



UNIVERSITÀ DEGLI STUDI DI PALERMO  
FACOLTÀ DI INGEGNERIA

---

*LAUREA MAGISTRALE IN INGEGNERIA INFORMATICA*

**PROGETTAZIONE E IMPLEMENTAZIONE DI  
UN SISTEMA MOBILE CROWDSENSING PER  
UTENTI BASATI SU RAZIONALITÀ LIMITATA**

*Tesi di Laurea*

Dott. Salvatore Vigano'

*Relatore:*

Ch.mo Prof. Pierluca Ferraro

*Controrelatore:*

*Correlatore:*

---

Anno Accademico 2025/2026

UNIVERSITÀ DEGLI STUDI DI PALERMO  
FACOLTÀ DI INGEGNERIA

---

*LAUREA MAGISTRALE IN INGEGNERIA INFORMATICA*

PROGETTAZIONE E IMPLEMENTAZIONE DI UN SISTEMA  
MOBILE CROWDSENSING PER UTENTI BASATI SU  
RAZIONALITÀ LIMITATA

*Tesi di Laurea di*  
Dott. Salvatore Vigano'

*Relatore:*  
Ch.mo Prof. Pierluca Ferraro

*Controrelatore:*

*Correlatore:*

---

## Sommario

Il presente lavoro di tesi si concentra sulla progettazione e validazione di un sistema di **Mobile Crowdensing (MCS)** basato sull'algoritmo **IMCU (Incentive Mechanism for Crowdensing Users)**, testandone la tenuta su dati reali di mobilità urbana relativi a una flotta di taxi operanti a Roma. Una prima analisi, condotta ipotizzando utenti perfettamente razionali in un contesto di **asta inversa veritiera**, conferma la solidità del modello teorico: il sistema risulta stabile e le strategie degli utenti convergono all'equilibrio di Nash. Tuttavia, l'introduzione di scenari più realistici caratterizzati da **razionalità limitata** — inclusi errori di valutazione e strategie subottimali — fa emergere criticità significative. In queste condizioni, l'efficienza del meccanismo IMCU subisce una flessione evidente, rivelando una vulnerabilità strutturale del modello classico. Per superare tali limiti, il lavoro propone e sperimenta **GAP**, un meccanismo adattivo che sfrutta l'apprendimento automatico per monitorare il comportamento degli utenti. Modulando dinamicamente incentivi e regole di selezione, GAP permette di "recuperare" l'efficienza perduta, ristabilendo la stabilità del sistema anche in contesti comportamentali complessi. In definitiva, lo studio dimostra che la sostenibilità delle piattaforme di crowdsensing non dipende solo dalla correttezza teorica, ma dalla capacità degli algoritmi di adattarsi elasticamente alla natura imperfetta del comportamento umano.

# Indice

<b>1</b>	<b>Introduzione</b>	<b>3</b>
1.1	Contesto del Mobile Crowdsensing . . . . .	3
1.2	Incentivi e Problema dei Costi . . . . .	5
1.3	Obiettivi della Ricerca . . . . .	7
1.3.1	Fase 1: Validazione della Baseline . . . . .	8
1.3.2	Fase 2: Analisi della Razionalità Limitata . . . . .	9
1.3.3	Meccanismo Adattivo GAP . . . . .	10
1.3.4	Sintesi del Piano Sperimentale . . . . .	10
1.4	Contributi del Lavoro . . . . .	11
1.5	Organizzazione della Tesi . . . . .	12
<b>2</b>	<b>Stato dell'Arte e Fondamenti Teorici</b>	<b>13</b>
2.1	Evoluzione dei Sistemi di Sensing . . . . .	13
2.1.1	Definizione di Mobile Crowdsensing . . . . .	14
2.1.2	Architettura e tassonomia dei sistemi MCS . . . . .	15
2.1.3	Acquisizione dei dati: Participatory vs Opportunistic Sensing . . . . .	19
2.1.4	Scenari applicativi . . . . .	20
2.1.5	Limiti principali per la sostenibilità delle piattaforme . . . . .	21
2.2	Teoria dei Giochi nei sistemi MCS . . . . .	22
2.2.1	Giochi cooperativi e Non-Cooperativi . . . . .	23
2.2.2	Giochi Bayesiani . . . . .	24
2.2.3	Equilibrio di Nash e Strategie Dominanti . . . . .	24
2.2.4	Giochi di Stackelberg . . . . .	25
2.2.5	Il meccanismo VCG . . . . .	26

2.3	Sfide in contesti reali per un sistema MCS . . . . .	27
2.3.1	Qualità dell'Informazione (QoI) . . . . .	27
2.3.2	Trustworthiness e Privacy . . . . .	28
2.3.3	Vincoli Energetici . . . . .	28
2.3.4	Copertura Spaziale e Mobilità . . . . .	29
2.3.5	Equità Distributiva: Indice di Gini . . . . .	29
2.3.6	Analisi dei Pattern Temporali e Segmentazione dei Regimi . . . . .	31
2.4	Razionalità Limitata (Bounded Rationality) . . . . .	32
2.4.1	Bias Cognitivi e Limitazione: il modello di Herbert Simon . . . . .	32
2.4.2	Modelli Fast-and-Frugal Trees (FFTs) . . . . .	34
2.4.3	Impatto sui Meccanismi d'Asta . . . . .	35
2.5	Submodularità e Ottimizzazione . . . . .	36
2.5.1	Proprietà di Submodularità . . . . .	36
2.5.2	Algoritmi Greedy . . . . .	38
2.5.3	Selezione dei Vincitori (Winner Determination) . . . . .	38
<b>3</b>	<b>Modello Matematico del Sistema e degli Agenti</b>	<b>40</b>
3.1	Introduzione al Modello Computazionale . . . . .	40
3.2	L'Entità Task: Formalizzazione della Domanda . . . . .	42
3.3	L'Entità Utente: L'Agente Razionale . . . . .	43
3.4	Modello di Costo e Strategie di Bidding . . . . .	45
3.4.1	Funzione di Costo Privato ( $c_i$ ) . . . . .	45
3.4.2	Strategia di Bidding in Regime di Razionalità Perfetta . . . . .	49
3.4.3	Funzione di Utilità dell'Agente . . . . .	52
<b>Elenco delle figure</b>		<b>55</b>
<b>Bibliografia</b>		<b>57</b>

