

Spatial Crowdsourcing: A Survey

Yongxin Tong · Zimu Zhou · Yuxiang Zeng · Lei Chen · Cyrus Shahabi

Received: date / Accepted: date

Abstract Crowdsourcing is a computing paradigm where humans are actively involved in a computing task, especially for tasks that are intrinsically easier for humans than for computers. Spatial crowdsourcing (SC) is an increasing popular category of crowdsourcing in the era of mobile Internet and sharing economy, where tasks are spatiotemporal and must be completed at a specific location and time. In fact, spatial crowdsourcing has stimulated a series of recent industrial successes including sharing economy for urban services (Uber and Gigwalk) and spatiotemporal data collection (OpenStreetMap and Waze).

This survey dives deep into the challenges and techniques brought by the unique characteristics of spatial

crowdsourcing. Particularly, we identify four core algorithmic issues in spatial crowdsourcing: (1) task assignment, (2) quality control, (3) incentive mechanism design and (4) privacy protection. We conduct a comprehensive and systematic review of existing research on the aforementioned four issues. We also analyze representative spatial crowdsourcing applications and explain how they are enabled by these four technical issues. Finally, we discuss open questions that need to be addressed for future spatial crowdsourcing research and applications.

Keywords Spatial crowdsourcing · Task assignment · Quality control · Incentive mechanism · Privacy protection

Y. Tong

State Key Laboratory of Software Development Environment,
Beijing Advanced Innovation Center for Big Data and Brain
Computing and International Research Institute for Multi-
disciplinary Science, Beihang University, Beijing, China
E-mail: yxtong@buaa.edu.cn

Z. Zhou

Computer Engineering and Networks Laboratory, ETH
Zurich, Zurich, Switzerland
E-mail: zzhou@tik.ee.ethz.ch

Y. Zeng

Department of Computer Science and Engineering, The Hong
Kong University of Science and Technology, Clear Water Bay,
Kowloon, Hong Kong SAR, China
E-mail: yzengal@cse.ust.hk

L. Chen

Department of Computer Science and Engineering, The Hong
Kong University of Science and Technology, Clear Water Bay,
Kowloon, Hong Kong SAR, China
E-mail: leichen@cse.ust.hk

C. Shahabi

Department of Computer Science, University of Southern
California, California, USA
E-mail: shahabi@usc.edu

1 Introduction

Crowdsourcing is a computing paradigm where humans actively or passively participate in the procedure of computing, especially for tasks that are intrinsically easier for humans than for computers. It has attracted extensive attention from both the academia and the industry [59, 69, 99, 141], and there have been many successful crowdsourcing platforms such as Amazon Mechanical Turk (MTurk) [2] and Upwork [28].

With the development of mobile Internet and sharing economy, traditional web-based crowdsourcing has shifted to spatial crowdsourcing¹ (*a.k.a.* mobile crowdsourcing) [57, 132, 207]. As with traditional crowdsourcing, spatial crowdsourcing involves three components, *tasks*, *workers* and *the platform*. Fig. 1 shows the typical workflow of spatial crowdsourcing. The roles of these component are as follows.

¹ The term was coined for the first time in [132].

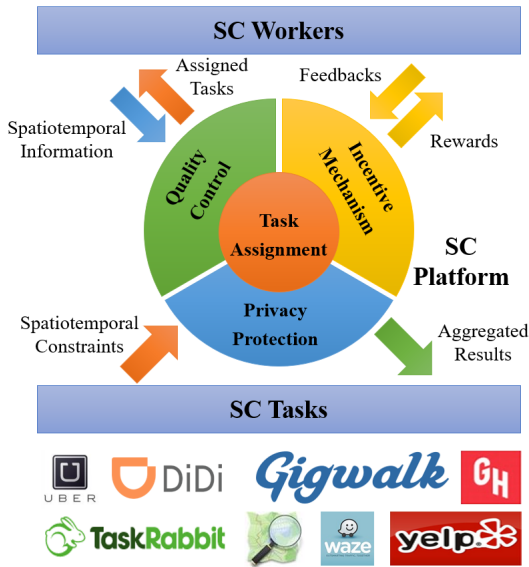


Fig. 1: Key components and workflow in spatial crowdsourcing.

- **Tasks.** Tasks with spatiotemporal constraints (*e.g.*, the positions and deadlines of tasks) are submitted to the platform. To complete a task, a worker has to physically move to the position of the task.
- **Workers.** Workers submit their spatiotemporal information such as their positions and deadlines to the platform. Depending on the concrete applications, workers either are assigned to tasks or can choose tasks by themselves.
- **The Platform.** The spatial crowdsourcing platform (platform for short) connects tasks and workers. Its core functions include assigning tasks to suitable works, aggregating the results submitted by workers, setting rewards for workers and protecting the privacy of the tasks and workers.

The major difference between spatial crowdsourcing and web-based crowdsourcing is that the former requires each worker to move in the physical world to perform tasks [132]. Hence spatiotemporal information such as location, mobility and the associated contexts plays a crucial role. Its natural connection with the physical world makes spatial crowdsourcing a computing paradigm for a wide spectrum of daily applications including real-time ride-hailing services, *e.g.*, Uber [17] and DiDi Chuxing [4], product placement checking supermarkets, *e.g.*, Gigwalk [8] and TaskRabbit [15], on-wheel meal-ordering services, *e.g.*, GrubHub [10] and Meituan [26], and citizen sensing services, *e.g.*, OpenStreetMap [13] and Waze [19]².

² Sometimes Waze is also viewed as a crowdsensing application, which leverages users' sensor-equipped mobile devices

The emphasis on spatiotemporal dynamics calls for new designs in crowdsourcing theories and systems. The aim of this survey is to provide a comprehensive review on the core algorithmic issues in spatial crowdsourcing from the perspective of the platform.

Task Assignment. In practice, a spatial crowdsourcing platform needs to manage massive tasks and workers every day. For example, in 2017, DiDi Chuxing needs to serve 25 million ride requests every day with the registered over 21 million drivers, which eventually produces over 70TB spatiotemporal data every day [21]. Thus, the first challenge of the spatial crowdsourcing platforms is how to assign the large-scale tasks to their workers, *i.e.*, task assignment. The platforms usually aims to arrange the tasks to suitable workers with different optimization objectives such as maximizing the total number of assigned tasks or the total payoff of the tasks to their assigned workers, minimizing the total travel cost of the allocated workers.

Quality Control. As with most crowdsourcing applications, results collected from workers in spatial crowdsourcing vary in quality. The aim of quality control is to quantify the quality of workers and tasks and effectively aggregate results to ensure high-quality task completion. Both the quality models and the aggregation techniques are tied to spatiotemporal information, which imposes unique challenges.

Incentive Mechanism. Proper Incentive mechanisms help attract workers to participate in spatial crowdsourcing. Dedicated incentive mechanism design is needed because the spatiotemporal factors and the relative relation between supply and demand in spatial crowdsourcing are dynamic. For example, if there are only a few workers in some area, the tasks posted in this area should provide more reward.

Privacy Protection. Privacy protection is particularly crucial in spatial crowdsourcing. Spatiotemporal information of workers, tasks and intermediate results needs to be properly transformed to avoid privacy leakage while allowing efficient information processing such as task assignment. Dedicated techniques and frameworks need to be designed to balance between the strength of privacy protection and the efficiency of other spatial crowdsourcing operations.

Contributions over Existing Surveys. There are some general surveys [34, 69, 99, 141] or tutorials [58, 59, 142] on traditional web-based crowdsourcing. Our survey focuses on the spatiotemporal factors and the new algorithmic designs on crowdsourcing due to these factors. There are also some surveys or tutorials

to collect and share data. Spatial crowdsourcing is a general framework and can subsume crowdsensing or participatory sensing [132].

Table 1: A time-line of milestone papers of spatial crowdsourcing.

Year	Reference	Influence
2012	[132]	First work of spatial crowdsourcing
2013	[78]	First work of static task matching (see Sec. 3.3) in spatial crowdsourcing
2013	[133]	First work of quality control (see Sec. 4) in spatial crowdsourcing
2014	[197]	First work of privacy protection (see Sec. 6) in spatial crowdsourcing
2014	[63]	First work of general spatial crowdsourcing platform
2015	[144]	First work of dynamic task planning (see Sec. 3.6) in spatial crowdsourcing
2016	[206]	First work of dynamic task matching (see Sec. 3.4) in spatial crowdsourcing
2016	[205]	First experimental work of dynamic task matching in spatial crowdsourcing
2018	[211]	First work of incentive mechanism (see Sec. 5) in spatial crowdsourcing
2018	[202]	First work of privacy protection in dynamic scenario

on spatial crowdsourcing. For example, Guo *et al.* [104] and Tong *et al.* [204] review task allocation of spatial crowdsourcing; To *et al.* review privacy protection of spatial crowdsourcing in Chapter 7 of [196]; Zhang *et al.* [240] review the incentive mechanisms in spatial crowdsourcing; Zhao *et al.* [245] give a brief survey on spatial crowdsourcing, which only sketches out a few representative works. Compared with [104, 196, 240, 245], we provide a comprehensive and holistic review on the latest progress on spatial crowdsourcing research. Chen *et al.* [57] also conduct a survey on spatial crowdsourcing. However, our work is more systematic in classifying the techniques and also covers the most recent literature in the last three years. Tong *et al.* give a tutorial on spatial crowdsourcing in [207]. This survey is its holistic and systematic extension and update.

Bibliography Methodology. We select papers primarily from top venues in the database communities such as SIGMOD, VLDB, ICDE and TKDE. We also include some representative works from the spatial and mobile computing communities since some important algorithmic issues in spatial crowdsourcing also stemmed from there (although as the topic of crowdsensing, which has a slightly different focus). In Table 1 we list the milestone papers during the development of spatial crowdsourcing and their influence on this research area.

In the rest of this survey, we first present the preliminaries in Sec. 2 and review the representative research on the four core issues in spatial crowdsourcing in Sec. 3 to Sec. 6. We then study some killer applications of spatial crowdsourcing in Sec. 7 and discuss future challenges and opportunities in Sec. 8. Finally we conclude in Sec. 9.

2 Preliminaries

This section introduces the models of tasks, the models of workers and the practical constraints that will be frequently used in this survey.

2.1 Task Modeling

In spatial crowdsourcing, a task is also known as a spatial task [132], a crowdsourced task [97], a spatial crowdsourced task [96], or a request [154]. The user, who submits the task on such platforms, is called task requester [187] or requester [132]. In real-world applications, a task can be a taxi calling request in ride sharing platform (*e.g.*, Uber [17] and Didi Chuxing [4]), a take-out order in food delivery platform (*e.g.*, GrubHub [10] and Seamless [27]), a last-mile delivery request in urban logistic platform (*e.g.*, UPS [18] and FedEx [6]), and other general tasks like taking photos of landmarks and appliance repairment in Gigwalk [8] and TaskRabbit [15]. For example, the number of food delivery orders has been increased to 10 billion in China by the end of 2017 [227]. The main reason is that crowdsourcing these tasks can usually result in higher quality task completion (*e.g.*, low latency) at a lower cost due to the large scale of workers.

After receiving the task issued by the requester, the platform will know the following major information about this task.

- **Arrival time** indicates when the task is submitted.
- **Location** represents the spatial information of the task. Some task (*e.g.*, a taxi calling request or a food delivery order) contains two types of locations, **origin** (pickup location) and **destination** (delivery location). To complete such a task, a worker needs to first come to the origin and then take to the destination.
- **Deadline** represents the expired time of the task.
- **Radius** restricts a circular range whose center is the location of the task.
- **Reward** is the payoff to the worker if he/she completes the task. The amount of reward is either directly decided by the requester or determined by the platform based on its incentive mechanisms.

A few other attributes of tasks are also considered in some studies, *e.g.*, required skills [66] (the requirement

Table 2: A summary about the attributes of tasks and workers used in the four core issues.

	Task					Worker				
	Arrival Time	Location	Deadline	Radius	Reward	Arrival Time	Location	Deadline	Radius	Capacity
Task Assignment	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Quality Control	✓	✓	✓	✗	✗	✓	✓	✗	✓	✓
Incentive Mechanism	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Privacy Protection	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗

of skills to perform the task), arrival rate [84] (the probability of appearance in a unit time), etc.

Similar to crowdsourcing [141], the tasks in spatial crowdsourcing can be also classified into two kinds in terms of granularity: *i.e.*, *macro-tasks* and *micro-tasks*. A macro-task in spatial crowdsourcing often involves a wider space and requires more time to complete. In contrast, a micro-task in spatial crowdsourcing usually involves much fewer locations and needs less time to complete. For example, mapping a city belongs to a macro-task whereas geotagging a landmark of this city is a micro-task. As most existing studies in spatial crowdsourcing focus on micro-tasks, this survey also mainly restricts to the scope of micro-tasks and only briefly discusses the issues of task assignment, quality control and incentive mechanism in macro-tasks.

2.2 Worker Modeling

In spatial crowdsourcing, a worker is also known as a spatial worker [33], a crowd worker [206], a mobile worker [112], a service provider [205], or an agent [55]. To join the platform and perform tasks, a worker usually shares his/her spatiotemporal information with the platform. The commonly used attributes include:

- **Arrival time** indicates when the worker appears on the platform.
- **Location** is the spatial information of the worker.
- **Deadline** restricts the leaving time of the worker.
- **Radius** represents a circular range whose center is the location of the worker.
- **Capacity** is the maximum number of tasks that he/she can perform before the deadline.

From historical data, the platform will also know the acceptance ratio of the worker [238] (the percentage of accepted ones among all the assigned tasks) and the reputation of the worker [108, 219]. Some works also consider a few other attributes of workers, *e.g.*, his/her skills [98, 188], travel budget [97, 246], etc.

Table 2 summarizes the attributes of tasks and workers used in the four core issues in spatial crowdsourcing that we will discuss in the subsequent sections.

2.3 Practical Constraints

The main characteristic of spatial crowdsourcing is the *spatial* factors (*e.g.*, location) and *temporal* factors (*e.g.*, deadline). These factors are important when the platform makes task assignment, controls the quality, designs the incentive mechanism and protects the privacy. Thus, existing works usually consider three types of constraints to satisfy the dynamics in spatial crowdsourcing, *i.e.*, **spatial constraints**, **temporal constraints** and **other constraints**. We list the major ones as follows.

Spatial Constraints.

- **Range constraint**: the task assigned to a worker is within his/her restricted range; the worker assigned to a task is within its restricted range.
- **Travel budget constraint**: the total travel cost of the worker should be under his/her travel budget.

Temporal Constraints.

- **Deadline constraint**: the task will expire after the corresponding deadline; the worker will leave the platform after his/her deadline.
- **Real-time constraint** (*a.k.a.* instantaneous constraint): once a task appears, a worker must be assigned to it before the next task appears.

Other Constraints.

- **Capacity constraint**: the number of tasks assigned to a worker cannot exceed his/her capacity.
- **Invariable constraint**: once a task is assigned to a worker, the allocation between the task and the worker cannot be changed.
- **Reward budget constraint**: the total payoff to the assigned workers should be under the reward budget of the requester.

- **Skill constraint:** each required skills of a task is covered by the skills of at least one worker.
- **Reliability constraint:** the probability of the task being performed correctly should be larger than a threshold of reliability.

Some of these constraints are more widely used in all four core issues, *e.g.*, range constraint, deadline constraint and capacity constraint. Some other constraints are only used in specific scenario, *e.g.*, skill constraint is often used when a task has a specific requirement about the skills of the assigned workers.

3 Task Assignment

Task assignment is considered as the most fundamental challenge in spatial crowdsourcing [57, 207]. This is because all the other core issues in spatial crowdsourcing are connected with task assignment, as we will discuss in the next several sections. In this section, we first define the task assignment problem (Sec. 3.1) and then categorize existing research from two dimensions (Sec. 3.2): the arrival of input (*i.e.*, static or dynamic) and the algorithmic assignment model (*i.e.*, matching or planning). Accordingly, we introduce existing research from four categories: static matching (Sec. 3.3), dynamic matching (Sec. 3.4), static planning (Sec. 3.5), and dynamic planning (Sec. 3.6). Finally, we summarize these studies in Sec. 3.7.

3.1 Generic Definition

Task assignment aims to arrange tasks to suitable workers for different objectives. A generic definition of task assignment in spatial crowdsourcing is as follows.

Given a set of tasks and a set of workers, task assignment refers to the process to make an arrangement between tasks and workers for specific objectives, while satisfying spatial constraints, temporal constraints and (or) other constraints.

In terms of objectives, there are mainly two types of optimization goals: **total utility** and **total cost**. Below are the general definitions of **utility** and the **cost**.

- **Utility.** *It is a value that measures the utility of an assignment between a task and its assigned worker.* Utility can be a constant value of 1, the reward of the task, the acceptance ratio of the worker, or even the reward times the acceptance ratio. Accordingly, the **total utility** will represent the total number of performed tasks [78], the total acceptance ratio of the workers [238], the total rewards of the assigned tasks [243] or the total expected rewards of the assigned tasks [206, 222].

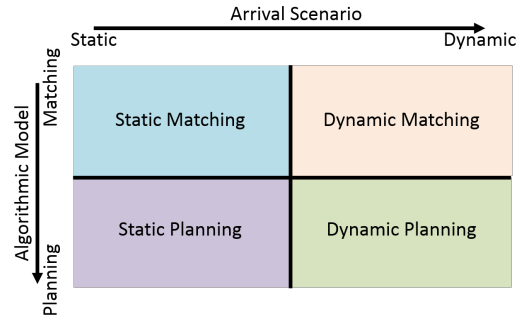


Fig. 2: Matrix overview of task assignment research.

- **Cost.** *It is a value to measure the cost of the assignment between a task and its assigned worker.* Cost can be the travel distance (time) of the worker to the task, or the delay of the task from its arrival time to the completion time. Accordingly, the **total cost** will represent the total travel distance of the workers [205] or the total delay of all the tasks [64].

Stable assignment is another objective for certain spatial crowdsourcing applications [225, 243]. It is motivated by *stable marriage* [94], which aims to minimize the number of *unstable pairs* (*a.k.a.* blocking pairs). A pair of task and worker (except its current assigned worker) is an unstable pair if the following conditions are satisfied: (1) the task prefers the worker more than its current assigned worker; (2) the worker prefers the task more than his/her current assigned task.

3.2 Categories of Existing Research

Fig. 2 summarizes the taxonomy of task assignment research. Existing studies can be categorized from two dimensions: the **arrival scenario**, which can be *static* or *dynamic*; and the **algorithmic model**, which can be *matching* or *planning*.

Arrival Scenario: Static vs. Dynamic.

- **Static.** In the static scenario (*a.k.a.* offline scenario), the platform is assumed to know all the spatiotemporal information of the tasks and workers at the beginning, which includes the arrival times and locations of tasks and workers.
- **Dynamic.** In the dynamic scenario (*a.k.a.* online scenario), the spatiotemporal information of either tasks or workers is only known upon their arrival.

Intuitively, the dynamic scenario is more practical yet more challenging than the static one, since tasks and workers in the dynamic scenario need to be assigned based on only partial information.

Algorithmic Model: Matching vs. Planning.

- **Matching.** In the matching model, task assignment is often formulated as a bipartite graph based problem. Workers and tasks can be represented by the vertices in the bipartite graph and utility or cost between a worker and a task can be denoted by the weight of the edges. Then the problem is to obtain an optimal matching in the bipartite graph.
- **Planning.** In the planning model (*a.k.a.* scheduling model), task assignment aims to plan a route for each worker to perform a sequence of tasks.

