	4.70
Food	0.78	20.80
Nightlife Spot	0.06	6.10
Outdoors & Recreation	83.49	5.20
Professional & Other Places	0.17	23.80
Residence	0.00	0.90
Shop & Service	0.55	26.10
Travel & Transport	6.15	7.60

follow their interests, as expressed in Flickr, and mainly carry out tasks in the top *Outdoors & Recreation* category, followed by *Arts & Entertainment* and *Travel & Transport*. As a result, the contributions are concentrated on the few venues that are of interest to the contributors, while the majority of the venues partly suffer. Clearly,

the interest-based rule is not the recommended option when the interests of the contributors do not match the theme of the available tasks. Finally, the help-the-weakest assignment rule follows the trend of the proximity-based one, achieving higher task quality for the three most frequent venue categories. Given that the help-the-weakest rule almost always dominates the proximity one, and the achieved quality is more evenly distributed among the available tasks (see Fig. 8c), we consider it to be the best choice overall.

5. Related work

5.1. From online to offline crowdsourcing

Online crowdsourcing has been used in a wide range of domains (Little, 2009; Kittur et al., 2011; Irani et al., 2013), including natural language processing (Callison-Burch, 2009; Bernstein et al., 2010), annotation of images (Rashtchian et al., 2010), and evaluation of creativity for multimedia content (Schifanella et al., 2015). There are still challenges to be addressed though, not least on how to formulate a

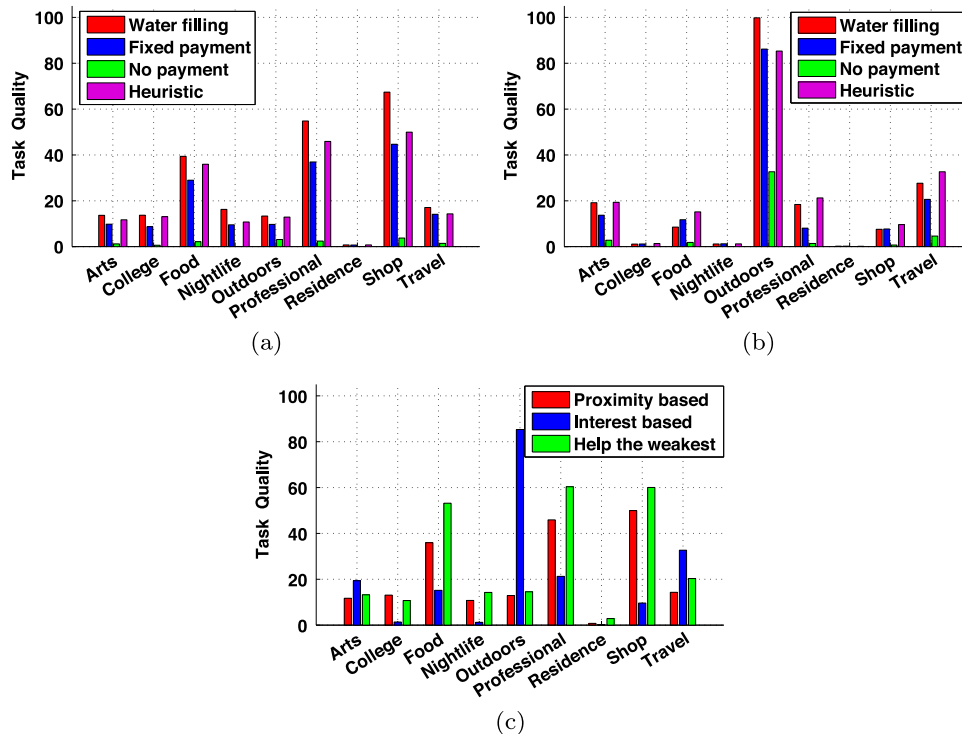


Fig. 8. Quality scores for tasks of different thematic categories under the (a) proximity-based rule, (b) interest-based rule and (c) all the task assignment rules, for  $K=1$  and  $B=200$ .

task (Kittur et al., 2008; Mason and Suri, 2012), how to select crowdworkers (Mashhadi and Capra, 2011), and how to assess the quality of crowdsourced works (Hsueh et al., 2009).

As opposed to online crowdsourcing, situated crowdsourcing deals with tasks that need to be executed offline (in situ) (Hosio et al., 2014). It has been applied to the estimation of queue size (Goncalves et al., 2015), the grading of exams (Heimerl et al., 2012), opinion polling (Hosio et al., 2015), and the gathering of collective emotions (Goncalves et al., 2014).