# Capitolo 1

## Introduzione

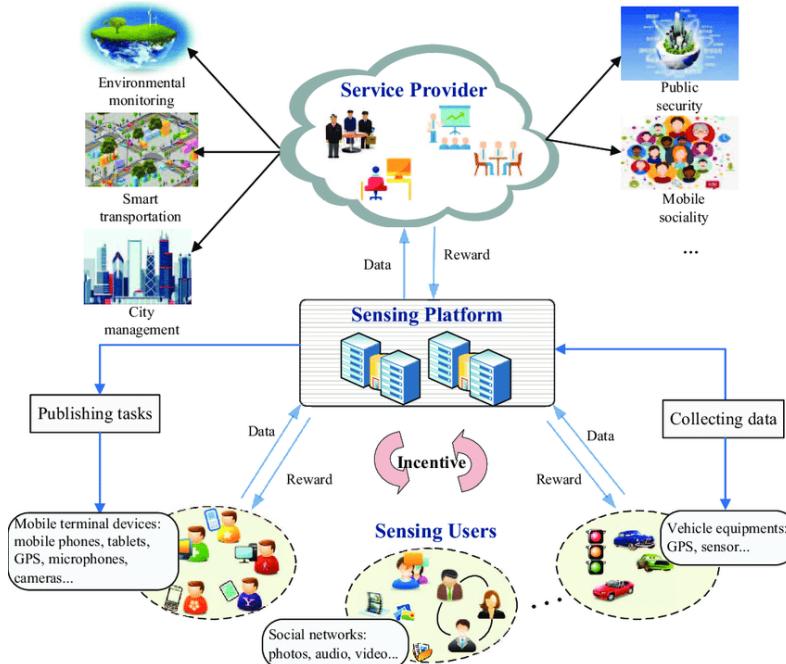
### 1.1 Contesto del Mobile Crowdsensing

Negli ultimi anni si è verificato un cambiamento radicale nel modo in cui raccogliamo dati sul mondo che ci circonda. Gli smartphone e i dispositivi mobili moderni integrano numerosi sensori — ricevitori GPS, accelerometri, giroscopi, magnetometri, microfoni, fotocamere ad alta risoluzione — supportati da capacità di calcolo e connettività in continua crescita. Questa infrastruttura sensoriale distribuita ha reso possibile un nuovo approccio alla raccolta di dati territoriali [10].

Su questa base tecnologica si è sviluppato il paradigma del **Mobile Crowdsensing (MCS)**, che sfrutta la mobilità degli utenti per raccogliere informazioni su larga scala senza dover installare infrastrutture dedicate. A differenza delle tradizionali reti di sensori wireless statici (WSN), che richiedono investimenti significativi per il dispiegamento e la manutenzione dell'hardware, l'MCS delega la raccolta dati alla "folla" (*crowd*), in cui gli utenti contribuiscono attivamente alla generazione delle informazioni [6].

La letteratura distingue due modalità operative principali [3]: il **Rilevamento Partecipativo** (*Participatory Sensing*), in cui l'utente interviene attivamente nel processo (ad esempio, fotografando una buca stradale o segnalando manualmente il prezzo del carburante), e il **Rilevamento Opportunistico** (*Opportunistic Sensing*), dove la raccolta avviene automaticamente in background durante le normali attività quotidiane, come il campionamento delle reti Wi-Fi o il monitoraggio delle vibrazioni stradali.

L'architettura generale di un sistema MCS, schematizzata in Figura 1.1, ruota attorno a una piattaforma centrale (tipicamente in cloud). La piattaforma pubblica task geolocalizzati, gli utenti mobili raccolgono dati tramite i sensori dei propri dispositivi e li trasmettono al server, che li aggrega e fornisce servizi ai consumatori finali (amministrazioni pubbliche, enti di ricerca, aziende).



**Figura 1.1:** Architettura concettuale di un sistema di Mobile Crowdensing. La piattaforma centrale coordina tre attori principali: (i) pubblica task geolocalizzati agli utenti mobili, (ii) raccoglie dati dai sensori dei dispositivi, (iii) aggrega le informazioni e fornisce servizi ai consumatori finali (amministrazioni pubbliche, enti di ricerca, aziende) [6].

Le applicazioni concrete sono numerose: dal monitoraggio del traffico in tempo reale alla mappatura dell'inquinamento acustico, dal controllo della qualità dell'aria alla verifica della copertura delle reti cellulari [17, 11].

In questo contesto, Roma rappresenta un caso di studio particolarmente interessante. La sua complessa topografia urbana — un centro storico medievale che si intreccia con quartieri moderni — e i flussi di mobilità eterogenei la rendono un banco di prova significativo per analizzare le dinamiche dei sistemi MCS. Il dataset utilizzato in questo lavoro, CRAWDAD *roma/taxi* [2], traccia continuativamente gli spostamenti di 316 veicoli per 28 giorni nel febbraio 2014, generando circa 11 milioni di punti GPS su un'area di oltre 1200 km<sup>2</sup>.

## 1.2 Incentivi e Problema dei Costi

Nonostante il potenziale tecnologico, la sostenibilità a lungo termine di un sistema MCS dipende dalla partecipazione attiva e continuativa degli utenti; pertanto, entra in gioco un problema fondamentale: **la fornitura di dati di qualità comporta costi non trascurabili per i partecipanti** [15].

Quando un utente partecipa a una campagna di sensing, sostiene diversi tipi di costi. In primo luogo, vi sono i *costi digitali*: consumo di batteria (talvolta significativo, se sensori come GPS e accelerometri rimangono attivi per lunghi periodi), utilizzo del piano dati mobile, potenza di calcolo. Nel caso di task che richiedono spostamenti fisici verso aree specifiche — quelli che la letteratura definisce rilevamento *location-dependent* — i costi includono carburante, usura del veicolo e tempo impiegato. Un esempio tipico in questo contesto riguarda il tassista che deve deviare dalla rotta ottimale per esigenze di rilevamento: ciò implica il sostenimento di costi aggiuntivi e il rischio di rinunciare a corse potenzialmente redditizie.