Before introducing solutions to each category of task assignment research, we list the evaluation metrics for a task assignment algorithm. In terms of *efficiency* of the algorithm, time and memory costs are used. In terms of *effectiveness*, **approximation ratio** and **competitive ratio** are standard to assess the theoretical guarantee of the offline algorithms and online algorithms, respectively. Specifically, the approximation ratio represent the effectiveness that an *offline* algorithm can guarantee in the worst case. The competitive ratio represents the effectiveness that an *online* algorithm can guarantee but under various analysis models. In the task assignment research in spatial crowdsourcing, the following analysis models are covered:

- **Adversarial Order Model (AO).** It considers the worst case of the online algorithm.
- **Random Order Model (RO).** It considers the average case of the online algorithm, *i.e.*, the arrival order of inputs is uniformly sampled from all possible permutations.
- **I.I.D Model (IID).** It assumes that the dynamically arrived vertices (*i.e.*, workers or tasks) are identical and independent distributed but unknown to the algorithms.
- **Known I.I.D Model (KIID).** It also assumes that workers or tasks are *i.i.d.*, but the algorithm knows the distribution.
- **Known Adversarial Distribution Model (KAD).** It is a generalization of the KIID model. However, each task and worker in this model is sampled according to an arbitrary distribution, which is known to the algorithm. Different from KIID model, the distributions of the tasks (workers) may be different from each other.

Now we review existing studies in four categories: static matching (Sec. 3.3), dynamic matching (Sec. 3.4), static planning (Sec. 3.5), and dynamic planning (Sec. 3.6).

3.3 Static Matching

This subsection reviews research on task matching in the static scenario, where information of workers and

tasks is known before assignment. We discuss existing studies in terms of different objectives, which include utility maximization (Sec. 3.3.1), cost minimization (Sec. 3.3.2), and stable matching (Sec. 3.3.3).

3.3.1 Utility Maximization

In practice, utility can represent the constant value 1 (*i.e.*, the number of assigned task) or the payoff to the worker. Accordingly, the objective of utility maximization is equivalent to either maximizing *the total number of assignments* or *the total payoff*. We discuss existing solutions to these two objectives separately.

Maximizing Total Number of Assignments. Solutions to static matching problem that maximizes the number of assigned tasks are either *exact* or *greedy based* approximate algorithms.

- **Exact.** Since task matching can be formulated as a bipartite graph, the maximum cardinality bipartite matching of the graph yields the assignment with the maximum total number. Hence exact algorithms (*e.g.*, Hungarian algorithm [52]) can optimally solve the problem. Alternatively, Kazemi *et al.* [132] reduce the bipartite graph into an instance of the *maximum flow* problem [32] and use the Ford-Fulkerson algorithm [93] to obtain the exact result. They also consider some practical issues. For example, a task may have fewer workers around and hence it should be assigned with higher priority. Therefore, the authors borrow the idea of location entropy [74] to represent this priority. Location entropy measures the total number of workers near that location as well as the relative proportion of their future visits to that location. Another heuristic strategy is to iteratively assign the task to its nearest worker (*i.e.*, Nearest Neighbor Priority (NNP)).
- **Greedy based.** To reduce the computation cost of exact solutions (*i.e.*, Hungarian [52] and Ford-Fulkerson [93] algorithms), various greedy based methods are proposed. Both [200] and [213] maximize the total number while considering a budget constraint. They extend the idea of location entropy [132] to region entropy, *i.e.*, tasks in the spatial region with fewer workers inside (*i.e.*, less region entropy) should have a higher priority to be assigned. Then they greedily make assignment based on the current highest priority. Alfarrarjeh *et al.* [33] further design several partition based distributed implementations (*e.g.*, Spatial Partitioning Approach (SPA)) to improve the scalability of the solutions.

Maximizing Total Payoff. Solutions to static matching that maximizes the total payoff are also either *exact* or *greedy based* approximate algorithms.

- **Exact.** To *et al.* [198] extend the problem in [132] by assuming that a worker with better performance should be paid more. Accordingly, an instance of static matching can be reduced to an instance of maximum weighted bipartite matching [52]. Thus the Hungarian algorithm [52] can still be utilized to obtain an exact solution. Considering worker distribution and travel cost, they also reduce the original matching problem to the minimum-cost maximum weighted bipartite matching problem.
- **Greedy based.** Similarly, the greedy based algorithms are proposed to improve the efficiency of exact algorithms such as the Hungarian algorithm. She *et al.* [184] consider the conflicts among tasks and propose an approximation solution with ratio $\frac{1}{1+C_{max}}$, where C_{max} is the maximum capacity of workers. Cheng *et al.* [66] study the settings where a task has requirements on the worker’s skills under a budget constraint. Assuming a batch mode [200], they propose a greedy based method and further develop a new algorithm with an adaptive cost model.

3.3.2 Cost Minimization

Since the platform tends to serve more tasks, the actual objective is often to find a matching with maximum cardinality and minimum cost. Hence the corresponding static matching problem can be transferred to the *minimum-cost maximum-flow (MCMF)* problem [32]. The problem can again be solved by exact algorithms such as the Hungarian algorithm [52] and Successive Shortest Path Algorithm (SSPA) [81]. To improve the efficiency, Hou *et al.* [214] leverage indexing and I/O optimization techniques. Specifically, they introduce incremental SSPA-based exact solutions with R-tree indexing. A heuristic algorithm is also designed to achieve better efficiency with an approximate result.

Bei *et al.* [48] study static matching with cost minimization where each worker can be assigned to at most two tasks at any time. The problem is formulated as a variant of the 3-dimensional matching (3DM) [101]. A matching in 3DM is a triple which consists of two tasks and one worker. To solve the problem, they first pack every two tasks and then make an arrangement between the workers and the pairs of tasks after packing. This two-phase based algorithm achieves an approximation ratio of 2.5 when the number of tasks is exactly twice of the number of workers.

Long *et al.* [156] also focuses on the static matching with cost minimization. Differently, they want to find a

matching with maximum cardinality which minimizes the maximum travel cost among all the assignments. They devise a scalable algorithm called Swap-Chain to efficiently get the optimal solution.

3.3.3 Stable Matching

Some other research studies the static stable matching problem in spatial data. Solutions to the stable marriage problem can be applied to static stable matching. For example, the problem can be solved by the Gale-Shapley algorithm [94], which takes $O(|T||W|)$ time when T is the set of tasks and W is the set of workers. Another intuitive solution is to iteratively select the closest pair from the remaining tasks and workers [72, 229]. To improve the efficiency, Wong *et al.* [225] reduce the concept of “mutual nearest neighbor” to the Bichromatic Mutual NN Search problem, and propose an NN search based Chain algorithm with a time complexity of $O((|T| + |W|) \cdot (\log^{O(1)} |T| + \log^{O(1)} |W|))$.

3.3.4 Summary on Static Matching

Table 3 lists the representative studies on static matching. Solutions to the static matching problem are the basis of many other more complex task assignment problems in spatial crowdsourcing. Cheng *et al.* [67] conduct a comprehensive evaluation on mainstream static matching algorithms. They conduct the experiments on the open datasets [9] collected by the gMission and the toolbox called SCAWG [199] for spatial crowdsourcing. According to their experimental results, LLEP [198] is a good choice to maximize the total utility and NNP [132] is closely effective but more efficient.

Besides, solutions to the static matching problem can be also extended for the dynamic scenario via the batch based mode [132, 200]. In the batch mode, an arrangement is made between tasks and workers in a fixed time interval (*i.e.*, a batch). However, the response time of tasks or workers tends to be long in the batch mode. Hence online algorithms directly designed for dynamic matching are desired, which we present below.

3.4 Dynamic Matching

This subsection reviews task matching research in the dynamic scenario, where the information of either tasks or workers is unknown beforehand. Dynamic matching can be further classified into *one-sided dynamic matching* and *two-sided dynamic matching*. In one-sided dynamic matching, only the information of workers or

Table 3: Comparison of existing solutions to task assignment as a static matching problem.

Method	Objective	Constraints ¹	Time Complexity ²	Ratio
GR [132]	maximizing total number	deadline, range	-	optimal
SP-WR-A [33]		range	-	heuristic
Temporal [200]		deadline, range, budget	-	heuristic
Greedy-GEACC [183]	maximizing total payoff	capacity	$O(n^3)$	$1/(1 + C_{max})$
g-D&C[66]		deadline, range, skill, budget	-	heuristic
ADAPTIVE [66]		deadline, range, skill, budget	-	heuristic
Basic [198]		deadline, range	-	optimal
IDA [214]	minimizing total distance	-	-	optimal
CA [214]		-	-	heuristic
Allocation [48]		capacity	$O(n^3)$	2.5
Swap-Chain [156]	minimizing maximum distance	capacity	$O(R \cdot T (T + W))$	optimal
Gale-Shapley [94]	minimizing #blocking pair	capacity	$O(T W)$	optimal
Closest Pair [72, 229]		capacity	$O(T W ^2)$	optimal
Chain [225]		capacity	$O((T + W) \cdot (\log^{O(1)} T + \log^{O(1)} W))$	optimal

¹ In the column of constraints, we use “-” to represent that the method supports no aforementioned constraints in Sec. 2.3. We use “range” to denote range constraints, “deadline” to denote deadline constraints (see Sec. 2.3 for more details).

² In the column of time complexity, we use “-” to represent the case when time complexity is not given in the paper. We use T and W to denote the set of tasks and the set of workers respectively. Hence n is $\max\{|T|, |W|\}$ and R is a parameter such that $R \ll |T||W|$.

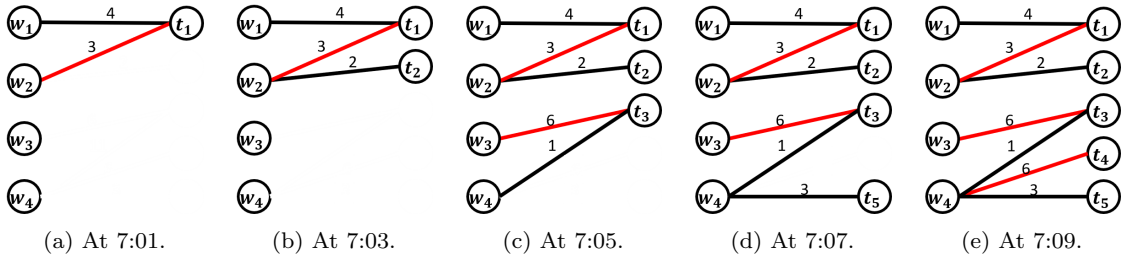


Fig. 3: An example of one-sided dynamic matching, where only tasks on the right appear dynamically. The information of workers on the left is known in advance.

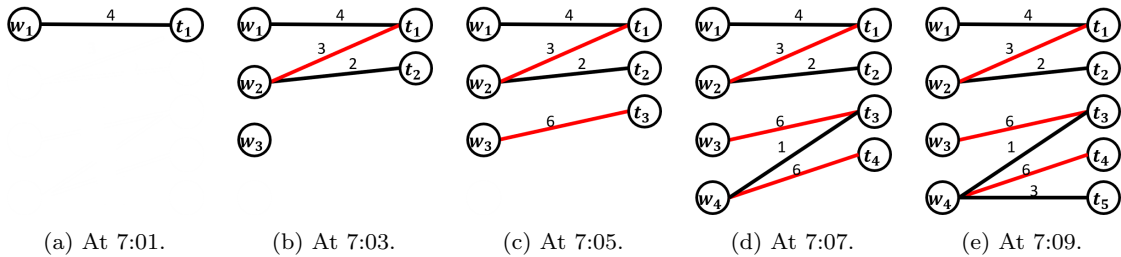


Fig. 4: An example of two-sided dynamic matching, where both workers and tasks appear dynamically and the assignment is made immediately.

tasks is unknown (*e.g.*, parcel delivery), while in two-sided dynamic matching, the information of both workers and tasks is unknown (*e.g.*, on-demand taxi dispatching). Fig. 3 and Fig. 4 show the examples of one-sided and two-sided dynamic matching, respectively. As in static matching, we review prior works based on their objectives: *utility maximization* (Sec. 3.4.1), *cost minimization* (Sec. 3.4.2) and *stable matching* (Sec. 3.4.3).

3.4.1 Utility Maximization

As in static matching, the utility between a task and its assigned worker in dynamic matching can also represent a constant value 1 and the payoff of the task. Since some real-world applications of dynamic matching allow workers to decide whether to accept the assigned task or not, the utility in dynamic matching can additionally

represent the accepted ratio of the worker and further the payoff times the acceptance ratio, *i.e.*, expected payoff. Accordingly, the main objectives for dynamic matching with utility maximization include *maximizing the (expected) total number of assigned tasks* and *maximizing the (expected) total payoff of assigned tasks*.

Maximizing Total Number of Assignments. Dynamic matching with this objective is also known as the online bipartite matching problem. Most research along this line focuses on the *one-sided online bipartite matching* problem [50, 82, 91, 102, 121, 129, 159], while relatively a few have investigated *two-sided online bipartite matching* [120, 209, 221].

(i) **Solutions to One-sided Scenario.** We discuss solutions optimized for worst-case performance and average performance, respectively.

- **Optimized for Worst-case Performance.** Karp *et al.* [129] propose three algorithms, *i.e.*, GREEDY, RANDOM and RANKING. GREEDY assigns a new task to an arbitrarily chosen available worker. RANDOM differs in that the worker is uniformly sampled. The competitive ratios of GREEDY and RANDOM are both 0.5 under the adversarial order model. RANKING is a two-phase algorithm. In the first phase, a random permutation of workers is picked and it represents the priority (*i.e.*, rank) of the workers. In the second phase, a newly appeared task will be assigned to the available worker with the highest rank. RANKING yields a competitive ratio $1 - 1/e \sim 0.632$ under the adversarial order model. The ratio is proven to be the lower bound of any online algorithm [50, 82, 102].
- **Optimized for Average Performance.** While RANKING has achieved the lower bound under the adversarial order model, it is unknown whether it is the most effective under other models such as the random order model and the i.i.d model [102]. Feldman *et al.* [91] devise the *suggested matching* algorithm with a higher ratio 0.67. The idea is to guide the online algorithm by offline solutions (“offline-guide-online”). Fig. 5 shows the procedure of the offline-guide-online technique. Specifically, they first predict the spatiotemporal information of tasks and workers (*i.e.*, learn the distribution). Next, an offline matching algorithm is used to obtain the optimal matching by using the predicted inputs. Finally, they use the offline matching to guide the online matching policy. When a new task appears, an eligible worker is usually sampled according to the chosen probability of this assignment in the offline solution. Following the same idea, Manshadi *et al.* [159] improve the ratio to 0.702 using Monte Carlo sampling. Jaillet *et al.* [121] apply linear program

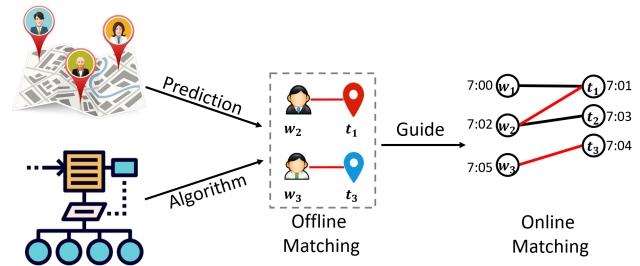


Fig. 5: Procedure of offline-guide-online technique.

as the offline algorithm and obtain the best known competitive ratio of 0.706.

(ii) **Solutions to Two-sided Scenario.** Many solutions to the two-sided scenario are built upon those to the one-sided scenario.

GREEDY can be applied to two-sided scenario and achieves a competitive ratio of 0.5 under the adversarial order model. Its effectiveness can be further improved by two methods, the randomized primal dual technique [120, 221] and the offline-guide-online technique [209].

In particular, the charging based framework [128] can be extended to achieve better effectiveness. Its idea is to increase the probability of each potential assignment whenever a worker or a task appears on the platform. At the deadline of each vertex, the algorithm will determine the final assignment of this vertex based on the probability. Wang *et al.* [221] extend the framework with the water-filling algorithm and obtain a better ratio of 0.526 than GREEDY under adversarial order model. Huang *et al.* [120] extend RANKING into the two-sided scenario where the vertex with higher rank has a larger probability to be matched. The extended RANKING algorithm achieves the currently best-known competitive ratio of 0.554.

Tong *et al.* [209] apply the *offline-guide-online* technique to a different setting where a worker can move in advance to other locations so as to increase the potential number of assignments. Their solution first predicts the spatiotemporal information of tasks and workers and guides workers to locations where there will be tasks in the future, and then makes assignments based on an offline guide. The proposed POLAR and POLAR-OP algorithms yield competitive ratios of 0.399 and 0.47 under the i.i.d model.

Maximizing Expected Total Number of Assignments. When workers are allowed to reject the assigned tasks, the objective above is replaced by maximizing the expected total number of assigned tasks.

Hassan *et al.* [111] use multi-armed bandit [176] to model the problem and apply a contextual bandit algorithm [143] to determine the assignments. Zhang *et al.*

[238] focus on predicting the acceptance ratio of workers in taxi dispatching via machine learning techniques. Note that tasks rejected by workers can be considered as new tasks and can be re-assigned to other workers.

Maximizing Total Payoff. Dynamic matching that maximizes the total payoff can be considered as an *online vertex-weighted bipartite matching* problem, where the weight of each edge in bipartite graph is represented by the weight of one-sided vertex. There are also two versions of this problem, *i.e.*, one-sided [31, 51, 195] and two-sided [84, 195].

(i) **Solutions to One-sided Scenario.** In [31], Aggarwal *et al.* study the problem where the information of tasks is known. A Perturbed-Greedy algorithm is proposed which achieves a competitive ratio $1 - 1/e \sim 0.632$ under adversarial order model. Specifically, the algorithm first perturbs each weight of vertices identically and independently by a function $\psi(x) = 1 - e^{-(1-x)}$. Then it sorts the vertices in the order of decreasing perturbed weights, which forms a rank. Finally, it utilizes the strategy of RANKING [129] to make the final decision. The authors prove that no randomized algorithm can obtain a higher ratio than 0.632 under the adversarial order model. Ting *et al.* [195] devise a randomized algorithm Greedy-RT to achieve the ratio of $\frac{1}{2e \ln \lceil U_{max} + 1 \rceil}$ under the adversarial order model, where U_{max} denotes the upper bound of the utility between a worker and a task. They first randomly sample a threshold and then match the new vertex to any existing vertex whose weight is higher than the threshold. Under the known i.i.d model, Brubach *et al.* [51] propose the VW algorithm with a competitive ratio of 0.729.

(ii) **Solutions to Two-sided Scenario.** For the two-sided scenario, Ting *et al.* [195] prove that Greedy-RT can still achieve the ratio of $\frac{1}{2e \ln \lceil U_{max} + 1 \rceil}$ under the adversarial order model. They also prove that no randomized algorithm can achieve a higher ratio than $\frac{2}{\lceil \log U_{max} \rceil + 1}$ under the adversarial order model. Dickerson *et al.* [84] design an algorithm ADAP based on the offline-guide-online technique. They first solve a linear program benchmark and then use the offline solution to simulate the online matching procedure. Finally, they prove that the competitive ratio of ADAP is 0.343 under the random order model.

Maximizing Expected Total Payoff. Similar to the case of maximizing the expected total number of assignments, this thread of research assumes the worker can reject the task with some probability. In this case, the weight of edges in the bipartite graph is determined by both the payoff of the task and the acceptance ratio of the worker. Hence the problem is similar to the **online edge-weighted bipartite matching** problem.

Again, there are two versions of this problem, *i.e.*, one-sided [51, 83, 134, 139] and two-sided [84, 206].

(i) **Solutions to One-sided Scenario.** Prior works either borrow the idea from the secretary problem [92] or use the offline-guide-online technique [51, 83].

Korula *et al.* [139] first propose a Sample-And-Price algorithm that has a competitive ratio of 0.125 under the random order model. The idea is to iteratively find a global assignment by GREEDY whenever a new vertex appears, and then sample an assignment with some probability. Kesselheim *et al.* [134] are also motivated by the secretary problem and devise the BOM algorithm, which improves the competitive ratio to $1/e \sim 0.367$ under the random order model. Different from Sample-And-Price, BOM skips the first $\lfloor (|W| + |T|)/e \rfloor$ vertices and finds a global optimal matching by the Hungarian method.