A number of mobile crowdsourcing platforms have emerged recently (e.g., TaskRabbit, GigWalk, and OpenStreetMap), and research in the area has followed (Goodchild, 2007; Sheppard et al., 2014; Teodoro et al., 2014). Many of those systems suffer from low geographic coverage (Haklay, 2010; Quattrone et al., 2014) and uneven spatial distribution of contributions (Mashhadi et al., 2013; Teodoro et al., 2014).

### 5.2. Task-to-crowdworker assignment

One of the initial steps in crowdsourcing is to assign tasks to crowdworkers. Several strategies to do so have been recently proposed. The assumption is often that a central authority with global knowledge coordinates the assignment. Under that assumption, Kazemi and Shahabi (2012) maximized the number of assigned tasks; Li et al. (2015) minimized the cost of crowd-workers; Ho et al. (2013) matched the maximum number of tasks crowd-workers are willing to perform; and Shirani-Mehr et al. (2009) broke the assignment into two steps: in the first step, the set of points in the city is selected, and, in the second step, the selected points are assigned to users (users submit information to the platform about their starting point, destination and maximum available time to spend, and the platform assigns users to the optimal paths).

Another way of assigning tasks is to account for the abilities and skills of crowd-workers. Reddy et al. (2010) proposed a recruitment framework to identify skilled users. The selection of users is done based on their past whereabouts and availability. In a similar way, our work has used the number of positive feedbacks (i.e., picture favourites) to identify skilled contributors. This is a conservative choice. To see why, consider that previous work (Schifanella et al., 2015) has found not only that, as one would expect, quality and popularity in Flickr pictures are correlated on average (rank correlation 0.4), but also that pictures with many favourites were of quality. Nothing could be said about pictures with no favourite: they were of high quality as much as of low quality. Therefore, considering only the contributors with a considerable number of favourites is a conservative approach.

### 5.3. Incentive allocation

After assigning tasks to crowd-workers, those workers need to be incentivized somehow to carry out those tasks (Restuccia et al., 2015). Under the assumption of a centralized platform, traditional auction

models have been applied to mobile crowdsourcing (Lee and Hoh, 2010; Jaimes et al., 2012; Luo et al., 2014), in which workers bid for their prices. Then, as a further incentive, Peng et al. (2015) proposed to pay workers depending on the estimated quality of their contributions.

There are a few works that combine those two steps: task-to-crowdworker assignment, and incentive allocation. In Gao et al. (2015)'s work, the platform selects a set of workers, and those workers are compensated based on a Lyapunov based VGC auction policy. In He et al. (2014)'s proposal, instead, the service provider selects a set of workers based on estimated quality and willingness to travel, and those workers bargain their compensation with the platform.

Our work has proposed a new framework that combines task-to-crowdworker assignment and incentive allocation, and it does so to optimally allocate tasks while ensuring high geographic coverage and high task quality.

## 6. Conclusion

We have proposed a framework that incentivizes social media users to perform mobile crowdsourcing tasks. First, user information is extracted from publicly available social media profiles. Then, this information guides the assignment of tasks to users who, in turn, receive monetary incentives to complete the tasks. Algorithms for the assignments of tasks and payments are proposed under different application scenarios, i.e., with or without complete spatio-temporal information about potential contributors, with take-it-or-leave-it offers or offers with multiple alternatives. Upon real datasets, our evaluation has shown that, compared to schemes unaware of user skills, those algorithms significantly enhance contribution quality.

There are several ways this work could be extended. First, one could explore alternative ways of modeling the likelihoods that workers will complete their tasks. For example, tasks could be broken down into attributes, and those likelihoods will depend on the extent to which the composing attributes match each user's needs (Karaliopoulos et al., 2012). Second, one could explore additional objective functions. Ours maximized the aggregate expected quality and, as such, is a "social welfare" type of function; alternatively, one could minimize payments to workers given some quality guarantees. Finally, one could produce further quantitative results about the applicability of our framework to platforms other than Flickr. In theory, any platform that comes with quality scores related to geo-referenced contributions could be fit for purpose. In practice, we expect that geographically salient (e.g., Facebook Places) and hyper-local platforms (e.g., TripAdvisor) might lend themselves to our analyses more than what generic online platforms (e.g., Twitter) might do.