Anche la *privacy* costituisce una barriera significativa: la condivisione di tracce GPS, registrazioni audio o fotografie geolocalizzate solleva legittime preoccupazioni. Infatti, sorgono diversi quesiti: dove vengono memorizzati questi dati? Chi può accedervi? Possono essere utilizzati per ricostruire abitudini quotidiane e pattern comportamentali sensibili? [4].

La *gamification* (badge, classifiche, punteggi di reputazione) può funzionare inizialmente per attrarre utenti, ma diversi studi hanno evidenziato come tali incentivi intrinseci perdano efficacia quando i costi tangibili superano la gratificazione psicologica. Quando il rapporto costo-beneficio percepito diventa negativo — ad esempio, sostenere costi di carburante significativi per ottenere solo riconoscimenti virtuali — il tasso di abbandono aumenta drasticamente. La progettazione di **meccanismi di incentivazione monetaria** robusti ed efficienti risulta fondamentale per garantire la sostenibilità economica di una piattaforma MCS professionale [8].

Progettare questi meccanismi in un contesto reale solleva però questioni complesse. Si identificano tre criticità fondamentali [19, 18]:

1. **Asimmetria informativa:** la piattaforma conosce il valore dei dati che intende raccogliere, ma non conosce i costi privati che ogni singolo utente deve sostenere per fornirli.
2. **Comportamento strategico:** gli utenti, agendo come agenti economici razionali (o tentando di farlo), possono essere incentivati a dichiarare costi gonfiati per massimizzare il proprio profitto a spese dell'efficienza globale del sistema.
3. **Vincoli di budget:** la piattaforma opera con risorse finanziarie limitate e deve massimizzare l'utilità dei dati raccolti rispettando vincoli di bilancio.

Per affrontare queste criticità, *Yang et al.* hanno proposto il meccanismo **IMCU (Incentive Mechanism for Crowdsensing Users)**, basato sulla teoria delle aste inverse (*Reverse Auctions*), che garantisce proprietà teoriche fondamentali come la veridicità (*Truthfulness*) e la razionalità individuale (*Individual Rationality*) [18]; tuttavia, la letteratura esistente si concentra prevalentemente su analisi teoriche o simulazioni in ambienti semplificati, trascurando le complessità comportamentali degli utenti reali.

Questo lavoro affronta quindi una domanda centrale: **in che misura le performance del meccanismo IMCU si mantengono quando il parametro di configurazione (come il raggio di copertura dei task) viene variato, e quali sono i trade-off che emergono tra efficienza economica e inclusività della partecipazione?**

### 1.3 Obiettivi della Ricerca

Questo lavoro integra competenze di ingegneria del software, teoria dei giochi ed economia comportamentale. L’obiettivo principale è valutare quanto i meccanismi teorici di incentivazione mantengano le loro proprietà in scenari operativi reali, caratterizzati da parametri di configurazione che influenzano in modo significativo le dinamiche competitive.

È necessario definire con precisione i confini della ricerca. Sebbene la privacy e la sicurezza dei dati siano temi critici per l’MCS, questo lavoro **non** sviluppa tecniche crittografiche avanzate: la privacy viene modellata implicitamente nel costo privato  $c_i$  sostenuto dall’utente, assumendo che gli utenti richiedano un compenso maggiore per task che espongono dati più sensibili. Questa astrazione economica permette di concentrarsi sulla dinamica di incentivazione senza affrontare le tecniche di protezione dei dati, che costituirebbero un filone di ricerca separato.

L’attenzione è quindi focalizzata esclusivamente sulla dinamica economica e algoritmica dell’allocazione dei task, sulla modellazione dei costi operativi e sul comportamento strategico (o sub-strategico) degli utenti.

Questo lavoro si articola in **tre fasi sperimentali**, che corrispondono alle tre parti della struttura della tesi:

- **Fase 1 (Parte I):** Validazione della baseline teorica;
- **Fase 2 (Parte II):** Analisi della robustezza alla razionalità limitata;
- **Fase 3 (Parte III):** Progettazione di un meccanismo adattivo (GAP).

### 1.3.1 Fase 1: Validazione della Baseline

Nella prima fase è stato progettato e implementato un simulatore MCS completo, calibrato su un dataset reale di mobilità taxi nella città di Roma (febbraio 2014, oltre 300 taxi tracciati). L’obiettivo è riprodurre fedelmente il meccanismo IMCU assumendo una popolazione di agenti a **razionalità perfetta**.

È opportuno chiarire operativamente il concetto di razionalità perfetta nel contesto della simulazione. Nel nostro caso, significa che ogni utente  $i$  sottomette un bid  $b_i$  esattamente uguale al proprio costo reale  $c_i$ . Questo rappresenta il comportamento di equilibrio previsto dalla teoria dei giochi quando tutti gli altri utenti agiscono razionalmente e il meccanismo è veritiero (*truthful*): in queste condizioni, dire la verità sul proprio costo diventa la strategia dominante.

È importante sottolineare che in questa fase **non si intende dimostrare** che gli utenti scelgano spontaneamente di essere veritieri — questa proprietà è già garantita teoricamente dal disegno del meccanismo IMCU [18]. L’obiettivo è piuttosto quello di **validare empiricamente** che, dato il comportamento di equilibrio ( $b_i = c_i$ ), il meccanismo mantiene le proprietà di:

- *Individual Rationality* (nessun utente opera in perdita:  $p_i \geq b_i$ );
- *Profitability* (la piattaforma non va in perdita:  $u_0 \geq 0$ ).

Un aspetto metodologico rilevante di questa fase è l’analisi dell’impatto del parametro *raggio di copertura* dei task sull’efficienza del sistema. Attraverso tre configurazioni sperimentali (raggio 1.5 km, 2.5 km, 4.0 km), vengono quantificati empiricamente i trade-off che emergono: l’aumento del raggio incrementa l’efficienza economica ma riduce drasticamente il numero di partecipanti selezionati, con implicazioni significative per la sostenibilità sociale della piattaforma.

I risultati di questa fase costituiscono la baseline quantitativa di riferimento per le fasi successive.