Solutions which use the offline-guide-online technique to obtain more promising results. In [51, 83], the authors first formulate the predicted instance with linear programming (LP), and then solve it by existing LP solver (*e.g.*, CPLEX [20]). They finally use the result of the LP solver to guide the online matching procedure. Under the known i.i.d model, the SW algorithm [51] can achieve a competitive ratio of 0.632 and an optimized algorithm EW [51] can obtain a ratio of 0.705. In [83], the authors exploit the fact that a worker tends to be re-assigned a new task right after he/she finishes the last task, and define the online matching with (offline) reusable resources problem. They propose a Monte-Carlo simulation based algorithm $ADAP(\gamma)$, which achieves a competitive ratio of 0.5 under the known adversarial distribution model.

(ii) **Solutions to Two-sided Scenario.** In [206], the authors extend the Greedy-RT algorithm [195] and prove its competitive ratio still holds. They also borrow the idea of secretary problem and devise a two-phase based framework. In the first half of vertices, GREEDY is used to determine the final assignment. In the other half of vertices, they first find a global matching and then determine the final assignment based on the global matching. Their proposed algorithms achieve competitive ratios of 0.25 and 0.125 under the random order model. Dickerson *et al.* [84] further use the offline-guide-online technique to improve the ratio to 0.295. Song *et al.* [187] study a variant of the problem for applications such as InterestingSport [11] and Nanguache [12], where workplaces, workers and tasks should all be considered. For example, InterestingSport needs to find suitable trainers (*i.e.*, workers) and book the corresponding sports facilities (*i.e.*, workplaces) for its users. The problem is modeled as online trichromatic match-

ing. A threshold-based randomized framework is proposed to solve the problem with a ratio of $\frac{1}{3e \ln[U_{max}+1]}$.

Summary. Most of the efforts on dynamic matching with utility maximization can be modeled as a variant of online bipartite matching problem. The offline-guide-online technique [91] is useful to achieve good competitive ratios, *e.g.*, [84, 209]. However, the common assumption is that the spatiotemporal distribution of either tasks or workers is completely predictable, which may be impractical in real-world applications.

3.4.2 Cost Minimization

The cost between a task and a worker can represent the travel distance (time) between the location of the worker and the location of the task, or the delay of the task from release time to completion time. Hence cost minimization indicates that tasks will be served more rapidly. We discuss solutions that minimize the *total travel distance* and the *total delay* separately.

Minimizing Total Travel Distance. Dynamic matching with this objective can be modeled as a variant of one-sided online minimum bipartite matching, where tasks dynamically appear on the platform. Existing solutions can be classified into two categories, *greedy* [127] and *HST based* [46, 161] algorithms.

- **Greedy based.** Kalyanasundaram *et al.* [127] propose Permutation, a $(2n - 1)$ -competitive ratio algorithm, where n is the number of workers to be matched. They also introduce Greedy, which greedily assigns a task to its closest worker (*i.e.*, nearest neighbor) and randomly picks one if there is a tie. Despite its efficiency, Greedy has a competitive ratio of $2^n - 1$ under the adversarial order model. In order to distinguish from the GREEDY algorithm [129] on utility maximization, we call this nearest neighbor based greedy method as NN-Greedy.
- **HST based.** Hierarchically Separated Tree (HST) [224] is a special type of tree metrics. Meyerson *et al.* [161] consider a randomized greedy algorithm, HST-Greedy, by extending the Permutation algorithm [127] into HST structures and it yields an expected competitive ratio of $O(\log^3 n)$ on any metric space. The $O(\log^3 n)$ bound is further improved in [46] by the HST-Reassignment algorithm, which is $O(\log n)$ -competitive on 2-HST metrics, and thus $O(\log^2 n)$ -competitive on general metrics.

The above studies focus on analyzing the competitive ratios in worst cases *i.e.*, under the adversarial order model. To evaluate the performance of these algorithms in practice, Tong *et al.* [205] present a comprehensive experimental comparison of some represen-

tative algorithms. The experiments show that the NN-Greedy, which has always been considered as the worst method due to its exponential competitive ratio ($2^n - 1$), significantly outperforms the others in terms of effectiveness. In particular, the worst case in the adversarial order model of NN-Greedy has a constant competitive ratio, 3.195 under the random order model.

Minimizing Total Delay. The delay of a task is the duration from its release time to its completion time. Unlike minimizing the total travel distance, minimizing the total delay is a group of problems where once a task appears, it can be kept waiting for potential better assignments instead of being matched immediately. The cost incurred is the sum of travel distances between matched worker-task pairs (the travel cost), and the sum of the tasks' response time (the waiting cost). The rationale is to trade off between the cost of instant assignments and that of waiting for better assignments.

- **Solutions to One-sided Scenario.** Emek *et al.* [88] present a randomized algorithm with competitive ratio $O(\log^2 n + \log \Delta) \sim O(\log^2 n)$ on n -point metric spaces with the longest distance Δ . Ashlagi *et al.* [42] prove the same ratio with a simpler analysis and Azar *et al.* [43] further improve the ratio to $O(\log n)$ under the adversarial order model.
- **Solutions to Two-sided Scenario.** Chen *et al.* [64] study the problem of minimizing the maximum delay among all matches while both tasks and workers dynamically appear. They present an HST-based algorithm MMD-HST, which has better effectiveness than a greedy based baseline.

Summary. Dynamic matching with cost minimization is usually modeled as an online minimum bipartite matching problem. There are mainly two kinds of solutions to this problem, greedy based and HST based algorithms. Compared with greedy based algorithms, HST based algorithms tend to have better competitive ratios in worst case analysis. However, since NN-Greedy is demonstrated to be effective on both synthetic and real datasets [205], it is still an open problem whether the competitive ratio of NN-Greedy is arbitrarily bad (*i.e.*, $2^n - 1$) under other analysis models (*e.g.*, random order model or known i.i.d model).

3.4.3 Stable Matching

Many studies on dynamic matching have integrated the preferences of *either* workers *or* tasks into their optimization objectives. For example, some studies [83, 206] aim to maximize the total utility obtained from all successful assignments, where the utility represents the workers' preference on payoff.

Table 4: Comparison of existing solutions to task assignment as a dynamic matching problem.

Method	Objective	Constraints	Time Complexity ¹	Analysis Model ²	Ratio
GREEDY [129]	maximizing total number	one-sided, range	$O(n^2)$	AO	0.5
GREEDY [102]		one-sided, range	$O(n^2)$	RO, IID	0.632
RANKING [129]		one-sided, range	$O(n^2)$	AO	0.632
suggested matching [91]		one-sided, range	$O(n)$	IID	0.632
MC sampling [159]		one-sided, range	$O(n)$	IID	0.702
Random Lists [121]		one-sided, range	$O(n^2)$	IID	0.706
GREEDY [120]		two-sided, range, deadline	$O(n^2)$	AO	0.5
water-filling [221]		two-sided, range	$O(n^2)$	AO	0.526
ext-RANKING [120]		two-sided, range, deadline	$O(n^2)$	AO	0.554
POLAR-OP [209]		two-sided, range, deadline	$O(n^2)$	IID	0.47
contextual bandit [111]	maximizing expected total number	two-sided, range, deadline	$O(n^2)$	-	heuristic
hill-climbing [238]		two-sided, range	$O(n^3)$	-	heuristic
Perturbed-Greedy [31]	maximizing total payoff	one-sided, range	$O(n \log n)$	AO	0.632
VW [51]		one-sided, range	$O(n^2)$	KIID	0.729
Greedy-vRT [195]		two-sided, range	$O(n^2)$	AO	$\frac{1}{2e \ln[U_{max} + 1]}$
ADAP [84]		two-sided, range, deadline	$O(n^2)$	KIID	0.343
Sample-And-Price [139]	maximizing expected total payoff	one-sided, range	$O(n^3)$	RO	0.125
BOM [134]		one-sided, range	$O(n^4)$	RO	0.367
SW,EW [51]		one-sided, range	$O(n^2)$	KIID	0.632,0.705
ADAP(γ) [84]		two-sided, range	$O(n)$	KAD	0.5
TGOA [206]		two-sided, range, deadline	$O(n^4)$	RO	0.25
TGOA-Greedy [206]		two-sided, range, deadline	$O(n^3 \log n)$	RO	0.125
NADAP [84]		two-sided, range, deadline	$O(n)$	KIID	0.295
Permutation [127]		minimizing total distance	one-sided	$O(n^3)$	AO
NN-Greedy [127]	one-sided		$O(n)$	AO	$2^n - 1$
HST-Greedy [161]	one-sided		$O(n)$	AO	$O(\log^3 n)$
HST-Reassignment [46]	one-sided		$O(n^2)$	AO	$O(\log^2 n)$
stilt-walker [88]	minimizing total delay	one-sided	-	AO	$O(\log^2 n)$
saturated [43]		one-sided	-	AO	$O(\log n)$
TGM [41]		one-sided	-	AO	$O(\log n)$
MMD-HST [64]	minimizing maximum delay	two-sided	-	AO	$O(\log n)$
FCFS-Greedy [135]	minimizing #blocking pair	one-sided	$O(n^2)$	AO	$O(E)$
LP-ALG [243]	maximizing total payoff & minimizing #blocking pair	one-sided, range	$O(n)$	KIID	0.632 $0.6 E $

¹ In the column of constraints, we use “range” to denote range constraints, “deadline” to denote deadline constraints (see Sec. 2.3 for more details).

² In the column of time complexity, we use “-” to represent that the time complexity is not given in the paper. We use n to denote the maximum value between the number of tasks and the number of workers.

³ In the column of analysis model, we use “-” to represent that the paper has no competitive analysis under specific models.

Zhao *et al.* [243] first consider the preferences of both workers and tasks in dynamic task matching and formulate the task assignment problem as a variant of online stable matching problem. The online stable matching problem is first studied by Khuller *et al.* [135]. They prove that the “first come, first served” method (FCFS-Greedy) produces $O(n \log n)$ blocking pairs on average and $O(n^2)$ blocking pairs in worst case. Zhao *et al.* [243] study a more difficult version since they also aim to maximize the total utility at the same time. They use the offline-guide-online technique [91] and propose an LP based algorithm LP-ALG, which achieves a competitive ratio of $1 - 1/e \sim 0.632$ for maximizing total

utility with no more than $0.6|E|$ blocking pairs under the known i.i.d model.

3.4.4 Summary on Dynamic Matching

Table 4 compares the representative research on dynamic matching with three objectives (utility maximization, cost minimization and stable matching). As is shown, the competitive ratio is often constant in solutions to utility maximization and different analysis models are often used to obtain a better result. However, fewer studies focus on minimizing the total cost or online stable matching. In particular, most research [46, 88, 161] uses the adversarial order model to

analyze the effectiveness of the algorithm in the worst case, which is more difficult to obtain a promising result. Thus, it is still an open problem whether it is possible to design an algorithm with a constant competitive ratio under the random order model or the known i.i.d model. Finally, it is worth mentioning that many research (*e.g.*, [145, 209, 243]) conducts the experiments of dynamic matching on the datasets collected by DiDi Chuxing [4]. DiDi Chuxing has so far already released many open datasets in their GAIA initiative [22]. These real datasets can usually be used to validate the performances of the dynamic matching algorithms for different objectives.

3.5 Static Planning

Task assignment in the real applications such as ride sharing and food delivery is a planning problem, where a route (*i.e.*, a sequence of tasks) should be planned for workers. This subsection reviews studies on static planning, which fall into two categories, **One-Worker-To-Many-Tasks Static Planning** (Sec. 3.5.1), which plans a route for one single worker, and **Many-Workers-To-Many-Tasks Static Planning** (Sec. 3.5.2), which plans routes for multiple workers.

3.5.1 One Worker To Many Tasks

In One-Worker-To-Many-Tasks Static Planning, most studies aim to find a route for one worker such that the number of performed tasks is maximized under the travel budget constraint. This problem is closely related to the orienteering problem [215]. The major differences include: (1) the utility value of each matching is often zero or one, and (2) the end vertex of the route is not given. Thus, the utility often represents a constant value 1 in the majority of works [78, 80] and only [73] considers the more general utility (*i.e.*, payoff). We discuss existing works based on their objectives.

Maximizing Total Number of Assignments. Deng *et al.* [78] first study static planning which maximizes the total number of performed tasks under the travel budget and deadline constraints. They name it the Maximum Task Scheduling (MTS) problem and prove its NP-hardness. There are two kinds of solutions to the this problem: *exact* and *greedy based* algorithms.

- **Exact.** To address the MTS problem, Deng *et al.* [78] propose several exact solutions. They first propose a dynamic programming algorithm, MST-DP, with a time complexity of $O(n^{2^2n})$ and a space complexity of $O(n^{2^n})$. They further propose a branch-and-bound based algorithm MST-BB, which has a

time complexity of $O(n!)$ and a space complexity of $O(n^2)$. They also propose several pruning strategies to improve the actual running time.

- **Greedy based.** Deng *et al.* [78] also propose several greedy based heuristics, including Nearest Neighbor Heuristic (NNH), Most Promising Heuristic (MPH) and Least Expiration Time Heuristic (LEH). Among these solutions, NNH is the most efficient and effective. To achieve a better trade-off between efficiency and effectiveness, they further present Beam Search Heuristic (BSH) [80]. It expands the cardinality of candidate set to a given threshold instead of one in NNH. BSH then invokes MST-BB with this candidate set to select proper tasks. Even though BSH is less efficient than NNH, it is more effective in experimental evaluations.

Maximizing Total Payoff. Costa *et al.* [73] study static planning which maximizes the total payoff. They assume that a worker may be on his/her preferred path and is willing to consider the trade-off between payoff and the travel cost. Due to its NP-hardness, they propose a Detour Oriented Heuristic (DOH) to find all non-dominated routes and recommend them to the workers.

3.5.2 Many Workers To Many Tasks

Although it is already NP-hard to plan a route for a single worker, a few efforts have explored Many-Workers-To-Many-Tasks Static Planning. Research on Many-Workers-To-Many-Tasks Static Planning mainly focuses on maximizing the general utility (*e.g.*, satisfaction score [97, 182], payoff [113]) while only [79] aims at maximizing the total number of performed tasks.

Maximizing Total Number of Assignments. Deng *et al.* [79] extend their Maximum Task Scheduling problem to the multiple-workers version. They devise a new three-phase framework called Global Assignment and Local Scheduling (GALS). The first two phases are static matching and One-Worker-To-Many-Tasks Static Planning. The last phase is to refine the matching result with the updated routing result. The last two phases repeat until no more tasks can be performed. The complexity of GALS is $O(n^4)$, which is relatively high in practice. Thus, they propose the Local Assignment Local Scheduling (LALS) algorithm based on the similar idea to improve the efficiency.

Maximizing Total Payoff/Satisfaction. In practice, the utility function can represent the satisfaction score [97, 182] between workers and tasks or the payoff of the worker by performing the task [113]. There are mainly two types of solutions to this problem, *greedy based* [97, 182] and *local ratio based* [113, 182] algorithms.

Table 5: Comparison of existing solutions to task assignment as a static planning problem.

Method	Objective	Constraints	Time Complexity ¹	Ratio
MST-BB [78]	maximizing total number	deadline	$O(n!)$	optimal
NNH [78]		deadline	-	heuristic
GALS [79]		deadline	$O(n^4)$	heuristic
LRBA [113]	maximizing total utility	travel budget	-	5
DeDPO [182]		deadline, travel budget	$O(n^3)$	0.5
SCUP [97]		deadline, travel budget, skill	$O(n^2)$	heuristic

¹ In the column of constraints, we use “deadline” to denote deadline constraints (see Sec. 2.3 for more details).

² In the column of time complexity, we use “-” to represent that the time complexity is not given in the paper. We use n to denote the maximum value between the number of tasks and the number of workers.

- **Greedy based.** She *et al.* [182] propose the problem of Utility-aware Social Event-participant Planning (USEP), which maximizes the total satisfaction of all the users considering the travel budget constraint. They propose a greedy based algorithm, RatioGreedy, which considers the utility-cost ratio of each worker-task pair and adds the pair with the largest ratio into the planning. Gao *et al.* [97] study a variant of the problem, where tasks may impose different skill requirements on the workers. They first form a set of worker with minimum cardinality to cover the skill requirement of tasks and then greedily assign the worker with the largest satisfaction to the tasks.
- **Local Ratio based.** The main idea of the local ratio framework [47] is to first decompose the problem into several simpler sub-problems and then eliminate the conflict of these sub-problems. She *et al.* [182] propose a two-phase algorithm called DeDP. It achieves the approximation ratio of 0.5 with time complexity of $O(n^3)$ and space complexity of $O(n^2)$. To improve the efficiency, they further devise an optimized algorithm DeDPO. To maximize the reward of the performed tasks, He *et al.* [113] propose a local ratio based algorithm, LRBA. They also use the same technique to prove that the approximation ratio of LRBA is 5. Experimental results show that LRBA outperforms a greedy based algorithm.

3.5.3 Summary on Static Planning

Table 5 summarizes existing works on static planning. Static planning in spatial crowdsourcing has been studied in two settings, One-Worker-To-Many-Tasks Static Planning and Many-Workers-To-Many-Tasks Static Planning. Since One-Worker-To-Many-Tasks Static Planning is NP-hard, the greedy based solutions are proposed to improve the efficiency of exact solutions. However, all greedy based solutions have no theoretical guarantee

in the effectiveness. The local ratio technique is often exploited to design an approximation solution. Experiments [113, 182] on the Meetup datasets in [153] show that the local ratio based algorithm is more effective than the greedy based solution.

3.6 Dynamic Planning

Dynamic planning is the planning problem where the information of workers or tasks is unknown in advance. It is more challenging than static planning since the routes of workers have to be planned when only partial information is available. As with static planning, we review research on dynamic planning in two categories: **One-Worker-To-Many-Tasks Dynamic Planning** (Sec. 3.6.1) and **Many-Workers-To-Many-Tasks Dynamic Planning** (Sec. 3.6.2).

3.6.1 One Worker To Many Tasks

Research on One-Worker-To-Many-Tasks Dynamic Planning often maximizes the total utility under the budget of travel cost. As before, we review two kinds of total utilities, *the total number of assignments* [144] and *the total payoff of the workers* [191].

Maximizing Total Number of Assignments. Li *et al.* [144] prove that under the adversarial order model, no deterministic algorithm has a constant competitive ratio. They propose several greedy based approaches such as Nearest Neighbor Heuristic (NN-Greedy) and Earliest Deadline Heuristic (ED-Greedy). They further propose a bi-directional search based algorithm to improve the effectiveness. The search begins with the origin and the destination of the worker. Some pruning strategies are proposed to reduce the searching space.

Maximizing Total Payoff. Sun *et al.* [191] extend the problem in [144] to maximize the total payoff to workers. They devise an NN-Greedy based algorithm

Table 6: Comparisons of existing solutions to task assignment as a dynamic planning problem.

Method	Objective	Constraints	Time Complexity ¹	Analysis Model ²	Ratio
Re-Route [144]	maximizing total number	deadline	-	AO	heuristic
Auction-SC [38]		deadline	-	-	heuristic
Fast-Planning [193]	maximizing total payoff	deadline	$O(n^3)$	AO	heuristic
APART [40]		deadline, budget	-	AO	heuristic
EPBR [191]		deadline, range	-	-	heuristic
PBM [248]		deadline, budget	$O(n^3)$	-	heuristic
t-share [158]	minimizing total travel distance	deadline	-	-	heuristic
kinetic [119]		deadline, budget	-	-	heuristic
pruneGreedyDP [212]	minimizing unified cost	deadline	$O(n^2 + n^2 \log n)$	AO	heuristic

¹ In the column of constraints, we use “range” to denote range constraints, “deadline” to denote deadline constraints (see Sec. 2.3 for more details).

² In the column of time complexity, we use “-” to represent that the time complexity is not given in the paper. We use n to denote the maximum value between the number of tasks and the number of workers.