## Acknowledgements

P. Micholia, M. Karaliopoulos and I. Koutsopoulos acknowledge the support of the European Commission through the Horizon 2020 project netCommons (Contract number 688768, duration 2016-2018).

### A.1. Derivation of payments under the water filling algorithm (5)

If  $\lambda$  is the Lagrange multiplier for the equality constraint in (5) and  $\mu_i$ ,  $i = 1, \dots, N$  are the KKT multipliers for the inequality constraints, the Lagrangian function is written as

$$L(\mathbf{p}, \bar{\lambda}, \bar{\mu}) = - \sum_{i=1}^N q_i (1 - e^{-\gamma_i^{\alpha_{ij} + \gamma_i^p p_{ij}}}) + \lambda_j \left( \sum_{i=1}^N p_{ij} - B_j \right) - \sum_{i=1}^N \mu_i p_{ij} \quad (\text{A.1})$$

with  $\lambda \geq 0$  and  $\mu_i \geq 0$ .

The first-order necessary KKT conditions for the existence of an optimum require that, for every  $i$ :



$$\nabla_{\mathbf{p}} L(\mathbf{p}, \bar{\lambda}, \bar{\mu}) = 0 \Rightarrow \frac{\partial L(\mathbf{p}, \bar{\lambda}, \bar{\mu})}{\partial p_{ij}} = 0 \Rightarrow -q_i \gamma_i^p e^{-(\gamma_i^\alpha \alpha_{ij} + \gamma_i^p p_{ij})} + \lambda_j - \mu_i = 0, \quad (\text{A.2})$$

and

$$\mu_i p_{ij} = 0. \quad (\text{A.3})$$

Solving for  $\mu_i$ , we get:

$$\mu_i = \lambda_j - q_i \gamma_i^p e^{-(\gamma_i^\alpha \alpha_{ij} + \gamma_i^p p_{ij})}. \quad (\text{A.4})$$

Eq. (A.3) implies three possible combinations of the values of  $p_{ij}$  and  $\mu_i$ , which leads to as many possible cases for Eq. (A.4):

1.  $p_{ij} > 0$  and  $\mu_i = 0$ . This implies that

$$\lambda_j = q_i \gamma_i^p e^{-(\gamma_i^\alpha \alpha_{ij} + \gamma_i^p p_{ij})} \Rightarrow p_{ij} = \frac{1}{\gamma_i^p} \left[ \ln \frac{q_i \gamma_i^p}{\lambda_j} - \gamma_i^\alpha \alpha_{ij} \right]$$

Since  $p_{ij} > 0$ , it also holds that  $\lambda_j < q_i \gamma_i^p e^{-\gamma_i^\alpha \alpha_{ij}}$ .

2.  $p_{ij} = 0$  and  $\mu_i > 0$ . This implies  $\lambda_j > q_i \gamma_i^p e^{-\gamma_i^\alpha \alpha_{ij}}$ .

3.  $\lambda_j = q_i \gamma_i^p e^{-\gamma_i^\alpha \alpha_{ij}}$ . In that case, replacing in (A.3), we obtain that  $p_{ij} = 0$ .

The combination of those three cases yields the formulation in (5).