### 1.3.2 Fase 2: Analisi della Razionalità Limitata

Nella seconda fase si introduce un elemento di realismo comportamentale ispirato alla teoria della *Bounded Rationality* di Herbert Simon [16]. La letteratura sulla mobilità urbana documenta ampiamente che i tassisti, pur essendo professionisti esperti, adottano euristiche semplificate piuttosto che ottimizzazioni matematiche perfette nella scelta delle corse.

Questa osservazione solleva una questione metodologica rilevante: se i tassisti manifestano comportamenti subottimali nelle operazioni quotidiane, è ragionevole assumere che adottino strategie perfettamente razionali in un contesto MCS?

Gli agenti della Fase 2 presentano quindi profili comportamentali eterogenei, caratterizzati da:

- **Errori nella stima dei costi** (sottostima o sovrastima sistematica delle distanze o dei tempi di percorrenza).
- **Euristiche decisionali semplificate** (strategie "sufficientemente buone" piuttosto che ottimali).
- **Comportamenti opportunistici non ottimali** (tentativi di manipolazione maldestri del sistema, che non necessariamente massimizzano il profitto individuale).

L'obiettivo è quantificare, attraverso metriche rigorose, il deterioramento dell'efficienza del meccanismo IMCU quando le assunzioni di perfetta razionalità vengono meno. L'ipotesi di lavoro è che tale deterioramento non sia trascurabile, a dimostrazione della potenziale fragilità dei modelli teorici classici in contesti reali.

Tra le metriche considerate rientrano il tasso di rottura dei contratti, l'efficienza realizzata rispetto alla baseline teorica e misure di robustezza del meccanismo rispetto a diversi profili comportamentali.

### 1.3.3 Fase 3: Meccanismo Adattivo GAP

Per affrontare le criticità emerse nella Fase 2, la terza fase propone lo sviluppo di un nuovo meccanismo algoritmico denominato **GAP (Game-theoretic Adaptive Policy)**. Il meccanismo GAP si basa sull'apprendimento dinamico dei pattern comportamentali degli utenti per adattare le strategie di selezione e incentivazione: invece di assumere che tutti gli utenti siano razionali (Fase 1) o limitarsi a constatare che non lo sono (Fase 2), GAP stima in tempo reale i profili comportamentali individuali — reputazione, affidabilità, pattern di bidding — e ricalibra le regole dell'asta di conseguenza. L'obiettivo è ripristinare l'efficienza e la stabilità del sistema anche in presenza di utenti con razionalità limitata, riducendo il gap tra le prestazioni teoriche e quelle osservate in scenari realistici.

### 1.3.4 Sintesi del Piano Sperimentale

La Tabella 1.1 sintetizza le tre fasi sperimentali della tesi, evidenziando gli obiettivi, le assunzioni comportamentali sugli utenti e le metriche chiave di valutazione.

Fase	Cap.	Assunzione Utenti	Obiettivo	Metriche Chiave
<b>Fase 1</b>	3–6	Razionalità perfetta ( $b_i = c_i$ )	Validazione empirica proprietà IMCU su dati reali	Efficienza $u_0/v(S)$ , Profitability, Indice di Gini
<b>Fase 2</b>	7	Razionalità limitata (errori stima, euristiche)	Valutazione del calo di efficienza con comportamenti realistici	Tasso rottura contratti, Efficienza realizzata, Robustezza
<b>Fase 3</b>	8	Eterogenei (razionali + limitati)	Ripristino efficienza tramite apprendimento e adattamento	Recovery rate, Convergenza, Overhead computazionale

**Tabella 1.1:** Sintesi delle tre fasi sperimentali della tesi, con indicazione dei capitoli corrispondenti, delle assunzioni sugli utenti, degli obiettivi e delle principali metriche di valutazione.

## 1.4 Contributi del Lavoro

I principali contributi di questo lavoro possono essere riassunti come segue.

- **Framework di simulazione data-driven:** a differenza della maggior parte della letteratura MCS, che si basa su simulazioni con dati sintetici o scenari semplificati, questo lavoro sviluppa un ambiente di simulazione che integra dati di mobilità reale (tracce GPS di taxi romani), topologia urbana effettiva e modelli economici calibrati su dati storici (costi operativi taxi Roma 2014). Il dataset copre 316 veicoli tracciati per 28 giorni consecutivi (febbraio 2014), generando circa 11 milioni di punti GPS su un'area di oltre 1200 km<sup>2</sup>. La validazione in un contesto urbano complesso come Roma fornisce evidenze quantitative dell'applicabilità operativa del meccanismo IMCU in scenari reali.
- **Quantificazione dell'impatto del raggio di copertura:** questo lavoro fornisce una prima quantificazione empirica dell'impatto del parametro *raggio di copertura* sulle performance di un'asta inversa MCS. I risultati della Fase 1 (Capitolo 6) evidenziano trade-off rilevanti: l'aumento del raggio da 1.5 km a 4.0 km porta a un incremento dell'efficienza economica, al prezzo però di una drastica contrazione della partecipazione, con implicazioni significative per la sostenibilità sociale della piattaforma.
- **Approccio adattivo GAP (proposto):** l'approccio adattivo proposto non richiede assunzioni rigide sul comportamento degli utenti, ma stima dinamicamente i loro profili per ottimizzare le decisioni allocative. Si progetta un sistema che apprende dai comportamenti osservati e si adatta di conseguenza, con l'obiettivo di ripristinare le prestazioni della baseline teorica anche in presenza di razionalità limitata.

## 1.5 Organizzazione della Tesi

Di seguito si riportano brevemente i contenuti dei singoli capitoli.

**Capitolo 2** Analizza lo stato dell'arte e i fondamenti teorici, spaziando dai concetti essenziali di Teoria dei Giochi (equilibri di Nash, strategie dominanti) alle sfide operative del Mobile Crowdsensing, inclusi i modelli di razionalità limitata.

**Capitolo 3** Definisce il modello matematico del sistema (task, utenti e funzioni di costo), formalizzando le assunzioni di razionalità perfetta per la **Fase 1** e motivando l'adozione dello *star routing* per la stima dei costi operativi.