³ In the column of analysis model, we use “-” to represent that the paper has no competitive analysis under specific models.

to balance three influence factors on a worker’s choice in terms of which task to undertake next. They further borrow the idea of offline-guide-online technique [91] to enhance the effectiveness and efficiency.

3.6.2 Many Workers To Many Tasks

Among the planning problems discussed in this survey, dynamic planning for multiple workers is the most challenging. We review existing literature with the objectives to maximize the total number of assignments [38], maximize the total payoff [40, 193, 248] or minimize the total travel distance [119, 158].

Maximizing Total Number of Assignments. In [38], the authors design an auction based framework. In the framework, workers give out their bids according to their best schedule if incorporating the new task and the platform then selects a worker for the task.

Maximizing Total Payoff. Tao *et al.* [193] devise two algorithms, Delay-Planning and Fast-Planning to solve the problem. In Delay-Planning, the worker, who has not finished his/her currently assigned tasks, will not be allocated to the newly arrived tasks. Instead, the route of a worker in Fast-Planning may be updated when new tasks arrive. Both [40] and [248] focus on maximizing the total payoff in another type of application, ride sharing. Asghari *et al.* [40] propose a branch-and-bound solution to find the optimal routes. Zheng *et al.* [248] devise an order matching based solution.

Minimizing Total Travel Distance. Both [158] and [119] aim to minimize the total travel distance while trying to serve all requests. Ma *et al.* [158] first study the dynamic task planning for ride sharing service on a road network. A filter-and-refine based framework *t-share* is devised with grid index. Based on a similar

framework, Huang *et al.* [119] design a trie based data structure called *kinetic tree*. The kinetic tree applies the procedure of *insertion* to update the route of each worker.

3.6.3 Summary on Dynamic Planning

Table 6 compares existing works on dynamic planning. Existing studies on dynamic planning, particularly those for ride sharing service, has two main limitations. First, the optimization objectives in some papers are conflicting (*e.g.*, [158] and [119]). Second, some solutions are inefficient. Specifically, some algorithms are inefficient when the capacity of workers becomes larger. For example, [248] restricts that the capacity is no more than 2 and [119] can not response in real time anymore when the capacity becomes 6 (see experiments in [212]). Major solutions rely on inefficient *insertion* procedure [119, 158]. To address these limitations, Tong *et al.* [212] abstract a unified formulation of dynamic planning in sharing transportation, *i.e.*, URPSM problem, which generalizes the previous two objectives. They further design a novel dynamic programming based *insertion* operation to improve the efficiency. They compare their solutions with the state-of-the-art algorithms on two large-scale datasets, *i.e.*, the GAIA datasets [22] collected by DiDi Chuxing and the NYC datasets [16] collected from the taxis in New York City. Experiments on these two datasets show that their framework *prune-GreedyDP* outperforms *t-share* [158] and *kinetic* [119].

3.7 Discussions

We summarize representative studies on each category of task assignment in Table 3 (static matching), Table 4

(dynamic matching), Table 5 (static planning) and Table 6 (dynamic planning). Almost all these papers focus on the micro-tasks rather than macro-tasks. This is because a macro-task (*e.g.*, mapping a city) is usually decomposed into large numbers of micro-tasks (*e.g.*, geotagging a landmark in this city) on real-world platforms. Then the algorithms can still be used to determine the allocation between workers and decomposed micro-tasks. Comparing these studies, many focus on the dynamic scenario instead of the static scenario and there are more papers on matching than planning. It seems that the offline-guide-online technique is helpful to obtain a better competitive ratio in dynamic task matching under the known i.i.d model or the known adversarial distribution model. We also observe that there is no competitive algorithm in dynamic planning. Thus, the offline-guide-online technique from dynamic matching may be a starting point to devise competitive algorithms for dynamic planning. Finally, despite extensive research on either static planning or dynamic planning, there is still no comprehensive evaluation on these solutions either empirically or theoretically.

4 Quality Control

One characteristic of crowdsourcing is that tasks are performed by workers of diverse quality. Quality control aims to ensure high-quality task completion in presence of diverse worker quality, which is achieved by allowing *multiple* workers to perform *the same* task. Quality control in traditional crowdsourcing roughly deals with two issues: (1) how to quantify the quality of workers and tasks; and (2) how to aggregate results from workers of diverse qualities to meet the quality requirements of tasks. The spatiotemporal factors add new dimensions in both issues, which we discuss in this section.

4.1 Quality Modeling

The definition of worker and task quality is application-specific. We focus on the worker and task quality related to spatiotemporal factors.

4.1.1 Quality of Worker

First we discuss worker quality used in traditional crowdsourcing (*inherent worker quality*) and then the new factors in spatial crowdsourcing (*spatiotemporal related worker quality*). Finally we briefly review the methods to estimate the quality of workers.

Inherent Worker Quality. Worker quality in traditional crowdsourcing can be modeled by worker probability [53, 105, 210], confusion matrix [216, 223] and diversity of skills [115, 247]. Specifically, the worker probability approach uses a single value to model the quality of a worker. The value can be the accuracy, confidence, experience or reputation of the worker. A large value normally means a high worker quality. However, the single-valued quality may not suffice to characterize the worker quality for some complex tasks. Hence multi-dimensional approaches such as vectors and confusion matrices are proposed to describe worker quality. The elements in the vectors or confusion matrices represent various skills of workers and the conditional probabilities with different truth values. For example, a normalized four-dimensional vector $(0.30, 0.78, 1.00, 0)^T$ may represent a worker’s abilities on Java, Python, Ruby and C#. Each row of a confusion matrix is the probability distribution under the condition of different correct answers. In general, the vector and matrix approaches characterize workers in more detail and outperform the single-valued worker probability model [249].

Spatiotemporal Related Worker Quality. In spatial crowdsourcing, quality of workers is often affected by extra spatiotemporal constraints. For example, in addition to an inherent quality as mentioned above, each worker is also assumed to have a distance-aware quality in crowdsourced POI labeling applications [117]. In fact, it is common for spatial crowdsourcing applications to assume that workers can only reliably perform tasks within a certain range [133, 206].

Assessment of Worker Quality. The assessment methods of worker quality vary for different aspects of worker quality. Assessment of the inherent quality is usually based on historical data [63, 65, 133, 238]. For example, the historical accuracy to perform tasks is used to estimate the accuracy of a worker to perform future tasks [65, 238]. Spatiotemporal related quality is often set via various spatiotemporal data processing models. For distance-aware quality, parameter estimation methods like Bayesian [100, 167, 168] and Maximum Likelihood estimation [117] are adopted to evaluate worker qualities with different distance sensitivities.

4.1.2 Quality of Task

On the one hand, similar to traditional crowdsourcing, the quality of a crowdsourcing task is evaluated by reliability, which is usually formalized as the probability that over 50% workers correctly answer the task [133, 181, 220] or the chance that at least one worker successfully completes the task [103, 241]. Specifically, [133] was the first work to consider the quality issue in spa-

tial crowdsourcing. These studies [133, 181, 220] focus on the spatial tasks that needs a qualified answer, *e.g.*, spatial data collection by taking photos. Therefore, the requester of the task usually has an expectation of the final answers. Differently, another type of tasks only needs to be successfully completed by one worker, *e.g.*, the on-demand taxi calling service in DiDi Chuxing [4]. Thus, such studies [103, 241] focus on the probability that at least one worker can eventually finish the task.

On the other hand, unlike the crowdsourcing tasks commonly seen in traditional crowdsourcing, the spatiotemporal factors may directly reflect the quality of tasks in spatial crowdsourcing.

Latency as Task Quality. Latency of tasks is closely related to the quality of service for a spatial crowdsourcing platform. Specifically, Zeng *et al.* [233] consider the maximum latency of all tasks as a criterion for task quality. This criteria is commonly used in real-world applications like Facebook Editor [5] and OpenStreetMap [13]. Differently, Das *et al.* [75] consider the average latency of all tasks as a criterion for task quality. The average latency is usually considered as the quality of tasks in taxi dispatching platform (*e.g.*, Uber [17] and DiDi Chuxing [4]) or food/parcel delivery platform (*e.g.*, Meituan [26] and Cainiao [3]).

Diversity as Task Quality. Diversity is particularly important for event detection or labelling applications. For example, a POI may need to be labelled multiple times by different workers so that reasonably accurate and complete information about the POI can be obtained [118]. Cheng *et al.* [65] first consider the diversity in the quality of tasks. They observe two types of diversity from the tasks in spatial crowdsourcing: *spatial diversity* and *temporal diversity*.

Specifically, spatial diversity is important when some tasks ask the workers to take photos/videos of the city landmarks from different angles. When there are r workers around the task, the authors use the entropy to define the spatial diversity (SD) as

$$SD = - \sum_{j=1}^r \frac{A_j}{2\pi} \cdot \log\left(\frac{A_j}{2\pi}\right), \quad (1)$$

where A_j is the angles between two results (photos).

Temporal diversity is important when some tasks require the workers to complete the tasks at different time intervals. For instance, a vacant parking space needs to be monitored at different time windows [65]. If there are r workers who will be working at each time interval of the whole working period T , the temporal diversity is also defined based on the idea of entropy as

$$TD = - \sum_{j=1}^{r+1} \frac{t_j}{T} \cdot \log\left(\frac{t_j}{T}\right), \quad (2)$$

where t_j is the j -th time interval.

The two kinds of diversity can also be combined to assess the spatiotemporal diversity (STD) of a task:

$$STD = \beta \cdot SD + (1 - \beta) \cdot TD, \quad (3)$$

where β is a parameter to balance the importance of spatial diversity and temporal diversity.

4.2 Result Aggregation

Given the worker quality and the results from multiple workers, aggregation techniques derive the final result for each task so that the quality requirements of tasks can be satisfied. Typical aggregation techniques [249] include Majority Voting [53, 140], Weighted Majority Voting [115, 147], Probabilistic Graphical Models [77, 174], etc. Aggregation techniques in spatial crowdsourcing need to account for spatiotemporal factors, which brings in new aggregation techniques.

In the task of real-time urban traffic speed estimation, workers are assigned to collect or voluntarily contribute traffic data in different locations, and the goal of the task is to reliably estimate the traffic speed in the road network. For example, in [116, 155], the systems recruit workers to probe the real-time traffic speed of some roads, while Waze [19] collects traffic data from users' mobile phones to estimate the average speed when its users drive around with the app turned on. Existing studies generally ignore the quality of workers, implicitly assuming that the data collected by workers are reliable. In addition, it is often the case that limited number of workers can be recruited to measure the traffic speed because of the budget constraint, *i.e.*, only the speeds of part of road segments are available. Therefore, the problem boils down to choosing the optimal subset of road segments to measure in order to maximize the quality of speed estimation of the entire road network. Hu *et al.* [116] study the real-time urban traffic speed estimation problem where only the speeds on a predefined number of roads (seeds) can be obtained by spatial crowdsourcing. They propose five algorithms (SUPGREEDY, RANDOM, MAXCOV, COVGREEDY, HYBRIDGREEDY) to select seeds and present a two-step model to estimate the speeds of other roads, taking advantage of the correlation among roads. Specifically, the first step constructs a probability graphical model to infer the traffic trend and the second step estimates the traffic speed using a hierarchical linear model. Evaluations on the taxi datasets [1] collected in Beijing and Nanjing show a traffic speed estimation accuracy around 80%. Similar to [116], Liu *et al.* [155] capture two statistical properties of speed, periodicity

Table 7: Comparison of representative studies on quality control in spatial crowdsourcing.

Reference	Quality Modeling		Aggregated Method ¹
	Worker	Task	
[133]	probability	reliability	majority voting
[233]	probability	latency and reliability	majority voting
[65]	probability	diversity and reliability	majority voting
[220]	probability	reliability	weighted majority voting
[103]	probability	reliability	-
[241]	probability	reliability	-
[100]	probability	reliability	bayesian estimation
[167, 168]	probability	reliability	bayesian estimation
[117]	probability and distance	reliability	expectation maximization

¹ In the column of aggregated method, we use “-” to represent that the paper aims to guarantee that at least one worker can successfully complete the task (*e.g.*, on-demand taxi-dispatching) and hence the proposed method does not need to consider the aggregation.

and correlation, using a probabilistic graphical model. They propose to select the best set of workers to probe the real-time traffic speed for the corresponding roads using a hybrid greedy-based algorithm with an approximation ratio above $(1 - \frac{1}{e})/2$. The traffic speed of the entire road network is then estimated using speed propagation based on the model constructed beforehand. The final false estimation rate of the proposed method on the gMission dataset [63] is around 0.08.

In the crowdsourced POI labeling task, a graphical probability model is proposed to deduce the correct labels [117]. Assuming that the labeling results follow a conditional distribution on worker quality, POI influence and the true labels, the authors propose a Maximum Likelihood Estimation (MLE) and Expectation Maximization (EM) method to estimate the unknown probability parameters and labeling results.

In the task of crowdsourced event detection, reports from different workers are aggregated to detect the true event [100, 168]. In [168], the problem is formulated as truth inference under missing or wrong reports. The authors model missing and wrong reports based on the location popularity, the truth of events and the participant reliability, and propose a recursive inference algorithm to infer the latent variables and the truth of events. The method is extended in [100] by considering the state of event as a function of time. The authors design inference algorithms to update the conditional probability of report and variables recursively until the true label of event converges. The Kalman filter is also used to improve the approximation to the event truth.

In the tasks of collaborative mapping, workers often voluntarily participate in map making without financial compensation. In such applications (*e.g.*, OpenStreetMap [13] and Wikimapia [29]), the major purpose of the macro task is to map a large region (*e.g.*, city), which can be decomposed into large numbers of micro-tasks (*e.g.*, mapping a landmark). Quality control for

such macro-tasks, *i.e.*, obtaining the qualified results of the macro-task, consists of three steps.

- **Assessment of worker/task quality.** Since the workers are usually volunteered, the qualities of workers and tasks may notably differ in practice [114, 179]. On one hand, the inherent quality of worker is usually based on historical records and the user profiles [236]. On the other hand, the quality of task can be evaluated based on spatiotemporal diversity [65, 70]. Existing work also uses the densities of the tasks to assess the quality of task [70], *e.g.*, the number of provided answers over the area of the region, the number of volunteered workers over the population of the region, etc.
- **Aggregation of micro-tasks.** With the decomposition of the macro-tasks, the results of each micro tasks can be independently aggregated. Therefore, typical aggregation techniques include voting [249] and rating [180] can be applied. Some platforms like OpenStreetMap [13] also allow the expert workers to help validate the aggregated answers.
- **Removal of inconsistencies.** Finally, the results of some micro-tasks may be conflicting from the global view of the macro-task, *e.g.*, administrative boundaries self-intersect or split instead of being closed-loop sequences of roads. Thus, existing work also investigates removing the consistencies between the micro tasks. KeepRight [24] is a data consistency check tool for OpenStreetMap which can detect errors in the map data, such as loops, overlapping ways and missing boundaries. Hashemi *et al.* [110] present a similarity-based framework to detect the logical, topological inconsistencies according to the spatial relationships of micro-tasks.

A few studies have also explored deep learning [56] in collaborative mapping.

4.3 Discussions

In a sense, quality control and task assignment in spatial crowdsourcing are interwoven. Table 7 summarizes existing studies on quality control.

On the one hand, the quality metrics of workers and tasks in Sec. 4.1 can be directly applied as either a constraint or an objective in the task assignment problems in spatial crowdsourcing. For example, in [65], maximizing the expected spatial/temporal diversities and the smallest reliability among all tasks are regarded as part the objective of task assignment. In the maximum correct task assignment problem [133], a correct match between a task and assigned workers should satisfy two spatial constraints: (i) tasks should be in the spatial region of assigned workers; (ii) aggregated reputation of workers should exceed a preset threshold of tasks.

On the other hand, the aggregation techniques in Sec. 4.2 can be combined with effective task assignment to further improve the quality of task completion. For example, in crowdsourced POI labeling, the authors divide the problem into label inference and task assignment [117]. In label inference, the accuracy of a label is determined by worker quality and POI influence. In task assignment, they use MLE to estimate the parameters mentioned above and the final results of labels. Then they adopt a greedy based algorithm which selects the assignment with maximum accuracy improvement for current workers. In [116], the speed estimation task is completed in two steps. The first step is task assignment which selects K roads that can best perform speed estimation. After obtaining the speeds of K roads, the second step is to infer the speed of other roads based on the these K roads.

5 Incentive Mechanism

Any crowdsourcing involves certain incentive mechanisms to attract active and qualified workers. Incentive mechanisms determine the rewards to workers such that more workers can be motivated to perform the tasks. Compared with the incentive mechanisms in traditional crowdsourcing, incentive mechanisms in spatial crowdsourcing not only need to attract the interests of workers (which is similar), but also to involve reliable workers to physically move to the location of tasks (which is unique). Since the locations of workers may change over time, the incentive mechanisms in spatial crowdsourcing also need to account for the spatiotemporal factors. In this section, we first introduce the commonly used evaluation metrics in the design of incentive mechanisms (Sec. 5.1). Next, we divide existing works into two

categories: **posted price models** (Sec. 5.2) and **auction based models** (Sec. 5.3). In posted price models, the platform first determines the reward for workers and workers can only accept it or not. Conversely, in auction-based models, workers can first submit their expected reward and the platform then determines the rewards to the workers afterwards. Finally, we compare existing studies in Sec. 5.4.

5.1 Evaluation Metrics

An incentive mechanism is assessed from two aspects, **algorithm metrics** and **mechanism metrics**.

Algorithm Metrics. In spatial crowdsourcing, an incentive mechanism is often an algorithm. Thus, the common algorithm metrics are also used to assess the efficiency and effectiveness of a mechanism.

- **Complexity.** Complexity analysis includes the running time and memory usage of the algorithm, which reflects the efficiency of an incentive mechanism. In particular, the *computational efficiency* of a mechanism represents whether the algorithm can be terminated in polynomial time.
- **Approximation/Competitive Ratio.** Approximation ratio and competitive ratio guarantee how bad an algorithm is compared with the optimal solution in the worst case in the offline scenario and the online scenario, which reflect the effectiveness of an incentive mechanism.

Mechanism Metrics. As a functional mechanism, an incentive mechanism should have the properties below.

- **Truthfulness.** A truthful mechanism guarantees that workers always submit the truthful information (*e.g.*, the expected reward based on his/her private evaluation) to the platform. In other words, they cannot obtain more revenue by submitting false information about themselves, where the revenue of a worker represents his/her reward minus his/her cost to perform the task.
- **Individual Rationality.** An individually rational mechanism guarantees that each participated worker will obtain a non-negative revenue, *i.e.*, the reward to the worker is no less than the cost of the worker to perform the task.
- **Budget Balance.** A budget-balanced mechanism guarantees that the total reward to workers does not exceed a given budget, *i.e.*, the mechanism does not need more budget from outside.

5.2 Posted Price Models

The posted price model is widely used in applications like taxi dispatching (*e.g.*, Uber [17]) and food delivery (*e.g.*, Meituan [26]). In this model, the platform determines the reward to the worker and the worker can only decide whether to accept the task or not. Incentive mechanisms following this model can be further divided into two types, *Supply-and-Demand-Aware Model* and *Quality-Aware Model*. In the first type, the rewards are mainly determined based on the comparison between supply (*i.e.*, the number of workers) and demand (*i.e.*, the number of tasks). In the second type, the rewards are mainly determined based on the quality of workers or the quality of tasks.

5.2.1 Supply-and-Demand-Aware Model

In spatial crowdsourcing applications, the **supply** (*i.e.*, the number of workers) and the **demand** (*i.e.*, the number of tasks) often vary in space and time [208]. The corresponding incentive mechanism should reflect the spatiotemporal dynamics between supply and demand. That is, the reward to the worker and the payment of the requester should be dynamic, *i.e.*, *dynamic pricing*. Compared with the traditional fixed price strategy (*i.e.*, static pricing), the incentive mechanisms based on this model are more likely to obtain higher total revenue, which has already been validated in real-world applications *e.g.*, the *surge pricing* in Uber [17].