## References

- Amintoosi, H., Kanhere, S.S., 2014. A reputation framework for social participatory sensing systems. *Mob. Netw. Appl.* 19 (February (1)), 88–100. <http://dx.doi.org/10.1007/s11036-013-0455-x>.
- Avriel, M., 2003. *Nonlinear Programming: Analysis and Methods*. Dover Books on Computer Science Series. Dover Publications, New York.
- Bernstein, M.S., Little, G., Miller, R.C., Hartmann, B., Ackerman, M.S., Karger, D.R., Crowell, D., Panovich, K., 2010. Soyent: a word processor with a crowd inside. In: *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology*. UIST '10. ACM, New York, NY, USA, pp. 313–322. <http://dx.doi.org/10.1145/1866029.1866078>.
- Callison-Burch, C., 2009. Fast, cheap, and creative: evaluating translation quality using amazon's mechanical turk. In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*. EMNLP '09. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 286–295. URL (<http://dl.acm.org/citation.cfm?id=1699510.1699548>).
- Gao, L., Hou, F., Huang, J., 2015. Providing long-term participation incentive in participatory sensing. *CoRR abs/1501.02480*. <http://arXiv.org/abs/1501.02480>.
- Goncalves, J., Kukka, H., Sánchez, I., Kostakos, V., 2016. Crowdsourcing queue estimations in situ. In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. CSCW '16. ACM, New York, NY, USA, pp. 1040–1051. <http://dx.doi.org/10.1145/2818048.2819997>.
- Goncalves, J., Pandab, P., Ferreira, D., Ghahramani, M., Zhao, G., Kostakos, V., 2014. Projective testing of diurnal collective emotion. In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '14. ACM, New York, NY, USA, pp. 487–497. <http://dx.doi.org/10.1145/2632048.2636067>.
- Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal* 69 (4), 211–221. <http://dx.doi.org/10.1007/s10708-007-9111-y>.
- Haklay, M., 2010. How good is volunteered geographical information? A comparative study of openstreetmap and ordnance survey datasets. *Environ. Plan. B: Plan. Des.* 37 (4), 682–703, URL (<http://epb.sagepub.com/content/37/4/682.abstract>).
- He, S., Shin, D.H., Zhang, J., Chen, J., 2014. Toward optimal allocation of location dependent tasks in crowdsensing. In: *IEEE Conference on Computer Communications*. INFOCOM '14, pp. 745–753, April.
- Heimerl, K., Gawalt, B., Chen, K., Parikh, T., Hartmann, B., 2012. Communitysourcing: engaging local crowds to perform expert work via physical kiosks. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. ACM, New York, NY, USA, pp. 1539–1548. <http://dx.doi.org/10.1145/2207676.2208619>.
- Ho, C.-J., Jabbari, S., Vaughan, J.W., 2013. Adaptive task assignment for crowdsourced classification. In: *Proceedings of the 30th International Conference on Machine Learning*, JMLR Workshop. ICML'13, vol. 28. JMLR.org, pp. 534–542.
- Hosio, S., Goncalves, J., Kostakos, V., Riecki, J., 2015. Crowdsourcing public opinion using urban pervasive technologies: lessons from real-life experiments in Oulu. *Policy Internet* 7 (2), 203–222. <http://dx.doi.org/10.1002/poi3.90>.
- Hosio, S., Goncalves, J., Lehdonvirta, V., Ferreira, D., Kostakos, V., 2014. Situated crowdsourcing using a market model. In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. UIST '14. ACM, New York, NY, USA, pp. 55–64. <http://dx.doi.org/10.1145/2642918.2647362>.
- Hsueh, P.-Y., Melville, P., Sindhvani, V., 2009. Data quality from crowdsourcing: a study of annotation selection criteria. In: *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing*. HLT '09. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 27–35. (<http://dl.acm.org/citation.cfm?id=1564131.1564137>).
- Irani, L.C., Silberman, M.S., 2013. Turkopticon: interrupting worker invisibility in amazon mechanical turk. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI'13. ACM, New York, NY, USA, pp. 611–620. <http://dx.doi.org/10.1145/2470654.2470742>.
- Jaimes, L., Vergara-Laurens, I., Labrador, M., 2012. A location-based incentive mechanism for participatory sensing systems with budget constraints. In: *IEEE International Conference on Pervasive Computing and Communications*. PerCom'12, pp. 103–108, March.
- Karaliopoulos, M., Koutsopoulos, I., Titsias, M., 2016. First learn then earn: optimizing mobile crowdsensing campaigns through data-driven user profiling. In: *Proceedings of 17th ACM MobiHoc*, pp. 271–280, July.
- Kazemi, L., Shahabi, C., 2012. Geocrowd: enabling query answering with spatial crowdsourcing. In: *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*. SIGSPATIAL '12. ACM, New York, NY, USA, pp. 189–198. <http://dx.doi.org/10.1145/2424321.2424346>.
- Kittur, A., Chi, E.H., Suh, B., 2008. Crowdsourcing user studies with mechanical turk. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI'08. ACM, New York, NY, USA, pp. 453–456. <http://dx.doi.org/10.1145/1357054.1357127>.
- Kittur, A., Smus, B., Khamkar, S., Kraut, R.E., 2011. Crowdforge: crowdsourcing complex work. In: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. UIST '11. ACM, New York, NY, USA, pp. 43–52. <http://dx.doi.org/10.1145/2047196.2047202>.
- Kosinski, M., Stillwell, D., Graepel, T., 2013. Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. U.S.A.* 110 (April(15)) 5802–5805.
- Krontiris, I., Freiling, F.C., 2010. Urban sensing through social networks: the tension between participation and privacy. In: *International Tyrrhenian Workshop on Digital Communications*. ITWDC'10.
- Lee, J.-S., Hoh, B., 2010. Dynamic pricing incentive for participatory sensing. *Pervas. Mob. Comput.* 6 (December (6)), 693–708. <http://dx.doi.org/10.1016/j.pmcj.2010.08.006>.
- Li, H., Li, T., Wang, Y., Oct 2015. Dynamic participant recruitment of mobile crowd sensing for heterogeneous sensing tasks. In: *IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems*. MASS'15, pp. 136–144.
- Little, G., Sep 2009. Turkit: tools for iterative tasks on mechanical turk. In: *IEEE Symposium on Visual Languages and Human-Centric Computing*. VL/HCC'09, pp. 252–253.
- Luo, T., Kanhere, S., Das, S., Tan, H.-P., 2014. Optimal prizes for all-pay contests in heterogeneous crowdsourcing. In: *IEEE 11th International Conference on Mobile Ad Hoc and Sensor Systems*. MASS'14, pp. 136–144, October.
- Martello, S., Toth, P., 1990. *Knapsack Problems: Algorithms and Computer Implementations*. John Wiley & Sons, Inc., New York, NY, USA.
- Mashhadi, A., Quattrone, G., Capra, L., 2013. Putting ubiquitous crowd-sourcing into context. In: *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. CSCW '13. ACM, New York, NY, USA, pp. 611–622. <http://dx.doi.org/10.1145/2441776.2441845>.
- Mashhadi, A.J., Capra, L., 2011. Quality control for real-time ubiquitous crowdsourcing. In: *Proceedings of the 2nd International Workshop on Ubiquitous Crowdsourcing*. UbiCrowd '11. ACM, New York, NY, USA, pp. 5–8. <http://dx.doi.org/10.1145/2030100.2030103>.
- Mason, W., Suri, S., 2012. Conducting behavioral research on amazon's mechanical turk. *Behav. Res. Methods* 44 (1), 1–23. <http://dx.doi.org/10.3758/s13428-011-0124-6>.
- Palomar, D.P., Fonollosa, J.R., 2005. Practical algorithms for a family of waterfilling