**Capitolo 4** Descrive il meccanismo *User-Centric* IMCU, dettagliando gli algoritmi greedy di allocazione e di calcolo dei pagamenti, e ne dimostra le proprietà teoriche fondamentali, inclusa la veridicità.

**Capitolo 5** Delinea la metodologia sperimentale della prima fase, descrivendo il dataset CRA-WDAD romा/taxi, le procedure di ETL e la discretizzazione spaziale utilizzata per generare gli scenari di simulazione.

**Capitolo 6** Discute i risultati della **Fase 1**, validando le prestazioni del meccanismo in condizioni ideali e quantificando l'impatto del raggio di copertura sull'efficienza economica e sulla partecipazione.

**Capitolo 7** Affronta la **Fase 2**, introducendo modelli di razionalità limitata per valutare il deterioramento dell'efficienza del sistema in presenza di comportamenti utente subottimali o opportunistici.

**Capitolo 8** Presenta la **Fase 3** e il meccanismo adattivo GAP (*Game-theoretic Adaptive Policy*), progettato per mitigare le inefficienze comportamentali attraverso tecniche di apprendimento dinamico.

**Capitolo 9** Conclude il lavoro sintetizzando i risultati ottenuti, discutendo le implicazioni pratiche per la progettazione di sistemi reali e delineando i possibili sviluppi futuri.

# Capitolo 2

## Stato dell'Arte e Fondamenti Teorici

### 2.1 Evoluzione dei Sistemi di Sensing

Nel contesto dell'intelligenza ambientale (*Ambient Intelligence*) e delle *Smart City*, la necessità di raccogliere dati territoriali in tempo reale e in modo capillare è emersa con forza a partire dalla fine degli anni '90. Le prime risposte tecnologiche a questa esigenza si sono basate sul paradigma delle **Wireless Sensor Network** (WSN), reti costituite da insiemi di nodi sensori statici dispiegati fisicamente sul territorio con un notevole sforzo logistico ed economico [10, 6]. Applicazioni tipiche includono la sensoristica industriale per la prevenzione sismica e i sistemi di monitoraggio strutturale (*Structural Health Monitoring*) di ponti, gallerie e infrastrutture critiche.

Il **modello architettonale WSN** assume in genere una topologia prevalentemente statica, con nodi stazionari e sincronizzati che comunicano secondo uno schema di tipo *point-to-sink*. Il deployment è vincolato da problemi severi di copertura radio e consumo energetico, che impongono una precisa calibrazione della densità dei nodi e della frequenza di campionamento  $f_t$ . In letteratura, viene messo in realtà come la densità di campionamento e la scalabilità delle WSN risultino, in pratica, limitate dagli elevati costi di installazione e manutenzione, con il rischio di ottenere *coperture frammentate* e una scarsa capacità di adattamento ai pattern dinamici e spesso imprevedibili della mobilità urbana [6, 3].

L'avvento dei dispositivi mobili multimodali (*smartphone*, *wearable*, *tablet*), dotati di un ampio set di sensori eterogenei — GPS, accelerometri, magnetometri, barometri, sensori di prossimità — e di interfacce di comunicazione a banda larga (Wi-Fi, LTE, 5G), ha progressivamente cambiato questo quadro. A partire dal 2010, le stime dell'ITU indicano la presenza di miliardi di dispositivi in uso attivo; ciascuno di essi rappresenta, di fatto, un potenziale *nodo sensoriale mobile* distribuito in modo casuale e capillare sull'intero territorio [10, 6].

La differenza sostanziale tra WSN e MCS non è quindi solo architettonica, ma riguarda anche la *natura della fonte del dato*:

- Le WSN raccolgono prevalentemente dati "ambientali" tramite hardware proprietario dedicato, installato per uno scopo specifico;
- Le reti di *Mobile Crowdsensing*, al contrario, acquisiscono dati "sociali", "comportamentali" e "spazio-temporali" sfruttando come vettore la mobilità naturale e l'*agency* della popolazione urbana.

Questo cambio di prospettiva rende possibile una copertura molto più densa e flessibile, ma introduce anche nuove sfide: *l'eterogeneità dei profili comportamentali* degli utenti, *l'imprevedibilità dei pattern* di copertura spazio-temporale e la necessità di rispettare vincoli stringenti di *privacy-by-design* per garantire l'affidabilità e l'accettabilità sociale del servizio [3].

### 2.1.1 Definizione di Mobile Crowdsensing

Per inquadrare con precisione il paradigma, è utile introdurre una definizione esplicita.

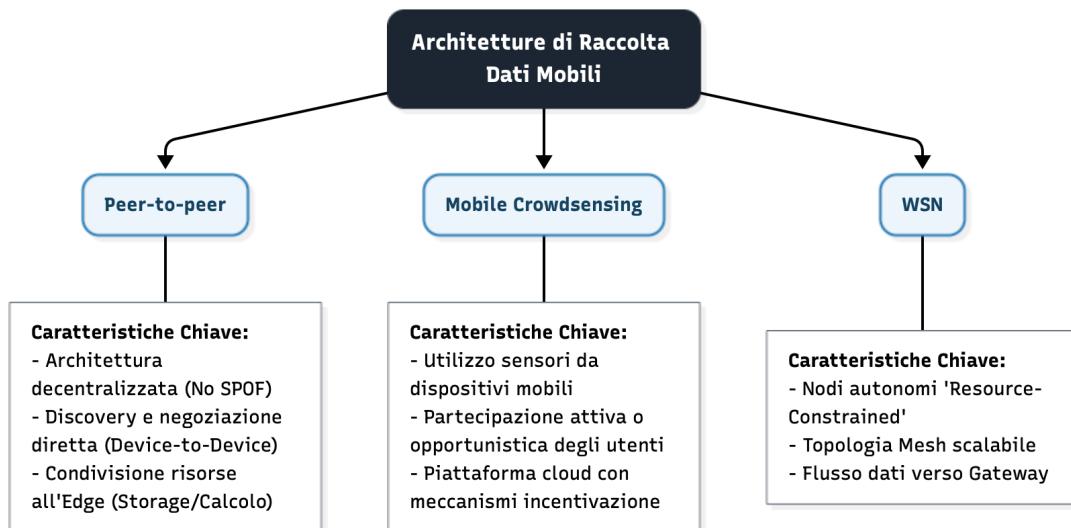
**Definizione 2.1** (Mobile Crowdsensing). Il **Mobile Crowdsensing (MCS)** è un paradigma di *sensing* partecipativo in cui una piattaforma digitale, generalmente centralizzata e basata su infrastrutture cloud, coordina e incentiva la raccolta, la trasmissione e l'aggregazione di dati provenienti dal multistrato sensoriale di una vasta popolazione di dispositivi mobili. I dispositivi sono detenuti da utenti umani (gli *agenti*), che compiono azioni volontarie o semi-automatiche (misurazioni, fotografie, *logging*, annotazioni semantiche). L'MCS sfrutta la capillarità e la mobilità intrinseca della popolazione per massimizzare al tempo stesso la granularità informativa e la copertura spazio-temporale delle osservazioni, superando i limiti di scalabilità delle reti di sensori fissi [6, 3].