In the model, a **base price** represents the long term unit price, which is usually determined based on prior knowledge of the markets. According to the dynamics of supply and demand, an incentive mechanism changes the unit reward on basis of the base price or the most recently used price.

A well-known adoption of this model is the *surge pricing* in Uber [17], which has been studied in [54, 61, 122, 138, 157]. Specifically, during times of high demand for rides, the unit fare may change by multiplying the base price with a multiplier accordingly to the incentive mechanism of *surge pricing*. Thus, the areas with higher multipliers usually indicate a steady stream of ride requests (*i.e.*, tasks), where drivers (*i.e.*, workers) will be attracted to come to. As a result, this incentive mechanism will eventually ensure that the pickup is quick and reliable. Experiments show that the surge pricing strategy not only reduces the waiting times of tasks, but also improves rewards for workers [157].

The supply-and-demand-aware model has also attracted extensive academic research.

Banerjee *et al.* [44] apply queuing theories to analyze the incentive mechanisms in ride sharing. They pro-

pose a *single-threshold* based dynamic pricing, where the unit fare for tasks reduces to a lower value if the number of workers is above the threshold. They find that the *single-threshold* dynamic pricing is robust and can be applied to find an optimal base price.

Both [45] and [60] apply Markov process to determine the fare to tasks and the reward to workers. Banerjee *et al.* [45] still assume that tasks appear on the platform following the queuing model and their pricing strategy is determined by Markovian transitions between independent state (*i.e.*, the distributions of workers on the platform). They obtain the approximate solution by relaxation techniques. Chen *et al.* [60] consider more spatiotemporal issues, *e.g.*, travel time and driver direction. They use Markov Decision Process (MDP) to formulate the problem, *i.e.*, the driver distribution on each vertex of graph as a state, the throughputs of tasks on each edge as actions, and the transitions between states as the revenue. Even though it is PSPACE-hard to solve MDPs, they design an polynomial-time algorithm to find an approximate result.

Differently, Tong *et al.* [211] use bipartite graphs to model the Global Dynamic Pricing (GDP) problem. They aim to find the optimal pricing strategy along with the task assignment. First, they propose a Myerson Reserve Price based algorithm to determine the base price for each urban area. Based on this base price, they further propose a matching based algorithm with an approximation ratio of $1 - 1/e \sim 0.632$ to dynamically adjust the unit price for each area according to the dynamics of supply and demand.

Other studies [39, 90, 185] focus on the incentive mechanisms specifically for ride sharing. Fang *et al.* [90] use *subsidies* to provide incentives to workers such that enough supplies can be ensured. Their experiments show that *subsidies* are effective to avoid supply shortages. Asghari *et al.* [39] take the future changes of supply and demand into consideration. Their intuition is that in regions where the supply is abundant, lowering the prices can lead to higher demand which in turn increases the number of requests.

Shen *et al.* [185] integrate the task planning into the design of incentive mechanisms in dynamic scenario. They develop an Integrated Online Ridesharing Mechanism (IORS), which satisfies desirable properties such as truthfulness, individual rationality, and budget balance. Their experiments show that compared to an auction-based mechanism [68] (which we will introduce later), IORS achieves a very close performance with substantially less computational time.

5.2.2 Quality-Aware Model

Sometimes tasks are expected to be accomplished with high quality, especially in applications like crowdsourced spatiotemporal data collection. The quality-aware model takes quality into account when providing incentives to workers. We focus on how to design effective incentives (*i.e.*, determine the reward to attract reliable workers), which is related to, but different from quality control in Sec. 4. According to the types of quality discussed in Sec. 4, we divide incentive mechanisms using quality-aware models into two types, *quality-of-worker-aware* [219, 226, 231] and *quality-of-task-aware* [151, 163]. Note that most of the studies above are under a reward budget constraint, *i.e.*, the total rewards of workers should not exceed the budget of the task.

Quality-of-Worker-aware. Studies of this type consider the reputation of workers [219, 231] or the willingness of workers in terms of spatial factors [226] when deciding the reward regarding the quality of workers.

Yu *et al.* [231] and Wang *et al.* [219] model the quality of workers with their reputation. They both assume that workers are classified into three kinds: *high* reputation, *medium* reputation or *low* reputation. The rewards of workers are determined by the reputation level, *i.e.*, the worker with higher reputation will obtain higher reward. However, a worker with low reputation will not be paid, since they assume the requester does not like to engage such a worker.

Wu *et al.* [226] consider the distance between workers and tasks. In general, workers prefer tasks nearby [173]. Therefore, in [226], extra remote subsidies should be paid if workers far away are selected. The subsidy increases linearly with the distance between the worker and the task but no higher than a threshold. The final reward for a worker consists of the base price (calculated with the local average payment per unit time), the subsidy, and the extra tips for more incentives.

Quality-of-Task-aware. Studies in this type consider the latency [163] or the spatial diversity [151] with regard to quality of tasks.

Mitsopoulou *et al.* [163] try to minimize the latency of tasks by incentive mechanisms. They propose an adaptive pricing policy. Specifically, workers will receive a penalty if they do not respond immediately, *i.e.*, workers providing responses with longer latency will get less reward, and the penalty increases with the latency. The parameters of the reward function can be tuned for every worker. So by adjusting the parameters, the platform can make more workers respond to the tasks, or make workers respond more quickly.

Liu *et al.* [151] provide incentives to workers in consideration of the spatial diversity. They study the case

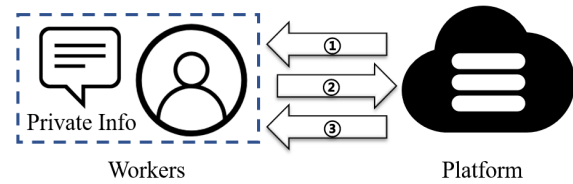


Fig. 6: Workflow of a basic auction model for incentive mechanism design in spatial crowdsourcing. Step (1): Announcement; Step (2): Bidding; Step (3): Rewarding.

where there is a task which needs to collect data from different places and propose a price adjustment function. This function allocates more money to the workers doing tasks in such places where data already collected is less than the expected amount. As a result, more workers will be attracted to such places, and the imbalance of data collection among different places can be mitigated. With enough data in each place required by the task, higher quality can be achieved.

Many real applications apply the idea of giving more rewards to remote places for more data, considering the spatial diversity. Pokémon Go [14] is among one of the most successful. In this gamification based spatial crowdsourcing platform, users can use their mobile phones to track and catch Pokémon (virtual monsters). According to a recent study [203], the platform can also collect the spatial data via GPS. By placing attractive Pokémons at different locations, the platform can stimulate players to go there. As a result, diversified and qualified data can be collected.

5.3 Auction based Model

Posted price models determine the reward to workers based on the *estimated* expectation of workers. When the estimation is wrong or unavailable, the reward might be improperly set. Auction based models overcome this disadvantage by permitting workers to bid the task with their own expectation (*i.e.*, private information including the reward they expect, etc.) and then determining the reward to the worker afterwards. We first introduce the workflow of the auction model in Sec. 5.3.1, and then review the representative works in Sec. 5.3.2.

5.3.1 Workflow

Fig. 6 illustrates the workflow of a basic auction model. It includes three steps.

- (1) **Announcement.** The platform first announces the task to the workers who are possible to complete it under the spatiotemporal constraints.

- (2) **Bidding.** After receiving the announcement from the platform, workers bid based on their private information (*e.g.*, submit their expected payment) to the platform. In this step, a worker can be strategic and selfish. Hence he may submit the fake information to earn higher reward (*i.e.*, an untruthful worker). The incentive mechanisms should guarantee the truthfulness of the worker.
- (3) **Rewarding.** The platform decides the reward according to the collected private information. Besides, the platform sometimes also needs to determine the task assignment along with the reward.

Next we review representative auction based incentive mechanisms. Since the first step of **announcement** is the same for most mechanisms, we mainly discuss the **bidding** and the **rewarding** procedures of different mechanisms as well as their performances.

5.3.2 Representative Auction based Mechanisms

We review auction based incentive mechanisms for two applications, ride sharing [37, 40, 237] and citizen sensing services [68, 244].

(i) **Auction based Incentives for Ride Sharing.** Both [40] and [37] focus on incentive mechanisms to maximize the total revenue of the platform, *i.e.*, total payment of requesters minus the total rewards to workers. They both use an auction based model to determine the reward to the worker along with the assignment of tasks. Specifically, after receiving the announcement from the platform, the worker will locally calculate the updated route which is the most profitable to complete this new task. Next, the worker will bid their expected reward and the calculated route to the platform. Finally, the platform will assign the task to the worker with the highest revenue from this task.

In [37, 40], the payment of requester is determined based on the calculated route of its assigned worker. In [40], the authors apply a first-price auction scheme (*i.e.*, to pay the highest reward to the worker with highest bid) to determine the reward to the worker while in [37] the researchers use the second-price auction scheme (*i.e.*, to pay the second highest reward to the worker with highest bid) to determine his/her reward. Since a second-price auction model can guarantee a few promising mechanism properties, the incentive mechanism used in [37], SPARP, is truthful and individually rational. Finally, they conduct the experiments on the real-world datasets of New York City's Taxi [16]. Experimental results show that [37] can obtain more total revenue than the first-price auction scheme.

Zhang *et al.* [237] adopt another auction model, *double auction*. In their incentive mechanism, both work-

ers and requesters provide their bids based on private information (*i.e.*, the expected reward to workers and the expected payment of requester) to the platform at the same time. After receiving the private information from the two sides of the market, the platform will make a task matching between workers and requesters, and determine the actual reward and the actual payment. They also design a discounted trade reduction mechanism to make a discount in both actual reward and actual payment, DTR, which is truthful, individually rational and budget balanced.

Cheng *et al.* [68] study the incentive mechanism design in the last-mile delivery service. In the step of bidding, the worker sends his/her direct travel distance and a compensation rate (*i.e.*, the cost in unit distance) to the platform. The authors devise the *bottom-up* mechanism to determine the actual reward along with routing plan. Their mechanism is truthful, individually rational and budget balanced.

(ii) **Auction based Incentives for Citizen Sensing.** Zhang *et al.* [239] propose an auction-based incentive mechanism in the online scenario. Specifically, each worker dynamically appears on the platform, and proposes his/her bid (including the expected reward) to the platform. The platform assigns the tasks to the selected workers and determines the reward to them, in order to maximize the total utility. Their incentive mechanism TOIM is computationally efficient, truthful, individually rational and profitable (*i.e.*, the platform will get a non-negative revenue from the mechanism).

Zhao *et al.* [244] also focus on the incentive mechanism design in the online scenario, with any monotone submodular [166] objective function and a budget constraint. In the bidding step, each worker submits his/her expected reward and the sets of tasks that he/she would like to accomplish. In the last step, the platform selects a subset of workers based on an adaptive threshold such that the total reward to these workers does not exceed the budget. They propose two incentive mechanisms, OMZ and OMG, to handle the cases when workers immediately leave the platform and the workers can stay for a time period, respectively. The mechanisms are truthful and individually rational with a competitive ratio of 0.25 and time complexity of $O(mn \min(m, n))$ at each time, where m and n are the number of requests and workers. To validate the effectiveness of the proposed mechanisms, they conduct the experiments in the Wi-Fi signal sensing application provided by [186]. The experimental results show that both OMZ and OMG achieve the approximate result as the offline optimal solution. In particular, OMZ is often better than OMG in terms of effectiveness.

Table 8: Comparison of representative studies on incentive mechanisms in spatial crowdsourcing.

Incentive Model	Reference	Time Complexity ¹	Ratio	Truthfulness	Individual Rationality	Budget Balance
auction	[40]	-	-	✗	✗	✗
auction	[37]	-	-	✓	✓	✓
auction	[237]	-	-	✓	✓	✓
auction	[68]	-	-	✓	✓	✓
auction	[239]	$O(\frac{1}{\epsilon} W ^3 \log(T))$	$O(\frac{1}{\epsilon})$	✓	✓	✗
auction	[244]	$O(T W \min(T , W))$	0.25	✓	✓	✓
quality-aware	[231]	-	-	✗	✗	✓
quality-aware	[219]	$O(W)$	-	✗	✗	✓
quality-aware	[226]	-	-	✗	✗	✓
quality-aware	[151]	$O(T \times (\# \text{ of Time Windows}))$	-	✗	✗	✓
quality-aware	[163]	-	-	✗	✗	✓
supply-and-demand-aware	[211]	$\frac{ G \log G + P \times \min(T , W)(\log G + E)}{}$	0.632	✗	✗	✗
supply-and-demand-aware	[44]	-	-	✓	✗	✗
supply-and-demand-aware	[60]	polynomial	-	✗	✗	✗
supply-and-demand-aware	[39]	$O(G ^3)$	-	✗	✗	✗
supply-and-demand-aware	[90]	-	-	✓	✗	✗

¹ In the column of time complexity, we use “-” to represent that the time complexity or ratio is not given in the paper. We use T and W to denote the set of tasks and the set of workers respectively.

² $|G|$ is the number of regions, $|P|$ is the number of discrete prices, and $|E|$ is the number of possible assigned pairs of tasks and workers.

5.4 Discussions

In summary, an incentive mechanism should motivate workers to participate in the tasks. In spatial crowdsourcing, different workers may have different interest in the task because of the variable spatial and temporal information of workers and tasks. Thus it has become a challenge how to design the incentive mechanism for spatial crowdsourcing.

An incentive mechanism is assessed using two types of metrics, *i.e.*, algorithm metrics and mechanism metrics. As shown in Table 8, most efforts focus on the algorithm metrics (especially the time complexity) of their mechanisms. Although many incentive mechanisms are computationally efficient (*i.e.*, able to terminate in polynomial time), the spatiotemporal dynamics may raise a real-time requirement for practical incentive mechanism design. Mechanism metrics are emphasized more in auction models than in posted price models. This is because the auction based model considers the participatory of the workers before pricing for workers, and as a result requires the mechanism metrics to guarantee the robustness.

Besides, the formulation of the incentive models notably varies even for the applications (*e.g.*, for taxi dispatching [45, 60, 211]). Hence it seems necessary to come up with a unified formulation such that the proposed incentive mechanisms can be fairly compared in

terms of effectiveness, efficiency and flexibility. Furthermore, many existing works focus on maximizing the revenue in short-term and it is still open how to design an incentive mechanism for the long-term revenue.

Finally, it is worth mentioning that there is another successful incentive besides the aforementioned ones: *volunteered based incentive*. In practice, when the scale of the whole task is large (*e.g.*, editing the whole map of world), it usually requires a large number of workers which often leads to a extremely high payment. Thus, a practical and efficient way to complete such tasks is to get help from volunteer workers [85]. For example, one of the biggest volunteer based community in spatial crowdsourcing is the Humanitarian OpenStreetMap Team (HOT) [23]. Since its foundation in 2010, HOT has already had 170,252 registered volunteers and together completed 1,933,608 tasks related to environmental and societal issues (*e.g.*, disaster response and risk reduction). The motivations of these volunteers are either contributing to the greater good (*e.g.*, users in HOT) or gaining something by taking part (*e.g.*, drivers in Waze [19]). However, rather than the algorithmic/theoretic aspects of incentive mechanisms, existing works on volunteer-based incentives usually focus on the supporting tool designs to attract volunteers [85, 137], which is not the major concern in this survey. We refer readers to [85, 136, 137, 152] on im-

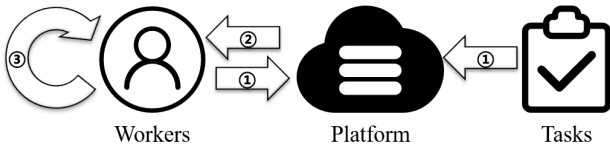


Fig. 7: Workflow of privacy protection for task assignment in spatial crowdsourcing. Step (1): transformation; Step (2): assignment; Step (3): refinement.

important issues in supporting tool designs for volunteer-based incentive mechanisms.

6 Privacy Protection

As in traditional web-based crowdsourcing, privacy is an important concern in spatial crowdsourcing. One particular interest in spatial crowdsourcing is to protect the location information of tasks and workers (and certain intermediate results) so that spatiotemporal tasks can be released and performed without exposing the physical locations of tasks and workers to malicious users. Overall, privacy protection research in spatial crowdsourcing is dedicated to design privacy-preserving frameworks and techniques compatible for the core issues in spatial crowdsourcing (*e.g.*, task assignment [171]).

6.1 Generic Framework

Most studies on privacy protection in spatial crowdsourcing focus on privacy-preserving task assignment. A generic privacy-preserving framework for task assignment in spatial crowdsourcing consists of three steps. Fig. 7 shows its workflow.

- (1) **Transformation.** The locations of workers and (or) tasks are transformed by some techniques.
- (2) **Assignment.** The spatial crowdsourcing platform performs task assignment based on the transformed locations of workers and (or) tasks.
- (3) **Refinement.** Workers confirm or refine the task assignment results based on their true locations.

Depending on the location transformation techniques and the assumptions on trusted parties, the step of refinement may be omitted. Furthermore, some privacy protection schemes may involve auxiliary trusted servers. In the context of spatial crowdsourcing, the spatial crowdsourcing platform (the platform for short) is usually assumed to be untrusted.

Below we review representative studies that exploit three categories of transformation techniques: *spatiotemporal cloaking*, *differential privacy* and *encryption*.

6.2 Spatiotemporal Cloaking based Transformation

Spatiotemporal cloaking protects location privacy by hiding the locations inside a cloaked region.

In [217], the locations of workers are first submitted to an extra trusted server. Then, the trusted server constructs a cloaked region around the worker’s actual location for each worker based on locality-sensitive hashing (LSH) [76], where both K -anonymity [192] and locality are preserved. The untrusted spatial crowdsourcing platform can only access the above transformed spatial cloak of each worker. Then an algorithm is devised for searching the k -nearest tasks of a worker with the help of the refinement by the trusted server, based on which task assignment can be performed.

In [130, 131], the authors assume that the workers trust each other but do not trust the spatial crowdsourcing platform. Each worker calculates his/her Voronoi cell in a distributed manner and forms the spatial cloak. Then a voting mechanism is designed through which a set of representative participants are selected whose cloaked regions should be sent out to the spatial crowdsourcing platform for querying the nearest tasks, during with K -anonymity is preserved. These query results will later be shared with the rest of the workers. As a result, all the tasks are assigned to the nearest workers.

In [170], instead of a spatiotemporal point, each worker submits a cloaked area including a spatiotemporal region a and the probability density function f of the worker at each point in a . Based on the cloaked locations of workers and exact locations of tasks, the spatial crowdsourcing platform performs uncertain task pre-assignment via the expected distances between the tasks and workers. The pre-assignment results are sent to the service of the workers and refined according to the exact locations of the workers. Based on the methods in [170], [172] proposes a demo of a location-based mobile Q&A application.

6.3 Differential Privacy based Transformation

Differential privacy [86] is a general-purpose approach for privacy protection and has emerged as the de facto standard for private data release. It releases data of a group such that what can be learned from the released data does not substantially differ regardless of the inclusion of a given individual’s data [87]. Below are some works with different implementations of differential privacy for task assignment in spatial crowdsourcing.

In [197], the locations of workers are first submitted to a trusted server and then transformed via *Private Spatiotemporal Decomposition (PSD)* proposed in

[71]. A PSD is a spatiotemporal index transformed according to differential privacy, where each index node is obtained by releasing a noisy count of the data points enclosed by that node's extent. The spatial crowdsourcing platform then performs task assignment as follows. For each task t the platform first queries the PSD released by the trusted server for a region where there are workers near t with a high probability. Then the platform geocasts the information of t to the workers in this region. The workers who are willing to perform t send a consent message back to the platform. The method can be also extended to dynamic data [201].

In [124, 125], the authors investigate privacy protection in crowdsourced dynamic spectrum sensing. The locations of workers are not transformed directly. Instead, the bids provided by all the workers are transformed, which represent the worker's cost for spectrum sensing and are closely tied to the worker's current locations. The transformation to protect differential privacy is based on the exponential mechanism [160]. Then task assignment is modeled as a reserve auction problem and privacy-preserving methods for worker selection are proposed based on the transformation.