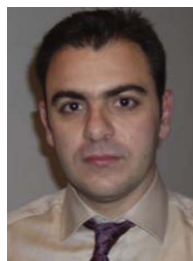
- solutions. *IEEE Trans. Signal Process.* 53 (2-1), 686-695, URL (<http://dblp.uni-trier.de/db/journals/tsp/tsp53.html#PalomarF05>).
- Peng, D., Wu, F., Chen, G., 2015. Pay as how well you do: a quality based incentive mechanism for crowdsensing. In: *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*. MobiHoc '15. ACM, New York, NY, USA, pp. 177-186. URL <http://doi.acm.org/10.1145/2746285.2746306>.
- Quattrone, G., Mashhadi, A., Capra, L., 2014. Mind the map: the impact of culture and economic affluence on crowd-mapping behaviours. In: *CSCW: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, New York, NY, USA, pp. 934-944. <http://dx.doi.org/10.1145/2531602.2531713>.
- Rashtchian, C., Young, P., Hodosh, M., Hockenmaier, J., 2010. Collecting image annotations using amazon's mechanical turk. In: *CSLDAMT: Proceedings of the NAACL HLT Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 139-147. URL (<http://dl.acm.org/citation.cfm?id=1866696.1866717>).
- Reddy, S., Estrin, D., Srivastava, M., 2010. Recruitment framework for participatory sensing data collections. In: *Proceedings of the 8th International Conference on Pervasive Computing*. Pervasive'10. Springer-Verlag, Berlin, Heidelberg, pp. 138-155. [http://dx.doi.org/10.1007/978-3-642-12654-3\\_9](http://dx.doi.org/10.1007/978-3-642-12654-3_9)
- Restuccia, F., Das, S.K., Payton, J., 2015. Incentive mechanisms for participatory sensing: survey and research challenges. *CoRR abs/1502.07687*. [arXiv:1502.07687](https://arxiv.org/abs/1502.07687).
- Schifanella, R., Redi, M., Aiello, L.M., 2015. An image is worth more than a thousand favorites: surfacing the hidden beauty of flickr pictures. In: *Proceedings of the 9th AAAI International Conference on Weblogs and Social Media*. ICWSM'15. AAAI. URL (<http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10547>).
- Sheppard, S.A., Wiggins, A., Terveen, L., 2014. Capturing quality: retaining provenance for curated volunteer monitoring data. In: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. CSCW '14. ACM, New York, NY, USA, pp. 1234-1245. <http://dx.doi.org/10.1145/2531602.2531689>.
- Shirani-Mehr, H., Banaei-Kashani, F., Shahabi, C., 2009. Efficient viewpoint assignment for urban texture documentation. In: *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. GIS '09. ACM, New York, NY, USA, pp. 62-71. <http://dx.doi.org/10.1145/1653771.1653783>.
- Teodoro, R., Ozturk, P., Naaman, M., Mason, W., Lindqvist, J., 2014. The motivations and experiences of the on-demand mobile workforce. In: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. CSCW '14. ACM, New York, NY, USA, pp. 236-247. <http://dx.doi.org/10.1145/2531602.2531680>.