Questa definizione mette in evidenza la *scalabilità sociale* del MCS: la *crowd*, composta da utenti reali che già portano con sé dispositivi mobili, può essere coinvolta — previo un adeguato meccanismo di incentivazione — per raccogliere informazioni ad alta risoluzione su fenomeni urbani complessi. In molti casi, ottenere una simile densità e varietà di dati con hardware dedicato richiederebbe investimenti proibitivi; il MCS consente di ridurre drasticamente tali barriere di ingresso, trasferendo parte dei costi infrastrutturali sul parco dispositivi esistente (cfr. Capitolo 1, Sezione 1.1).

## 2.1.2 Architettura e tassonomia dei sistemi MCS

### Tassonomia architetturale

I sistemi di raccolta dati wireless possono essere ricondotti a tre macro-architetture che rappresentano una vera e propria progressione storica e tecnologica [3]: le reti di sensori statiche (**WSN**), i sistemi di *peer-to-peer sensing* e le piattaforme di *Mobile Crowdsensing* (**MCS**). Ciascun paradigma adotta un diverso modello di coordinamento, richiede livelli differenti di investimento infrastrutturale e offre gradi diversi di flessibilità e scalabilità.



**Figura 2.1:** Schema delle principali architetture di raccolta dati mobili. La figura mostra la tassonomia fra WSN, paradigmi peer-to-peer e Mobile Crowdsensing, evidenziando per ciascuno le caratteristiche chiave di deployment e di funzionamento, in linea con le classificazioni proposte nelle principali survey di letteratura [10, 6, 3].

Nelle **WSN** tradizionali, la logica del sistema è fortemente centrata sull'infrastruttura installata ad hoc: i nodi sensori sono posizionati in punti specifici, spesso difficilmente accessibili, e inviano le misure a un unico punto di raccolta o a un insieme ristretto di gateway. Questo paradigma garantisce un buon controllo hardware e un flusso dati prevedibile, ma soffre per costi elevati di installazione e manutenzione, copertura rigida e difficoltà ad adattarsi rapidamente a cambiamenti nella domanda informativa o nei pattern di mobilità.

I sistemi di **peer-to-peer sensing** riducono il grado di centralizzazione, valorizzando la comunicazione diretta tra dispositivi vicini. In questo scenario la scoperta dei nodi, la negoziazione dello scambio dati e la condivisione delle risorse avvengono in modo distribuito, sfruttando collegamenti locali (ad esempio Bluetooth o Wi-Fi Direct). Questa scelta migliora resilienza e privacy a livello di prossimità, ma rende più complesso ottenere una visione globale coerente e garantire copertura omogenea sull'intero territorio urbano.

Le piattaforme di **Mobile Crowdsensing** rappresentano l'ultimo passo di questa evoluzione. Esse combinano un forte coordinamento centrale, tipicamente cloud-based, con una fase di sensing distribuita affidata ai dispositivi degli utenti. La piattaforma definisce e pubblica i task, raccoglie e valida i contributi e aggrega le osservazioni, mentre gli utenti partecipano fornendo dati ambientali e contestuali direttamente dai propri smartphone. Questo modello offre una scalabilità potenzialmente molto elevata e una copertura spazio-temporale fine, ma introduce sfide specifiche in termini di gestione degli incentivi, tutela della privacy e controllo della qualità del dato.

Architettura	Utilizzo	Dati	Tipologia
<b>WSN</b>	Rete di sensori fissi installati su infrastruttura dedicata	Misure fisiche e ambientali di natura oggettiva	Raccolta automatica, senza coinvolgimento diretto dell'utente
<b>Peer-to-peer Sensing</b>	Connessioni dirette tra dispositivi vicini, senza backend centrale	Informazioni locali di prossimità e di contesto	Partecipazione perlopiù passiva degli utenti-dispositivi
<b>Mobile Crowdsensing</b>	Dispositivi mobili personali che raccolgono dati durante gli spostamenti	Dati ambientali e contestuali, arricchiti da valutazioni soggettive	Partecipazione attiva o opportunistica degli utenti

**Tabella 2.1:** Sintesi dei tre principali paradigmi di raccolta dati mobili.

## Architettura a Tre Livelli

Nell'ambito dei sistemi di Mobile Crowdsensing (MCS) più evoluti, la letteratura scientifica converge verso un'architettura logica organizzata in una struttura stratificata a tre livelli: [3]

### 1. Layer Piattaforma (Platform Tier):

Rappresenta il cervello del sistema, generalmente ospitato su infrastrutture cloud. Questo livello ha la responsabilità globale dell'intera campagna di sensing e gestisce il ciclo di vita dei dati: dalla raccolta dei dati grezzi fino all'analisi dettagliata e specializzata. Le sue funzioni principali includono:

- **Gestione dei Task:** Creazione, pianificazione (*scheduling*) e allocazione ottimale dei compiti agli utenti più idonei (*task assignment*), basandosi su criteri di efficienza o copertura spaziale.
- **Elaborazione dei Dati:** Una volta ricevuti i contributi, la piattaforma non si limita ad archiviarli, ma esegue processi di aggregazione e validazione. Qui vengono spesso applicati algoritmi di *Truth Discovery* per filtrare dati rumorosi o maliziosi [6].
- **Gestione dell'Ecosistema:** Include la distribuzione degli incentivi (pagamenti, crediti o reputazione) e l'applicazione rigorosa delle politiche di sicurezza, controllo degli accessi e anonimizzazione dei dati sensibili.