In [218], the authors study privacy protection in crowdsourced urban sensing. To transform the locations of workers, the spatial crowdsourcing platform first provides an obfuscation matrix and a data adjustment function designed via differential privacy. The location information is transformed by the obfuscation matrix which encodes the probabilities of obfuscating any one region to another. The corresponding data is transformed by the data adjustment function. Note that the platform cannot obtain the original data although it provides the obfuscation matrix and data adjustment function. After the transformed locations and sensory data are uploaded to the spatial crowdsourcing platform, the platform can infer the distribution of the data in the sensed regions from the transformed data.

In [202], the authors propose a privacy-preserving one-sided online task assignment scheme where tasks appear dynamically. The locations of workers are transformed by Geo-indistinguishability (Geo-I) [35], which is a notion of location privacy based on differential privacy, and are submitted to the spatial crowdsourcing platform in advance. Once a task appears, the transformed location is submitted to the spatial crowdsourcing platform. Subsequently, the spatial crowdsourcing platform identifies a set of candidate workers who are most capable to perform the task and sends the information of these workers to the requester of the task. Finally, the requester of the task connects the workers one by one bypassing the spatial crowdsourcing platform and share the exact location to refine whom can com-

plete the task. The authors test their privacy-preserving schemes on a taxi datasets called T-Drive [232]. Their experimental results show the proposed techniques, algorithms, and heuristics achieve high effectiveness and low disclosure of the location information.

6.4 Encryption based Transformation

This line of research applies specific encryption techniques on the locations of tasks and workers such that particular calculation can still be performed on the encrypted data, *e.g.*, calculating the distance between two encrypted positions. Thus tasks can be assigned based on the calculation results.

In [150], the locations of tasks and workers are encrypted via a Paillier cryptosystem [169]. According to Paillier cryptosystem's characteristics, the exact distance between two encrypted positions can be calculated without releasing the plain data. Hence the distances between the tasks and workers can be obtained during task assignment. However the adoption of the Paillier cryptosystem significantly increases the computation overhead. Thus the authors in [150] extend KD-tree to SKD-tree in order to prune unnecessary Paillier cryptosystem query or computation.

In [149], the authors take the velocity of workers into consideration and propose a Paillier cryptosystem [169] and ElGamal cryptosystem [95] based privacy protection method. Specifically, the locations of tasks and workers are encrypted and the distance is computed in a privacy-preserving way. The velocity of workers is also encrypted and the travel time for a worker to move to the position of a task can also be computed with privacy preservation. Thus each task can be assignment to the worker with the minimum travel time.

In [148], researchers encrypt the locations of tasks and workers and calculate the distance secretly via the Paillier cryptosystem [169]. Distance comparison is performed via the Yao's protocol [146, 230], and each task is assigned to the nearest worker. To improve the efficiency, Geohash [7] is adopted to find the nearest workers approximately.

In [175, 234], the information of tasks including their positions are posed in the spatial crowdsourcing platform in advance and thus are not protected. The workers can browse the information and submit their travel costs for different tasks instead of their exact positions to the platform. The travel cost is encrypted into perturbed data via Bitwise XOR homomorphic cipher system [242]. After receiving all the perturbed data from the workers, the platform assigns tasks through a reverse auction based algorithm.

Table 9: Comparison of privacy protection techniques in spatial crowdsourcing.

Transformation Techniques	Reference	Protected Components	Arrival Scenario	Assignment Goal	Without Extra Server	Computation Overhead
spatiotemporal cloaking	[217]	workers	static	nearest	✗	low
	[130, 131]	workers	static	nearest	✓	low
	[170]	workers	static	minimizing total cost	✓	low
differential privacy	[197]	workers	static	maximizing total number	✗	low
	[201]	workers	dynamic	maximizing total number	✗	low
	[124, 125]	workers	static	minimizing total cost	✓	low
	[218]	workers	static	quality of sensed data	✓	low
	[202]	tasks & workers	dynamic	maximizing total number	✓	low
encryption	[150]	tasks & workers	static	nmaximizing total number & minimizing total cost	✗	high
	[149]	tasks & workers	static	nearest	✗	high
	[148]	tasks & workers	static	nearest	✗	high
	[234]	workers	static	maximizing total utility	✗	high

6.5 Discussions

Table 9 shows representative studies on privacy protection using different transformation techniques. We summarize their main characteristics below.

- **Protected Components.** Spatial cloaking based and most differential privacy based methods only protect workers. The reasons are two fold. (i) The location privacy of workers is more sensitive than that of tasks. (ii) It is non-trivial to extend spatial cloaking and differential privacy based transformation techniques to the cases where two components of data need to be protected.
- **Applicable Task Assignment Categories.** Most existing privacy-preserving task assignment schemes only apply to static arrival scenarios. Protecting dynamic location information is more difficult since many transformation techniques require all data to be protected to be known in advance. Since most studies integrate privacy protection into the task assignment framework, the applicable assignment goals are constrained by the privacy protection methods. For example, in [130, 131, 148, 149, 217] the goal is simply to assign the tasks to the nearest workers. For other more complex goals, the assignment methods are closely coupled with the privacy protection methods and most of them are heuristic.

- **Overhead.** It adds extra overhead to protect privacy in task assignment. Encryption based methods are often more computation-intensive than spatial cloaking and differential privacy based methods, due to the high cost of encryption/decryption and the need for extra trusted servers for key distribution.
- **Trade-off.** The impact of enforcing privacy protection on the locations of workers and/or tasks depends on the specific privacy-preserving technique. If the transformation is performed through spatiotemporal cloaking or differential privacy, the effect is mainly on the number of assigned pairs of tasks and workers, as some pairs which satisfies the range constraint may not satisfy it anymore after their locations are transformed. However, the quality of each assigned pair is not affected due to the refinement step. If the transformation is performed through encryption, the quality is not impacted as all the calculation is exact. However, the low efficiency of encryption is its main drawback.

In summary, a practical privacy-preserving task assignment scheme should protect at least the locations of workers. Privacy protection brings extra overhead and constraints to task assignment in spatial crowdsourcing.

In addition to task assignment, privacy protection is also necessary in other core issues in spatial crowdsourcing. For example, privacy protection is combined with quality control in [218], where an inference algo-

rithm is designed to improve the inference accuracy for the data transformed according to differential privacy. In [123], privacy protection is combined with incentive mechanism. Specifically, a reverse auction based incentive mechanism is designed when considering the privacy requirement of different workers. Then, the data collected via spatial crowdsourcing is published after transformation through differential privacy.

7 Applications

Spatial crowdsourcing is closely tied to the physical world and there have been various real-world applications. This section summarizes typical spatial crowdsourcing applications into two categories: *sharing economy based urban services* and *crowdsourced spatiotemporal data collection*.

7.1 Sharing Economy based Urban Services

Sharing economy based urban services refer to applications such as delivery and on-site services crowdsourced to freelancers. In the context of spatial crowdsourcing, a task in these applications is usually served by a single worker and thus often involves no explicit quality control (*i.e.*, result aggregation). Popular applications include on-demand taxi dispatching, ride sharing, food delivery and on-site micro services.

On-demand Taxi Dispatching. It is one of the earliest successful spatial crowdsourcing applications. Passengers appear dynamically and submit taxi requests to platforms such as Uber [17] and Didi Chuxing [4]. The platform assigns taxis to passengers in real time to pick up passengers. Hence in terms of task assignment, on-demand taxi dispatching can be modeled as a dynamic matching problem with diverse objectives such as maximizing the total payoff or minimizing the average latency of passengers.

Ride Sharing. This application is an emerging extension of on-demand taxi dispatching service often provided by the same companies, *e.g.*, Uber [17] and Didi Chuxing [4]. The key issue of ride sharing is to schedule a route, which consists of a sequence of pickup locations and delivery locations for each passenger to minimize the total travel cost of the drivers (*i.e.*, workers) [158] or the average latency of the passengers (*i.e.*, requesters) [228]. In terms of task assignment, ride sharing is often modeled as a dynamic planning problem [40, 119, 158, 212].

Food Delivery. Food delivery services such as Grubhub [10] and Meituan [26] are similar to ride sharing in

terms of task assignment. Customers dynamically submit food delivery requests to the platform. The platform then determines the price of the delivery requests for the requesters and the schedules of the delivery requests for the couriers. Similarly, food delivery services are often formulated as dynamic planning [154].

On-site Micro Services. On-site micro services are another successful adoption of spatial crowdsourcing. Platforms such as TaskRabbit [15] and Gigwalk [8] connect various domestic services, *e.g.*, house cleaning, with freelancers. Similar to on-demand taxi dispatching, task assignment in on-site micro services can be considered as a dynamic matching problem.

Discussions. Sharing economy for urban services often deal with highly dynamic data at urban scale. To provide better quality of services, more efficient and effective task assignment algorithms are needed. Sharing economy for urban services usually apply the incentive mechanisms based on supply-and-demand-aware models and provide certain degree of privacy protection. Nevertheless, there is a growing trend to introduce additional incentive mechanisms into these applications to consistently attract more users.

7.2 Crowdsourced Spatiotemporal Data Collection

Crowdsourced spatiotemporal data collection refers to applications that crowdsource collection of various spatiotemporal information to citizens. In the context of spatial crowdsourcing, a task in these applications is usually performed by multiple workers and involves certain spatiotemporal data processing. Tasks in this category vary in real-time requirement and degree of spatiotemporal data processing, but all involve quality control to aggregate highly qualified results.

Crowdsourced Event Detection and Labelling. It is natural to crowdsource detecting and labelling of urban spots or events to citizens. For instance, residents can contribute to POI labelling in the neighbourhood [117]. They can also report noise pollution [189], air pollution [109] and weather conditions [200] in the vicinity. Since such data are normally provided by unprofessional workers using noisy sensors, it is crucial to aggregate sensory data from workers to control the quality of the detection or labelling tasks. Truth inference is commonly used for quality control in crowdsourced event detection and labelling [162, 168].

Crowdsourced Map Applications. Spatial crowdsourcing can also be applied in more complex spatiotemporal data collection and processing such as map generation, real-time traffic speed estimation and road navigation. For example, OpenStreetMap [107] is already

Table 10: Categories of core issues in typical spatial crowdsourcing applications.

Category		Reference	Task Assignment	Quality Control	Incentive Mechanism	Privacy Protection ¹	
sharing economy based urban services	taxi dispatching	[4, 17]	dynamic matching	✗	supply-and-demand-aware	✓	
	ride sharing	[158, 212]	dynamic planning	✗	supply-and-demand-aware	✓	
	food delivery	[10, 26]	dynamic planning	✗	supply-and-demand-aware	✓	
	on-site micro services	[8, 15, 164]	dynamic matching	✗	supply-and-demand-aware	✓	
spatiotemporal data collection	event detection and labelling	POI labelling	[117]	static matching	expectation maximization	✗	✗
		pollution detection	[109, 189]	static matching	aggregated diversity	✗	✓
		event detection	[30, 168]	dynamic matching	bayesian estimation	✗	✗
	map application	map generation	[62, 107]	dynamic matching	aggregated diversity	quality-aware	✓
		speed estimation	[116, 155]	static matching	bayesian estimation	✗	✗
		road navigation	[19, 89]	static planning	aggregated diversity	✗	✗
		congestion alert	[36]	dynamic matching	expectation maximization	✗	✗
		path selection	[190, 235]	static matching	expectation maximization	✗	✗

¹ Some applications claim that privacy protection is considered but the detailed techniques are not specified.

the world’s largest crowdsourced mapping project that creates a free and collaboratively editable map of the world. Real-time traffic speed in a map can be inferred by crowdsourcing speed estimation of a portion of seed roads and jointly considering historical speed information [116, 155]. Crowdsourced road navigation is viable by collecting real-time traffic information, *e.g.*, using Waze [19] and constructing a landmark scoring model for route recommendation [89]. Some other functions in map applications, such as alerting traffic congestion [36] or answering path selection queries [190, 235] could also be crowdsourced by consulting nearby drivers and picking out desirable answers. Quality control in these applications is dedicated and sometimes is coupled with the underlying spatiotemporal data processing process.

Discussions. Table 10 summarizes the aforementioned applications in spatial crowdsourcing. Compared with the sharing economy based urban services, task assignment in crowdsourced spatiotemporal data collection depends on the specific data to collect and varies in the models. It can be formulated as static matching (*e.g.*, POI labelling [117]), static planning (*e.g.*, road navigation [89]), or dynamic matching (*e.g.*, map generation [107]). It is important for crowdsourced spatiotemporal data collection applications to attract the highly qualified workers. Hence the incentive mechanisms based on the quality-aware models are used to

motivate workers. While some pioneer studies have proposed privacy protection schemes for certain applications in this category (*e.g.*, pollution detection [162] and map generation [62]), it is unclear whether privacy protection methods suited for other applications have been designed.

8 Open Problems

In this section, we discuss some important open problems in spatial crowdsourcing.

More Effective Task Assignment Algorithms. Task assignment is central to spatial crowdsourcing, yet its effectiveness is still not satisfactory for many real-world applications. Particularly, emerging applications such as on-demand taxi dispatching and ride sharing require highly effective dynamic matching and planning algorithms. Yet the competitive ratios of the state-of-the-art algorithms for dynamic matching are often no higher than 0.5 unless under some strong assumptions (*e.g.*, one-sided [121], arrival rate [120]). It also seems hard to propose competitive solutions to dynamic task planning under extreme cases. In particular, the worst cases to prove the hardness result are usually impractical, *e.g.*, with the extremely short deadline [144, 193]. Thus, existing studies (*e.g.*, [119, 158, 193]) usually propose heuristics without any theoretical guarantee. One op-

portunity to improve the effectiveness of dynamic task assignment is that practical applications may not strictly require instant assignments. Therefore, it may be feasible to wait for a reasonably short period and make global assignments on a batch basis. However, it remains open how to theoretically select the best single batch or adapt the batch size in real-time to notably improve the effectiveness of task assignment algorithms.

Indices for Spatial Crowdsourced Data. Efficient spatial crowdsourcing requires not only efficient algorithms but also efficient data structures, *e.g.*, indices. Indices for spatial crowdsourced data need to be optimized for spatial queries and frequent updates. Some indices (*e.g.*, grid, R-tree [106], quadtree [178] and k -d tree [49]) are proposed for spatial queries. Others are proposed to handle the dynamics of spatiotemporal data, such as 3D R-tree [194], HR-tree [165] and TPR-tree [177]. Recently, Jonathan *et al.* [126] exploit a pyramid multi-resolution index to speed up the retrieval of workers in a given area. However, dedicated spatiotemporal indices are overlooked in existing spatial crowdsourcing algorithms. It is largely unexplored how to select or design suitable spatiotemporal indices and co-optimize the end-to-end efficiency of spatial crowdsourcing algorithms.

Benchmarks for Spatial Crowdsourcing. Standardized benchmarks are important for the continuous development of spatial crowdsourcing research. There have been many benchmarks for classical spatial data management. For example, DIMACS Implementation Challenge provides a set of benchmark instances for various shortest path problems. However, there is a lack of similar benchmarks for spatial crowdsourcing. Although there are a few synthetic data generators for spatial crowdsourcing [199], the lack of public real-world datasets still presents a challenge to the development of spatial crowdsourcing. The reasons of such a quandary are twofold. First, the owners of real data are usually commercial platforms that are not willing to share their data. Secondly, although there are open-source spatial crowdsourcing platforms such as gMission [63] and MediaQ [25], they cannot collect large amounts of data due to their limited scales.

9 Conclusion

In this paper, we surveyed the state-of-the-art research on spatial crowdsourcing, with comprehensive comparisons between spatial crowdsourcing and general-purposed crowdsourcing in terms of challenges and techniques. We summarized existing literature on spatial crowdsourcing algorithms into four categories: task assign-

ment, quality control, incentive mechanism design and privacy protection. Particularly, for task assignment, we reviewed matching and planning models in static and dynamic scenarios; for quality control, we discussed quality models of tasks/workers and result aggregation techniques; for incentive mechanism design, we presented posted price models and auction based models; for privacy protection, we offered a general privacy protection framework and compared three types of data transformation techniques. In addition, we studied emerging representative spatial crowdsourcing applications and explained how they are enabled by these techniques. Finally, we identified some open problems for future research in this active research area. We envision this survey as a timely reference and guideline for researchers and practitioners in spatial crowdsourcing.

Acknowledgements We are grateful to anonymous reviewers for their constructive comments. Yongxin Tong's work is partially supported by the National Science Foundation of China (NSFC) under Grant No. 61822201, U1811463 and 71531001, Science and Technology Major Project of Beijing under Grant No. Z171100005117001 and Didi Gaia Collaborative Research Funds for Young Scholars. Yuxiang Zeng and Lei Chen's works are partially supported by the Hong Kong RGC CRF C6030-18G Project, the National Science Foundation of China (NSFC) under Grant No. 61729201, Science and Technology Planning Project of Guangdong Province, China, No. 2015B010110006, Hong Kong ITC ITF grants ITS/212/16FP and ITS/470/18FX, Didi-HKUST joint research lab project, Microsoft Research Asia Collaborative Research Grant and Wechat Research Grant. Cyrus Shahabi's work has been funded in part by NSF grants IIS1320149 and CNS-1461963, the USC Integrated Media Systems Center (IMSC), and unrestricted cash gifts from Google and Oracle. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of any of the sponsors such as the National Science Foundation.