**Panagiota Micholia** received the Diploma degree in Electrical and Computer Engineering from the Democritus University of Thrace (DUTH), Greece, in 2010 and the M.S. degree in Advanced Sciences of Modern Telecommunications from the University of Valencia (UV), Spain, in 2012. She is currently a Ph.D. candidate at the Department of Informatics, Athens University of Economics and Business (AUEB), Athens, Greece. Her research interests focus on user recruitment and incentives in crowdsourcing systems and community networks.



**Merkouris Karaliopoulos** is a Senior Research Associate at the Athens University of Economics and Business, in Greece. He obtained his Ph.D. degree in Electronic Engineering from the University of Surrey, UK (2004). He has been a Postdoctoral researcher at the University of North Carolina at Chapel Hill (2005-2006), and a Senior Researcher and Lecturer in ETH Zurich (2007-2010). Prior joining AUEB, he was a Marie-Curie Fellow with the University of Athens from 2010-2012 and a Researcher with the Center of Research and Technology Hellas, Greece (2013-2015). His research interests lie in the broader area of wireless and mobile social networks.



**Iordanis Koutsopoulos** received the Diploma degree in Electrical and Computer Engineering from the National Technical University of Athens (NTUA), Greece, and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Maryland, College Park (UMCP), USA. He is now an Associate Professor at the Department of Informatics, Athens University of Economics and Business (AUEB). He received the single-investigator European Research Council (ERC) competition runner-up award for the project RECITAL: Resource Management for Self-coordinated Autonomic Wireless Networks. His research interests involve network control and optimization in wireless networks, social and community networks, crowd-sensing systems, smart-grid and cloud computing.



**Luca Maria Aiello** is Senior Research Scientist in the Social Dynamics group at Bell Labs in Cambridge, UK. He received his PhD in Computer Science from the University of Torino in 2012 and he has been a Research Scientist at Yahoo Labs for 4.5 years. He conducts interdisciplinary research in network science, computational social science, and urban informatics. He co-authored 50+ peer-reviewed papers and his work has been covered by more than 200 news articles published by prestigious news outlets worldwide including Wired, Wall Street Journal, and BBC. He serves in the PCs of multiple conferences including WWW, WSDM, ICWSM, and CIKM. He organized several academic events: organizer of the SNOW workshop at WWW, workshop and tutorial chair for ICWSM in 2015-16, and general chair of SocInfo in 2014. He led the work of Yahoo in the European project "SocialSensor". He is among the founding members of GoodCityLife.org.



**Gianmarco De Francisci Morales** is a Scientist at QCRI. Previously he was a Visiting Scientist at Aalto University, a Research Scientist at Yahoo Labs, and a Research Associate at ISTI-CNR. He received his Ph.D. in Computer Science and Engineering from the IMT Institute for Advanced Studies of Lucca, in 2012. His research focuses on scalable data mining, and data-intensive scalable computing systems. He is a member of the Apache Software Foundation, a committer for Apache Pig, and a lead developer of Apache SAMOA, a platform for mining big data streams. He serves on the PC of several conferences, including WSDM, KDD, CIKM, and WWW.



**Daniele Quercia** leads the Social Dynamics group at Bell Labs in Cambridge (UK). He has been named one of Fortune magazine's 2014 Data All-Stars, and spoke about happy maps at TED. His research has been focusing on the area of urban informatics and received best paper awards from Ubicomp 2014 and from ICWSM 2015, and an honourable mention from ICWSM 2013. He was a Research Scientist at Yahoo Labs, a Horizon Senior Researcher at the University of Cambridge, and Postdoctoral Associate at the Department of Urban Studies and Planning at MIT. He received his Ph.D. from UC London. His thesis was sponsored by Microsoft Research and was nominated for BCS Best British Ph.D. dissertation in Computer Science.