### 2. Layer Task (Task Tier):

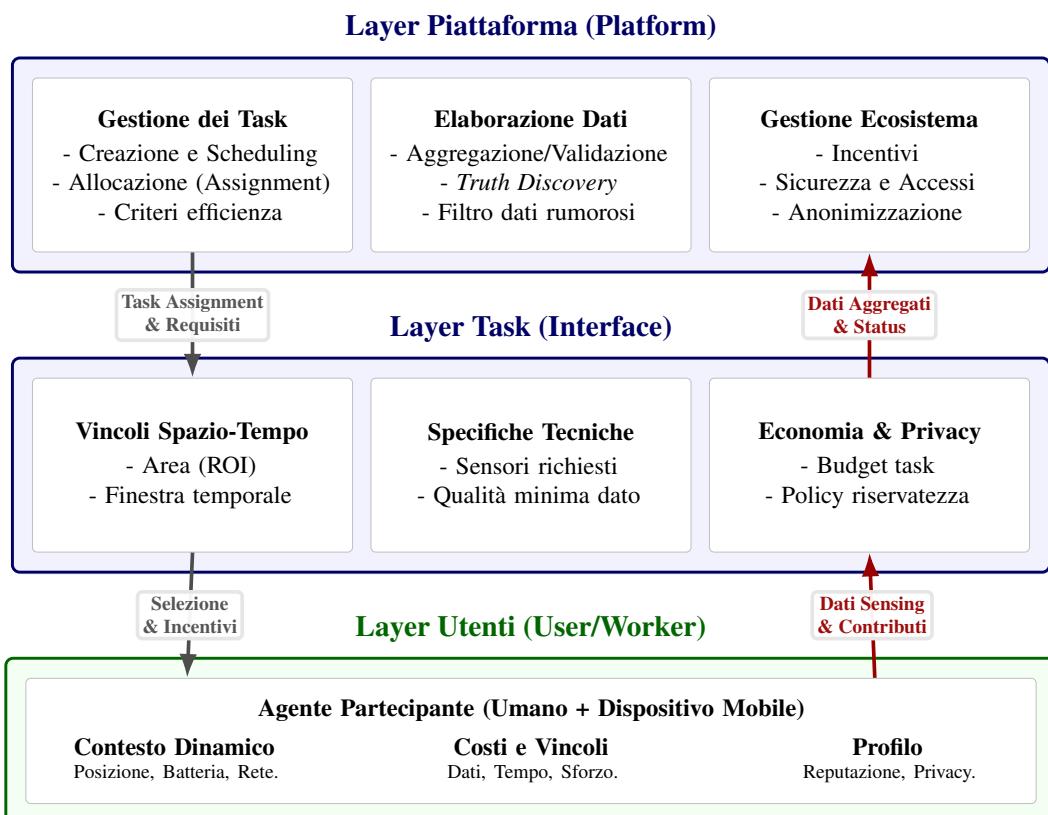
Questo livello agisce come un'interfaccia logica o un livello di astrazione tra le richieste della piattaforma e le capacità degli utenti. Non si tratta solo di una "lista di cose da fare", ma di una definizione formale dei requisiti della missione. Un task viene qui modellato attraverso parametri specifici:

- **Vincoli Spazio-Temporali:** L'area geografica di interesse (ROI) e la finestra temporale entro cui la rilevazione è valida.
- **Specifiche Tecniche:** La tipologia di sensori richiesti (es. accelerometro, fotocamera, microfono) e la qualità minima accettabile del dato.
- **Modello Economico e Privacy:** Il valore del budget allocato per il task (il "prezzo" del dato) e le policy di riservatezza richieste per quella specifica rilevazione.

### 3. Layer Utenti (User/Worker Tier):

Costituisce la base della piramide ed è composto dagli agenti partecipanti, intesi come l'unione inscindibile tra l'essere umano e il proprio dispositivo mobile (*smartphone* o *wearable*). Questo livello è intrinsecamente dinamico e imprevedibile. Ogni agente è caratterizzato da:

- **Contesto Dinamico:** Una posizione che varia continuamente nello spazio urbano e uno stato del dispositivo mutevole (livello di batteria, connettività disponibile).
- **Costi e Vincoli:** L'utente sostiene costi reali per partecipare, siano essi tangibili (consumo dati, batteria) o intangibili (tempo, sforzo fisico, attenzione).
- **Profilo Comportamentale:** Include lo storico dell'affidabilità (reputazione) e le preferenze personali di privacy che influenzano la decisione di accettare o rifiutare un task.



**Figura 2.2:** Architettura logica a tre livelli per sistemi MCS evoluti.

### 2.1.3 Acquisizione dei dati: Participatory vs Opportunistic Sensing

La modalità con cui avviene la raccolta del dato è un aspetto centrale nella progettazione dei sistemi MCS, perché incide sia sullo sforzo richiesto agli utenti sia sul tipo di informazione che la piattaforma può ottenere [3]. In letteratura si distinguono due modalità principali: il *Participatory Sensing*, in cui l'utente interviene in modo esplicito, e l'*Opportunistic Sensing*, in cui il dispositivo rileva dati in modo quasi trasparente rispetto alle azioni quotidiane [10].

**Participatory Sensing:** l'utente è parte attiva e consapevole del processo di sensing (*user-in-the-loop*). Riceve un task (notifica, richiesta nell'app), decide se accettarlo e quando eseguirlo e può anche controllare il dato prima dell'invio, ad esempio scegliendo quali foto o commenti condividere. Questa modalità è indicata per informazioni con forte contenuto semantico o soggettivo, come giudizi di qualità o segnalazioni puntuali di problemi. *Esempi:* scattare e caricare la foto di una buca stradale per una campagna di manutenzione; indicare tramite app il livello di affollamento di un mezzo pubblico [10]. Il vantaggio è il maggiore controllo per l'utente; il limite è che la partecipazione dipende dalla sua motivazione e disponibilità.