References

1. (2016) Datatang Taxi Dataset. URL <http://www.datatang.com/data/45888>
2. (2018) Amazon Mechanical Turk. URL <https://www.mturk.com/>
3. (2018) Cainiao. URL <https://www.cainiao.com/>
4. (2018) DiDi Chuxing. URL <https://www.didiglobal.com/>
5. (2018) Facebook Editor. URL <https://www.facebook.com/editor>
6. (2018) FedEx. URL <https://www.fedex.com/>
7. (2018) Geohash. URL <https://en.wikipedia.org/wiki/Geohash>
8. (2018) Gigwalk. URL <http://www.gigwalk.com>

9. (2018) gMission Dataset Generator. URL <https://github.com/gmission/SCDataGenerator>
10. (2018) GrubHub. URL <https://www.grubhub.com/>
11. (2018) InterestingSport. URL <http://www.quyundong.com/>
12. (2018) Nanguache. URL <http://www.nanguache.com/>
13. (2018) OpenStreetMap. URL <https://www.openstreetmap.org/>
14. (2018) Pokémon Go. URL <https://www.pokemongo.com/>
15. (2018) TaskRabbit. URL <http://www.taskrabbit.com>
16. (2018) TLC Trip Record Data. URL http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml
17. (2018) Uber. URL <https://www.uber.com/>
18. (2018) UPS. URL <https://www.ups.com/>
19. (2018) Waze. URL <http://www.waze.com/>
20. (2019) CPLEX. URL <https://www.ibm.com/analytics/cplex-optimizer>
21. (2019) Didi Chuxing Corporate Citizenship Report. URL <https://www.didiglobal.com/about-didi/responsibility>
22. (2019) GAIA Open Dataset. URL <https://outreach.didichuxing.com/research/opensource>
23. (2019) Humanitarian OpenStreetMap Team. URL <https://www.hotosm.org/>
24. (2019) keepright. URL <https://www.keepright.at/>
25. (2019) MediaQ. URL <http://mediaq.usc.edu/>
26. (2019) Meituan. URL <https://www.meituan.com/>
27. (2019) Seamless. URL <https://www.seamless.com>
28. (2019) Upwork. URL <https://www.upwork.com/>
29. (2019) Wikimapia. URL <https://www.wikimapia.org/>
30. Agapie E, Teevan J, Monroy-Hernández A (2015) Crowdsourcing in the field: A case study using local crowds for event reporting. In: Proceedings of the 3rd AAAI Conference on Human Computation and Crowdsourcing, pp 2–11
31. Aggarwal G, Goel G, Karande C, Mehta A (2011) Online vertex-weighted bipartite matching and single-bid budgeted allocations. In: Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms, pp 1253–1264
32. Ahuja RK, Magnanti TL, Orlin JB (1993) Network flows - theory, algorithms and applications. Prentice Hall
33. Alfarrarjeh A, Emrich T, Shahabi C (2015) Scalable spatial crowdsourcing: A study of distributed algorithms. In: 16th IEEE International Conference on Mobile Data Management, pp 134–144
34. Amsterdamer Y, Milo T (2014) Foundations of crowd data sourcing. SIGMOD Record 43(4):5–14
35. Andrés ME, Bordenabe NE, Chatzikokolakis K, Palamidessi C (2013) Geo-indistinguishability: differential privacy for location-based systems. In: 2013 ACM SIGSAC Conference on Computer and Communications Security, pp 901–914
36. Artikis A, Weidlich M, Schnitzler F, Boutsis I, Liebig T, Piatkowski N, Bockermann C, Morik K, Kalogeraki V, Marecek J, Gal A, Mannor S, Gunopulos D, Kinane D (2014) Heterogeneous stream processing and crowdsourcing for urban traffic management. In: Proceedings of the 17th International Conference on Extending Database Technology, pp 712–723
37. Asghari M, Shahabi C (2017) An on-line truthful and individually rational pricing mechanism for ride-sharing. In: Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp 7:1–7:10
38. Asghari M, Shahabi C (2017) On on-line task assignment in spatial crowdsourcing. In: 2017 IEEE International Conference on Big Data, pp 395–404
39. Asghari M, Shahabi C (2018) Adapt-pricing: a dynamic and predictive technique for pricing to maximize revenue in ridesharing platforms. In: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp 189–198
40. Asghari M, Deng D, Shahabi C, Demiryurek U, Li Y (2016) Price-aware real-time ride-sharing at scale: an auction-based approach. In: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp 3:1–3:10
41. Ashlagi I, Azar Y, Charikar M, Chiplunkar A, Geri O, Kaplan H, Makhijani RM, Wang Y, Wattenhofer R (2017) Min-cost bipartite perfect matching with delays. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2017), pp 1:1–1:20
42. Azar Y, Fanani AJ (2018) Deterministic min-cost matching with delays. In: 16th International Workshop on Approximation and Online Algorithms, pp 21–35
43. Azar Y, Chiplunkar A, Kaplan H (2017) Polylogarithmic bounds on the competitiveness of min-cost

- perfect matching with delays. In: Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms, pp 1051–1061
44. Banerjee S, Johari R, Riquelme C (2015) Pricing in ride-sharing platforms: A queueing-theoretic approach. In: Proceedings of the 16th ACM Conference on Economics and Computation, p 639
 45. Banerjee S, Freund D, Lykouris T (2017) Pricing and optimization in shared vehicle systems: An approximation framework. In: Proceedings of the 2017 ACM Conference on Economics and Computation, p 517
 46. Bansal N, Buchbinder N, Gupta A, Naor J (2014) A randomized $o(\log^2 k)$ -competitive algorithm for metric bipartite matching. *Algorithmica* 68(2):390–403
 47. Bar-Yehuda R, Bendel K, Freund A, Rawitz D (2004) Local ratio: A unified framework for approximation algorithms. *ACM Computing Surveys* 36(4):422–463
 48. Bei X, Zhang S (2018) Algorithms for trip-vehicle assignment in ride-sharing. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, pp 3–9
 49. Bentley JL (1975) Multidimensional binary search trees used for associative searching. *Communications of the ACM* 18(9):509–517
 50. Birnbaum BE, Mathieu C (2008) On-line bipartite matching made simple. *SIGACT News* 39(1):80–87
 51. Brubach B, Sankararaman KA, Srinivasan A, Xu P (2016) New algorithms, better bounds, and a novel model for online stochastic matching. In: 24th Annual European Symposium on Algorithms, pp 24:1–24:16
 52. Burkard RE, Dell’Amico M, Martello S (2009) *Assignment problems*. Springer
 53. Cao CC, She J, Tong Y, Chen L (2012) Whom to ask? jury selection for decision making tasks on micro-blog services. *PVLDB* 5(11):1495–1506
 54. Castillo J, Knoepfle D, Weyl G (2017) Surge pricing solves the wild goose chase. In: Proceedings of the 2017 ACM Conference on Economics and Computation, pp 241–242
 55. Chen C, Cheng S, Misra A, Lau HC (2015) Multi-agent task assignment for mobile crowdsourcing under trajectory uncertainties. In: Proceedings of the 14th International Conference on Autonomous Agents and MultiAgent Systems, pp 1715–1716
 56. Chen J, Zipf A (2017) Deepvgi: Deep learning with volunteered geographic information. In: Proceedings of the 26th International Conference on World Wide Web Companion, pp 771–772
 57. Chen L, Shahabi C (2016) Spatial crowdsourcing: Challenges and opportunities. *IEEE Data Engineering Bulletin* 39(4):14–25
 58. Chen L, Lee D, Zhang M (2014) Crowdsourcing in information and knowledge management. In: Proceedings of the 23rd ACM International Conference on Information and Knowledge Management
 59. Chen L, Lee D, Milo T (2015) Data-driven crowdsourcing: Management, mining, and applications. In: 31st IEEE International Conference on Data Engineering, pp 1527–1529
 60. Chen M, Shen W, Tang P, Zuo S (2018) Optimal vehicle dispatching for ride-sharing platforms via dynamic pricing. In: Companion of the The Web Conference, pp 51–52
 61. Chen MK (2016) Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform. In: Proceedings of the 2016 ACM Conference on Economics and Computation, p 455
 62. Chen X, Wu X, Li X, Ji X, He Y, Liu Y (2016) Privacy-aware high-quality map generation with participatory sensing. *IEEE Transactions on Mobile Computing* 15(3):719–732
 63. Chen Z, Fu R, Zhao Z, Liu Z, Xia L, Chen L, Cheng P, Cao CC, Tong Y, Zhang CJ (2014) gMission: A general spatial crowdsourcing platform. *PVLDB* 7(13):1629–1632
 64. Chen Z, Cheng P, Zeng Y, Chen L (2019) Minimizing maximum delay of task assignment in spatial crowdsourcing. In: 35th IEEE International Conference on Data Engineering, pp 1454–1465
 65. Cheng P, Lian X, Chen Z, Fu R, Chen L, Han J, Zhao J (2015) Reliable diversity-based spatial crowdsourcing by moving workers. *PVLDB* 8(10):1022–1033
 66. Cheng P, Lian X, Chen L, Han J, Zhao J (2016) Task assignment on multi-skill oriented spatial crowdsourcing. *IEEE Transactions on Knowledge and Data Engineering* 28(8):2201–2215
 67. Cheng P, Jian X, Chen L (2018) An experimental evaluation of task assignment in spatial crowdsourcing. *PVLDB* 11(11):1428–1440
 68. Cheng S, Nguyen DT, Lau HC (2014) Mechanisms for arranging ride sharing and fare splitting for last-mile travel demands. In: Proceedings of the 13th International Conference on Autonomous Agents and MultiAgent Systems, pp 1505–1506
 69. Chittilappilly AI, Chen L, Amer-Yahia S (2016) A survey of general-purpose crowdsourcing techniques. *IEEE Transactions on Knowledge and Data Engineering* 28(9):2246–2266
 70. Chuang T, Deng D, Hsu C, Lemmens R (2013) The one and many maps: participatory and tem-

- poral diversities in openstreetmap. In: Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, pp 79–86
71. Cormode G, Procopiuc CM, Srivastava D, Shen E, Yu T (2012) Differentially private spatial decompositions. In: 28th IEEE International Conference on Data Engineering, pp 20–31
 72. Corral A, Manolopoulos Y, Theodoridis Y, Vassilakopoulos M (2000) Closest pair queries in spatial databases. In: Proceedings of the 2000 ACM International Conference on Management of Data, pp 189–200
 73. Costa CF, Nascimento MA (2018) In-route task selection in crowdsourcing. In: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp 524–527
 74. Cranshaw J, Toch E, Hong JI, Kittur A, Sadeh NM (2010) Bridging the gap between physical location and online social networks. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing, pp 119–128
 75. Das A, Gollapudi S, Kim A, Panigrahi D, Swamy C (2018) Minimizing latency in online ride and delivery services. In: Proceedings of the 27th International Conference on World Wide Web, pp 379–388
 76. Datar M, Immorlica N, Indyk P, Mirrokni VS (2004) Locality-sensitive hashing scheme based on p -stable distributions. In: Proceedings of the 20th ACM Symposium on Computational Geometry, pp 253–262
 77. Demartini G, Difallah DE, Cudré-Mauroux P (2012) Zencrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. In: Proceedings of the 21st International Conference on World Wide Web, pp 469–478
 78. Deng D, Shahabi C, Demiryurek U (2013) Maximizing the number of worker’s self-selected tasks in spatial crowdsourcing. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp 314–323
 79. Deng D, Shahabi C, Zhu L (2015) Task matching and scheduling for multiple workers in spatial crowdsourcing. In: Proceedings of the 23rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp 21:1–21:10
 80. Deng D, Shahabi C, Demiryurek U, Zhu L (2016) Task selection in spatial crowdsourcing from worker’s perspective. *GeoInformatica* 20(3):529–568
 81. Derigs U (1981) A shortest augmenting path method for solving minimal perfect matching problems. *Networks* 11(4):379–390
 82. Devanur NR, Jain K, Kleinberg RD (2013) Randomized primal-dual analysis of RANKING for online bipartite matching. In: Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms, pp 101–107
 83. Dickerson JP, Sankararaman KA, Srinivasan A, Xu P (2018) Allocation problems in ride-sharing platforms: Online matching with offline reusable resources. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, pp 1007–1014
 84. Dickerson JP, Sankararaman KA, Srinivasan A, Xu P (2018) Assigning tasks to workers based on historical data: Online task assignment with two-sided arrivals. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, pp 318–326
 85. Dittus M, Quattrone G, Capra L (2016) Analysing volunteer engagement in humanitarian mapping: building contributor communities at large scale. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, pp 108–118
 86. Dwork C (2006) Differential privacy. In: International Colloquium on Automata, Languages and Programming, pp 1–12
 87. Dwork C (2008) Differential privacy: A survey of results. In: Theory and Applications of Models of Computation, pp 1–19
 88. Emek Y, Kutten S, Wattenhofer R (2016) Online matching: haste makes waste! In: Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, pp 333–344
 89. Fan X, Liu J, Wang Z, Jiang Y, Liu X (2017) Crowdsourced road navigation: Concept, design, and implementation. *IEEE Communications Magazine* 55(6):126–128
 90. Fang Z, Huang L, Wierman A (2017) Prices and subsidies in the sharing economy. In: Proceedings of the 26th International Conference on World Wide Web, pp 53–62
 91. Feldman J, Mehta A, Mirrokni VS, Muthukrishnan S (2009) Online stochastic matching: Beating $1-1/e$. In: 50th Annual IEEE Symposium on Foundations of Computer Science, pp 117–126
 92. Ferguson TS, et al. (1989) Who solved the secretary problem? *Statistical science* 4(3):282–289
 93. Ford LR, Fulkerson DR (1956) Maximal flow through a network. *Canadian journal of Mathe-*

- ematics 8(3):399–404
94. Gale D, Shapley LS (1962) College admissions and the stability of marriage. *The American Mathematical Monthly* 69(1):9–15
 95. Gamal TE (1985) A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transactions on Information Theory* 31(4):469–472
 96. Gao D, Tong Y, She J, Song T, Chen L, Xu K (2016) Top-k team recommendation in spatial crowdsourcing. In: 17th International Conference on Web-Age Information Management, pp 191–204
 97. Gao D, Tong Y, Ji Y, Xu K (2017) Team-oriented task planning in spatial crowdsourcing. In: Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data, pp 41–56
 98. Gao D, Tong Y, She J, Song T, Chen L, Xu K (2017) Top-k team recommendation and its variants in spatial crowdsourcing. *Data Science and Engineering* 2(2):136–150
 99. Garcia-Molina H, Joglekar M, Marcus A, Parameswaran AG, Verroios V (2016) Challenges in data crowdsourcing. *IEEE Transactions on Knowledge and Data Engineering* 28(4):901–911
 100. Garcia-Ulloa DA, Xiong L, Sunderam VS (2017) Truth discovery for spatiotemporal events from crowdsourced data. *PVLDB* 10(11):1562–1573
 101. Garey MR, Johnson DS (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman
 102. Goel G, Mehta A (2008) Online budgeted matching in random input models with applications to adwords. In: Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms, pp 982–991
 103. Gu L, Wang K, Liu X, Guo S, Liu B (2017) A reliable task assignment strategy for spatial crowdsourcing in big data environment. In: IEEE International Conference on Communications, pp 1–6
 104. Guo B, Liu Y, Wang L, Li VOK, Lam JCK, Yu Z (2018) Task allocation in spatial crowdsourcing: Current state and future directions. *IEEE Internet of Things Journal* 5(3):1749–1764
 105. Guo S, Parameswaran AG, Garcia-Molina H (2012) So who won?: dynamic max discovery with the crowd. In: Proceedings of the 2012 ACM International Conference on Management of Data, pp 385–396
 106. Guttman A (1984) R-trees: A dynamic index structure for spatial searching. In: Proceedings of the 1984 ACM International Conference on Management of Data, pp 47–57
 107. Haklay MM, Weber P (2008) Openstreetmap: User-generated street maps. *IEEE Pervasive Computing* 7(4):12–18
 108. Han S, Xu Z, Zeng Y, Chen L (2019) Fluid: A blockchain based framework for crowdsourcing. In: Proceedings of the 2019 ACM International Conference on Management of Data, pp 1921–1924
 109. Hasenfratz D, Saukh O, Sturzenegger S, Thiele L (2012) Participatory air pollution monitoring using smartphones. *Mobile Sensing* 1:1–5
 110. Hashemi P, Abbaspour RA (2015) Assessment of logical consistency in openstreetmap based on the spatial similarity concept. In: *OpenStreetMap in GIScience, Lecture Notes in Geoinformation and Cartography*, Springer, pp 19–36
 111. ul Hassan U, Curry E (2014) A multi-armed bandit approach to online spatial task assignment. In: 2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops, pp 212–219
 112. ul Hassan U, Curry E (2016) Efficient task assignment for spatial crowdsourcing: A combinatorial fractional optimization approach with semi-bandit learning. *Expert Systems with Applications* 58:36–56
 113. He S, Shin D, Zhang J, Chen J (2014) Toward optimal allocation of location dependent tasks in crowdsensing. In: 2014 IEEE Conference on Computer Communications, pp 745–753
 114. Heipke C (2010) Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing* 65(6):550–557
 115. Ho C, Jabbari S, Vaughan JW (2013) Adaptive task assignment for crowdsourced classification. In: Proceedings of the 30th International Conference on Machine Learning, pp 534–542
 116. Hu H, Li G, Bao Z, Cui Y, Feng J (2016) Crowdsourcing-based real-time urban traffic speed estimation: From trends to speeds. In: 32nd IEEE International Conference on Data Engineering, pp 883–894
 117. Hu H, Zheng Y, Bao Z, Li G, Feng J, Cheng R (2016) Crowdsourced POI labelling: Location-aware result inference and task assignment. In: 32nd IEEE International Conference on Data Engineering, pp 61–72
 118. Huang P, Zhu W, Liao K, Sellis T, Yu Z, Guo L (2018) Efficient algorithms for flexible sweep coverage in crowdsensing. *IEEE Access* 6:50055–

- 50065
119. Huang Y, Bastani F, Jin R, Wang XS (2014) Large scale real-time ridesharing with service guarantee on road networks. *PVLDB* 7(14):2017–2028
 120. Huang Z, Kang N, Tang ZG, Wu X, Zhang Y, Zhu X (2018) How to match when all vertices arrive online. In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pp 17–29
 121. Jaillet P, Lu X (2014) Online stochastic matching: New algorithms with better bounds. *Mathematics of Operations Research* 39(3):624–646
 122. Jiang S, Chen L, Mislove A, Wilson C (2018) On ridesharing competition and accessibility: Evidence from uber, lyft, and taxi. In: *Proceedings of the 2018 International Conference on World Wide Web*, pp 863–872
 123. Jin H, Su L, Xiao H, Nahrstedt K (2018) Incentive mechanism for privacy-aware data aggregation in mobile crowd sensing systems. *IEEE/ACM Transactions on Networking* 26(5):2019–2032
 124. Jin X, Zhang Y (2016) Privacy-preserving crowd-sourced spectrum sensing. In: *Proceedings of the IEEE International Conference on Computer Communications*, pp 1–9
 125. Jin X, Zhang Y (2018) Privacy-preserving crowd-sourced spectrum sensing. *IEEE/ACM Transactions on Networking* 26(3):1236–1249
 126. Jonathan C, Mokbel MF (2018) Stella: geotagging images via crowdsourcing. In: *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp 169–178
 127. Kalyanasundaram B, Pruhs K (1993) Online weighted matching. *Journal of Algorithms* 14(3):478–488
 128. Kalyanasundaram B, Pruhs K (2000) An optimal deterministic algorithm for online b-matching. *Theoretical Computer Science* 233(1-2):319–325
 129. Karp RM, Vazirani UV, Vazirani VV (1990) An optimal algorithm for on-line bipartite matching. In: *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing*, pp 352–358
 130. Kazemi L, Shahabi C (2011) A privacy-aware framework for participatory sensing. *SIGKDD Explorations* 13(1):43–51
 131. Kazemi L, Shahabi C (2011) Towards preserving privacy in participatory sensing. In: *9th Annual IEEE International Conference on Pervasive Computing and Communications*, pp 328–331
 132. Kazemi L, Shahabi C (2012) Geocrowd: enabling query answering with spatial crowdsourcing. In: *Proceedings of the 20th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp 189–198
 133. Kazemi L, Shahabi C, Chen L (2013) Geotrucrowd: trustworthy query answering with spatial crowdsourcing. In: *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp 304–313
 134. Kesselheim T, Radke K, Tönnis A, Vöcking B (2013) An optimal online algorithm for weighted bipartite matching and extensions to combinatorial auctions. In: *21st Annual European Symposium on Algorithms*, pp 589–600
 135. Khuller S, Mitchell SG, Vazirani VV (1994) Online algorithms for weighted bipartite matching and stable marriages. *Theoretical Computer Science* 127(2):255–267
 136. Kim S, Mankoff J, Paulos E (2013) Sensr: evaluating a flexible framework for authoring mobile data-collection tools for citizen science. In: *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, pp 1453–1462
 137. Kim S, Mankoff J, Paulos E (2015) Exploring barriers to the adoption of mobile technologies for volunteer data collection campaigns. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp 3117–3126
 138. Kooti F, Grbovic M, Aiello LM, Djuric N, Radosavljevic V, Lerman K (2017) Analyzing uber’s ride-sharing economy. In: *Proceedings of the 26th International Conference on World Wide Web Companion*, pp 574–582
 139. Korula N, Pál M (2009) Algorithms for secretary problems on graphs and hypergraphs. In: *36th International Colloquium on Automata, Languages, and Programming*, pp 508–520
 140. Kuncheva LI, Whitaker CJ, Shipp CA, Duin RPW (2003) Limits on the majority vote accuracy in classifier fusion. *Pattern Analysis and Applications* 6(1):22–31
 141. Li G, Wang J, Zheng Y, Franklin MJ (2016) Crowdsourced data management: A survey. *IEEE Transactions on Knowledge and Data Engineering* 28(9):2296–2319
 142. Li G, Zheng Y, Fan J, Wang J, Cheng R (2017) Crowdsourced data management: Overview and challenges. In: *Proceedings of the 2017 ACM International Conference on Management of Data*, pp 1711–1716
 143. Li L, Chu W, Langford J, Schapire RE (2010) A contextual-bandit approach to personalized news article recommendation. In: *Proceedings of the 19th International Conference on World Wide*

- Web, pp 661–670
144. Li Y, Yiu ML, Xu W (2015) Oriented online route recommendation for spatial crowdsourcing task workers. In: *International Symposium on Spatial and Temporal Databases*, pp 137–156
 145. Li Y, Fang J, Zeng Y, Maag B, Tong Y, Zhang L (2019) Two-sided online bipartite matching in spatial data: experiments and analysis. *GeoInformatica* pp 1–24
 146. Lindell Y, Pinkas B (2009) A proof of security of Yao’s protocol for two-party computation. *Journal of Cryptology* 22(2):161–188
 147. Littlestone N, Warmuth MK (1994) The weighted majority algorithm. *Information and Computation* 108(2):212–261
 148. Liu A, Li Z, Liu G, Zheng K, Zhang M, Li Q, Zhang X (2017) Privacy-preserving task assignment in spatial crowdsourcing. *Journal of Computer Science and Technology* 32(5):905–918
 149. Liu A, Wang W, Shang S, Li Q, Zhang X (2018) Efficient task assignment in spatial crowdsourcing with worker and task privacy protection. *GeoInformatica* 22(2):335–362
 150. Liu B, Chen L, Zhu X, Zhang Y, Zhang C, Qiu W (2017) Protecting location privacy in spatial crowdsourcing using encrypted data. In: *Proceedings of the 20th International Conference on Extending Database Technology*, pp 478–481
 151. Liu J, Ji Y, Lv W, Xu K (2017) Budget-aware dynamic incentive mechanism in spatial crowdsourcing. *Journal of Computer Science and Technology* 32(5):890–904
 152. Liu SB, Iacucci AA, Meier P (2010) Ushahidi haiti and chile: next generation crisis mapping. *ACSM Bulletin* 246
 153. Liu X, He Q, Tian Y, Lee W, McPherson J, Han J (2012) Event-based social networks: linking the online and offline social worlds. In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp 1032–1040
 154. Liu Y, Guo B, Du H, Yu Z, Zhang D, Chen C (2017) Foodnet: Optimized on demand take-out food delivery using spatial crowdsourcing. In: *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pp 564–566
 155. Liu Z, Chen L, Tong Y (2018) Realtime traffic speed estimation with sparse crowdsourced data. In: *34th IEEE International Conference on Data Engineering*, pp 329–340
 156. Long C, Wong RC, Yu PS, Jiang M (2013) On optimal worst-case matching. In: *Proceedings of the 2013 ACM International Conference on Management of Data*, pp 845–856
 157. Lu A, Frazier PI, Kislev O (2018) Surge pricing moves uber’s driver-partners. In: *Proceedings of the 2018 ACM Conference on Economics and Computation*, p 3
 158. Ma S, Zheng Y, Wolfson O (2013) T-share: A large-scale dynamic taxi ridesharing service. In: *29th IEEE International Conference on Data Engineering*, pp 410–421
 159. Manshadi VH, Gharan SO, Saberi A (2011) Online stochastic matching: Online actions based on offline statistics. In: *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms*, pp 1285–1294
 160. McSherry F, Talwar K (2007) Mechanism design via differential privacy. In: *48th Annual IEEE Symposium on Foundations of Computer Science*, pp 94–103
 161. Meyerson A, Nanavati A, Poplawski LJ (2006) Randomized online algorithms for minimum metric bipartite matching. In: *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp 954–959
 162. Mineraud J, Lancerin F, Balasubramaniam S, Conti M, Tarkoma S (2015) You are airing too much: Assessing the privacy of users in crowdsourcing environmental data. In: *2015 IEEE TrustCom/BigDataSE/ISPA*, pp 523–530
 163. Mitsopoulou E, Boutsis I, Kalogeraki V, Yu JY (2018) A cost-aware incentive mechanism in mobile crowdsourcing systems. In: *2018 19th IEEE International Conference on Mobile Data Management*, pp 239–244
 164. Musthag M, Ganesan D (2013) Labor dynamics in a mobile micro-task market. In: *2013 ACM SIGCHI Conference on Human Factors in Computing Systems*, pp 641–650
 165. Nascimento MA, Silva JRO (1998) Towards historical r-trees. In: *Proceedings of the 1998 ACM symposium on Applied Computing*, pp 235–240
 166. Nemhauser GL, Wolsey LA, Fisher ML (1978) An analysis of approximations for maximizing submodular set functions. *Mathematical programming* 14(1):265–294
 167. Ouyang WR, Srivastava MB, Toniolo A, Norman TJ (2014) Truth discovery in crowdsourced detection of spatial events. In: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pp 461–470
 168. Ouyang WR, Srivastava MB, Toniolo A, Norman TJ (2016) Truth discovery in crowdsourced

- detection of spatial events. *IEEE Transactions on Knowledge and Data Engineering* 28(4):1047–1060
169. Paillier P (1999) Public-key cryptosystems based on composite degree residuosity classes. In: *International Conference on the Theory and Applications of Cryptographic Techniques*, pp 223–238
 170. Pournajaf L, Xiong L, Sunderam VS, Goryczka S (2014) Spatial task assignment for crowd sensing with cloaked locations. In: *IEEE 15th International Conference on Mobile Data Management*, pp 73–82
 171. Pournajaf L, Garcia-Ulloa DA, Xiong L, Sunderam VS (2015) Participant privacy in mobile crowd sensing task management: A survey of methods and challenges. *Proceedings of the 2015 ACM International Conference on Management of Data* 44(4):23–34
 172. Pournajaf L, Xiong L, Sunderam VS, Xu X (2015) STAC: spatial task assignment for crowd sensing with cloaked participant locations. In: *Proceedings of the 23rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp 90:1–90:4
 173. Quattrone G, Mashhadi A, Capra L (2014) Mind the map: the impact of culture and economic affluence on crowd-mapping behaviours. In: *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pp 934–944
 174. Raykar VC, Yu S, Zhao LH, Valadez GH, Florin C, Bogoni L, Moy L (2010) Learning from crowds. *Journal of Machine Learning Research* 11:1297–1322
 175. Ren X, Yu C, Yu W, Yang S, Yang X, McCann JA, Yu PS (2018) Lopub: High-dimensional crowdsourced data publication with local differential privacy. *IEEE Transactions on Information Forensics and Security* 13(9):2151–2166
 176. Robbins H (1985) Some aspects of the sequential design of experiments. In: *Herbert Robbins Selected Papers*, pp 169–177
 177. Saltenis S, Jensen CS, Leutenegger ST, López MA (2000) Indexing the positions of continuously moving objects. In: *Proceedings of the 2000 ACM International Conference on Management of Data*, pp 331–342
 178. Samet H (1984) The quadtree and related hierarchical data structures. *ACM Computing Surveys* 16(2):187–260
 179. Senaratne H, Mobasher A, Ali AL, Capineri C, Haklay M (2017) A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science* 31(1):139–167
 180. Senaratne H, Mobasher A, Ali AL, Capineri C, Haklay MM (2017) A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science* 31(1)
 181. Shahabi C (2013) Towards a generic framework for trustworthy spatial crowdsourcing. In: *Proceedings of the 12th International ACM Workshop on Data Engineering for Wireless and Mobile Access*, pp 1–4
 182. She J, Tong Y, Chen L (2015) Utility-aware social event-participant planning. In: *Proceedings of the 2015 ACM International Conference on Management of Data*, pp 1629–1643
 183. She J, Tong Y, Chen L, Cao CC (2015) Conflict-aware event-participant arrangement. In: *IEEE 31st International Conference on Data Engineering*, pp 735–746
 184. She J, Tong Y, Chen L, Cao CC (2016) Conflict-aware event-participant arrangement and its variant for online setting. *IEEE Transactions on Knowledge and Data Engineering* 28(9):2281–2295
 185. Shen W, Lopes CV, Crandall JW (2016) An online mechanism for ridesharing in autonomous mobility-on-demand systems. In: *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, pp 475–481
 186. Sheng X, Tang J, Zhang W (2012) Energy-efficient collaborative sensing with mobile phones. In: *2012 IEEE Conference on Computer Communications*, pp 1916–1924
 187. Song T, Tong Y, Wang L, She J, Yao B, Chen L, Xu K (2017) Trichromatic online matching in real-time spatial crowdsourcing. In: *33rd IEEE International Conference on Data Engineering*, pp 1009–1020
 188. Song T, Xu K, Li J, Li Y, Tong Y (2019) Multi-skill aware task assignment in real-time spatial crowdsourcing. *GeoInformatica* pp 1–21
 189. Stevens M, DHondt E (2010) Crowdsourcing of pollution data using smartphones. In: *Workshop on Ubiquitous Crowdsourcing*
 190. Su H, Zheng K, Huang J, Jeung H, Chen L, Zhou X (2014) Crowdplanner: A crowd-based route recommendation system. In: *30th IEEE International Conference on Data Engineering*, pp 1144–1155
 191. Sun D, Xu K, Cheng H, Zhang Y, Song T, Liu R, Xu Y (2018) Online delivery route recommendation in spatial crowdsourcing. *World Wide Web* (11):1–22

192. Sweeney L (2002) k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10(5):557–570
193. Tao Q, Zeng Y, Zhou Z, Tong Y, Chen L, Xu K (2018) Multi-worker-aware task planning in real-time spatial crowdsourcing. In: *International Conference on Database Systems for Advanced Applications*, pp 301–317
194. Theodoridis Y, Vazirgiannis M, Sellis TK (1996) Spatio-temporal indexing for large multimedia applications. In: *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, pp 441–448
195. Ting H, Xiang X (2015) Near optimal algorithms for online maximum edge-weighted b-matching and two-sided vertex-weighted b-matching. *Theoretical Computer Science* 607:247–256
196. To H, Shahabi C (2018) Location privacy in spatial crowdsourcing, *Handbook of mobile data privacy* pp 167–194
197. To H, Ghinita G, Shahabi C (2014) A framework for protecting worker location privacy in spatial crowdsourcing. *PVLDB* 7(10):919–930
198. To H, Shahabi C, Kazemi L (2015) A server-assigned spatial crowdsourcing framework. *ACM Transactions on Spatial Algorithms and Systems* 1(1):2
199. To H, Asghari M, Deng D, Shahabi C (2016) SCAWG: A toolbox for generating synthetic workload for spatial crowdsourcing. In: *2016 IEEE International Conference on Pervasive Computing and Communication Workshops*, pp 1–6
200. To H, Fan L, Tran L, Shahabi C (2016) Real-time task assignment in hyperlocal spatial crowdsourcing under budget constraints. In: *IEEE International Conference on Pervasive Computing and Communications*, pp 1–8
201. To H, Ghinita G, Fan L, Shahabi C (2017) Differentially private location protection for worker datasets in spatial crowdsourcing. *IEEE Transactions on Mobile Computing* 16(4):934–949
202. To H, Shahabi C, Xiong L (2018) Privacy-preserving online task assignment in spatial crowdsourcing with untrusted server. In: *34th IEEE International Conference on Data Engineering*, pp 833–844
203. Tong X, Gupta A, Lo H, Choo A, Gromala D, Shaw CD (2017) Chasing lovely monsters in the wild, exploring players’ motivation and play patterns of pokémon go: Go, gone or go away? In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, Companion Volume*, pp 327–330
204. Tong Y, Zhou Z (2018) Dynamic task assignment in spatial crowdsourcing. *Proceedings of the 26rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* 10(2):18–25
205. Tong Y, She J, Ding B, Chen L, Wo T, Xu K (2016) Online minimum matching in real-time spatial data: Experiments and analysis. *PVLDB* 9(12):1053–1064
206. Tong Y, She J, Ding B, Wang L, Chen L (2016) Online mobile micro-task allocation in spatial crowdsourcing. In: *32nd IEEE International Conference on Data Engineering*, pp 49–60
207. Tong Y, Chen L, Shahabi C (2017) Spatial crowdsourcing: Challenges, techniques, and applications. *PVLDB* 10(12):1988–1991
208. Tong Y, Chen Y, Zhou Z, Chen L, Wang J, Yang Q, Ye J, Lv W (2017) The simpler the better: A unified approach to predicting original taxi demands based on large-scale online platforms. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp 1653–1662
209. Tong Y, Wang L, Zhou Z, Ding B, Chen L, Ye J, Xu K (2017) Flexible online task assignment in real-time spatial data. *PVLDB* 10(11):1334–1345
210. Tong Y, Chen L, Zhou Z, Jagadish HV, Shou L, Lv W (2018) SLADE: A smart large-scale task decomposer in crowdsourcing. *IEEE Transactions on Knowledge and Data Engineering* 30(8):1588–1601
211. Tong Y, Wang L, Zhou Z, Chen L, Du B, Ye J (2018) Dynamic pricing in spatial crowdsourcing: A matching-based approach. In: *Proceedings of the 2018 ACM International Conference on Management of Data*, pp 773–788
212. Tong Y, Zeng Y, Zhou Z, Chen L, Ye J, Xu K (2018) A unified approach to route planning for shared mobility. *PVLDB* 11(11):1633–1646
213. Tran L, To H, Fan L, Shahabi C (2018) A real-time framework for task assignment in hyperlocal spatial crowdsourcing. *ACM Transactions on Intelligent Systems and Technology* 9(3):37:1–37:26
214. U LH, Yiu ML, Mouratidis K, Mamoulis N (2008) Capacity constrained assignment in spatial databases. In: *Proceedings of the 2008 ACM International Conference on Management of Data*, pp 15–28
215. Vansteenwegen P, Souffriau W, Oudheusden DV (2011) The orienteering problem: A survey. *European Journal of Operational Research* 209(1):1–10

216. Venanzi M, Guiver J, Kazai G, Kohli P, Shokouhi M (2014) Community-based bayesian aggregation models for crowdsourcing. In: Proceedings of the 23rd International Conference on World Wide Web, pp 155–164
217. Vu K, Zheng R, Gao J (2012) Efficient algorithms for k-anonymous location privacy in participatory sensing. In: Proceedings of the IEEE International Conference on Computer Communications, pp 2399–2407
218. Wang L, Zhang D, Yang D, Lim BY, Ma X (2016) Differential location privacy for sparse mobile crowdsensing. In: IEEE 16th International Conference on Data Mining, pp 1257–1262
219. Wang Q, He W, Wang X, Cui L (2016) Quality-assure and budget-aware task assignment for spatial crowdsourcing. In: International Conference on Collaborative Computing: Networking, Applications and Worksharing, pp 60–70
220. Wang Q, He W, Wang X, Cui L (2016) Quality-assure and budget-aware task assignment for spatial crowdsourcing. In: 12th International Conference on Collaborate Computing: Networking, Applications and Worksharing, pp 60–70
221. Wang Y, Wong SC (2015) Two-sided online bipartite matching and vertex cover: Beating the greedy algorithm. In: 42nd International Colloquium on Automata, Languages, and Programming, pp 1070–1081
222. Wang Y, Tong Y, Long C, Xu P, Xu K, Lv W (2019) Adaptive dynamic bipartite graph matching: A reinforcement learning approach. In: 35th IEEE International Conference on Data Engineering, pp 1478–1489
223. Whitehill J, Ruvolo P, Wu T, Bergsma J, Movellan JR (2009) Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In: Advances in Neural Information Processing Systems, pp 2035–2043
224. Williamson DP, Shmoys DB (2011) The Design of Approximation Algorithms. Cambridge University Press
225. Wong RC, Tao Y, Fu AW, Xiao X (2007) On efficient spatial matching. In: Proceedings of the 33rd International Conference on Very Large Data Bases, pp 579–590
226. Wu P, Ngai EW, Wu Y (2018) Toward a real-time and budget-aware task package allocation in spatial crowdsourcing. Decision Support Systems 110:107–117
227. Xia H, Yang H (2018) Is last-mile delivery a ‘killer app’ for self-driving vehicles? Communications of the ACM 61(11):70–75
228. Xu Y, Tong Y, Shi Y, Tao Q, Xu K, Li W (2019) An efficient insertion operator in dynamic ridesharing services. In: 35th IEEE International Conference on Data Engineering, pp 1022–1033
229. Yang C, Lin K (2002) An index structure for improving closest pairs and related join queries in spatial databases. In: International Database Engineering & Applications Symposium, pp 140–149
230. Yao AC (1986) How to generate and exchange secrets (extended abstract). In: 27th Annual Symposium on Foundations of Computer Science, pp 162–167
231. Yu H, Miao C, Shen Z, Leung C (2015) Quality and budget aware task allocation for spatial crowdsourcing. In: Proceedings of the 14th International Conference on Autonomous Agents and MultiAgent Systems, pp 1689–1690
232. Yuan J, Zheng Y, Xie X, Sun G (2013) T-drive: Enhancing driving directions with taxi drivers’ intelligence. IEEE Transactions on Knowledge and Data Engineering 25(1):220–232
233. Zeng Y, Tong Y, Chen L, Zhou Z (2018) Latency-oriented task completion via spatial crowdsourcing. In: 34th IEEE International Conference on Data Engineering, pp 317–328
234. Zhai D, Sun Y, Liu A, Li Z, Liu G, Zhao L, Zheng K (2018) Towards secure and truthful task assignment in spatial crowdsourcing. World Wide Web pp 1–24
235. Zhang CJ, Tong Y, Chen L (2014) Where to: Crowd-aided path selection. PVLDB 7(14):2005–2016
236. Zhang G, Zhu A, Huang Z, Ren G, Qin C, Xiao W (2018) Validity of historical volunteered geographic information: Evaluating citizen data for mapping historical geographic phenomena. Transactions in GIS 22(1):149–164
237. Zhang J, Wen D, Zeng S (2016) A discounted trade reduction mechanism for dynamic ridesharing pricing. IEEE Transactions on Intelligent Transportation Systems 17(6):1586–1595
238. Zhang L, Hu T, Min Y, Wu G, Zhang J, Feng P, Gong P, Ye J (2017) A taxi order dispatch model based on combinatorial optimization. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 2151–2159
239. Zhang X, Yang Z, Zhou Z, Cai H, Chen L, Li X (2014) Free market of crowdsourcing: Incentive mechanism design for mobile sensing. IEEE Transactions on Parallel and Distributed Systems 25(12):3190–3200

240. Zhang X, Yang Z, Sun W, Liu Y, Tang S, Xing K, Mao X (2016) Incentives for mobile crowd sensing: A survey. *IEEE Communications Surveys and Tutorials* 18(1):54–67
241. Zhang X, Yang Z, Liu Y, Tang S (2019) On reliable task assignment for spatial crowdsourcing. *IEEE Transactions on Emerging Topics in Computing* 7(1):174–186
242. Zhang Y, Chen Q, Zhong S (2016) Privacy-preserving data aggregation in mobile phone sensing. *IEEE Transactions on Information Forensics and Security* 11(5):980–992
243. Zhao B, Xu P, Shi Y, Tong Y, Zhou Z, Zeng Y (2019) Preference-aware task assignment in on-demand taxi dispatching: An online stable matching approach. In: *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, pp 2245–2252
244. Zhao D, Li X, Ma H (2014) How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint. In: *Proceedings of the IEEE Conference on Computer Communications*, pp 1213–1221
245. Zhao Y, Han Q (2016) Spatial crowdsourcing: current state and future directions. *IEEE Communications Magazine* 54(7):102–107
246. Zhao Y, Li Y, Wang Y, Su H, Zheng K (2017) Destination-aware task assignment in spatial crowdsourcing. In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp 297–306
247. Zhao Z, Wei F, Zhou M, Chen W, Ng W (2015) Crowd-selection query processing in crowdsourcing databases: A task-driven approach. In: *Proceedings of the 18th International Conference on Extending Database Technology*, pp 397–408
248. Zheng L, Chen L, Ye J (2018) Order dispatch in price-aware ridesharing. *PVLDB* 11(8):853–865
249. Zheng Y, Li G, Li Y, Shan C, Cheng R (2017) Truth inference in crowdsourcing: Is the problem solved? *PVLDB* 10(5):541–552