**Opportunistic Sensing:** in questo caso il dispositivo esegue il campionamento in modo automatico, seguendo politiche impostate dalla piattaforma, e l'intervento dell'utente è minimo. I dati diventano un *byproduct* della routine: lo smartphone viene usato normalmente, mentre un servizio in background raccoglie periodicamente misure (GPS, accelerometro, microfono, sensori ambientali) e le invia quando vi sono condizioni favorevoli, ad esempio buona connettività o batteria sufficiente. Questa modalità è adatta a fenomeni che richiedono misure frequenti e diffuse, come il monitoraggio del traffico o del rumore urbano. *Esempi:* il logging continuo delle tracce GPS dei taxi (come nel dataset CRAWDAD) o la raccolta automatica dei livelli di rumore durante gli spostamenti quotidiani [10, 3]. Offre copertura ampia e continua, ma pone sfide legate a consumo energetico e tutela della privacy.

Nella pratica, molte piattaforme MCS combinano le due modalità. Una stessa applicazione può mantenere sempre attivo un livello opportunistico a bassa frequenza e, quando necessario, attivare campagne partecipative mirate per ottenere contributi più ricchi o specifici (ad esempio in caso di eventi imprevisti o criticità locali). Il bilanciamento tra componente partecipativa e opportunistica dipende dal dominio applicativo, dai vincoli di privacy ed energia e dal tipo di incentivi messi a disposizione degli utenti.

## 2.1.4 Scenari applicativi

L'adozione del paradigma MCS ha dato luogo a numerose applicazioni reali in ambiti anche molto diversi tra loro. Questi esempi mostrano come la raccolta diffusa di dati tramite dispositivi mobili possa diventare uno strumento concreto di supporto alle decisioni operative e alla pianificazione a livello urbano e territoriale.

- **Smart City e infomobilità:** piattaforme come *Waze* rappresentano uno dei casi di maggior successo di crowdsensing su larga scala. L'app raccoglie continuamente segnalazioni e tracce di percorrenza dagli automobilisti e le usa per proporre percorsi alternativi, avvisare in tempo reale su incidenti, lavori in corso e congestioni e stimare i tempi di viaggio. In questo modo il sistema diventa un supporto dinamico alla gestione della viabilità e alla comunicazione verso i cittadini.
- **Monitoraggio ambientale:** progetti come *NoiseTube* trasformano lo smartphone in un fonometro mobile, permettendo di costruire mappe acustiche dettagliate a partire dai contributi di molti utenti. Lo stesso approccio può essere esteso alla misura di altri inquinanti, come il particolato o alcuni gas atmosferici, aprendo la strada a forme di *citizen science* in cui i cittadini partecipano direttamente alla produzione di dati ambientali ad alta risoluzione spaziale.
- **Pianificazione dei trasporti:** l'analisi dei dati di mobilità raccolti da taxi, autobus o servizi di ride-sharing consente di ricostruire i flussi reali di spostamento all'interno della città. Queste informazioni permettono di individuare colli di bottiglia infrastrutturali, valutare l'efficacia delle linee esistenti e progettare interventi mirati sulla rete di trasporto pubblico, basandosi su evidenze empiriche anziché su modelli puramente teorici.
- **Sanità pubblica (m-Health):** nel campo della salute pubblica, sistemi MCS sono stati impiegati per monitorare abitudini a rischio, tracciare i pattern di mobilità durante epidemie o raccogliere indicatori distribuiti di benessere (ad esempio, livelli di attività fisica o esposizione al rumore). In questi scenari è centrale l'adozione di tecniche di raccolta e aggregazione che preservino la privacy, così da coniugare il valore informativo dei dati con la tutela dei singoli individui.

### 2.1.5 Limiti principali per la sostenibilità delle piattaforme

Nonostante i vantaggi rispetto alle architetture tradizionali, i sistemi MCS presentano criticità intrinseche che devono essere affrontate in modo sistematico [15, 3]:

1. **Qualità e affidabilità del dato (*Data Trustworthiness*):** l'assenza di un controllo diretto sulla manutenzione dell'hardware, sulla calibrazione dei sensori e sull'onestà delle rilevazioni rende necessario ricorrere a strategie di aggregazione robuste, come algoritmi di *Truth Discovery* e sistemi di punteggio reputazionale. La variabilità dei sensori *consumer-grade* e l'eterogeneità dei comportamenti umani introducono rumore e bias che possono degradare la *Quality of Information* complessiva [15].
2. **Privacy e sicurezza:** dataset che includono tracce GPS, registrazioni audio o dati ambientali ricchi di contesto possono rivelare abitudini, luoghi sensibili e identità degli utenti. Senza adeguati meccanismi di anonimizzazione, *privacy-preserving* e controllo degli accessi, i partecipanti sono esposti a rischi concreti di re-identificazione e profilazione. L'architettura deve quindi incorporare i principi di *privacy-by-design* fin dalle prime fasi di progettazione [4].
3. **Sostenibilità della partecipazione (*Incentive Mechanisms*):** la partecipazione comporta costi certi per l'utente (batteria, traffico dati, tempo, deviazioni di percorso). In assenza di incentivi economici o sociali adeguati, calibrati sul trade-off tra costi e benefici, l'entusiasmo iniziale tende a esaurirsi rapidamente, dando luogo a fenomeni di *user fatigue*. Lo studio delle aste inverse, del *auction design* e delle economie comportamentali rappresenta quindi una linea di ricerca centrale, nonché uno dei focus principali di questa tesi [8, 19].

Queste considerazioni motivano l'analisi approfondita dei meccanismi di incentivazione che verrà sviluppata nelle sezioni e nei capitoli successivi, con particolare attenzione alla loro robustezza in scenari realistici.

## 2.2 Teoria dei Giochi nei sistemi MCS

La **Teoria dei Giochi** fornisce il quadro matematico entro cui modellare le interazioni strategiche tra agenti razionali — utenti e piattaforma — nei sistemi di Mobile Crowdsensing (MCS). Utilizzando questo formalismo, è possibile affrontare in modo rigoroso il problema dell’allocazione ottima delle risorse e progettare meccanismi di incentivazione (*Mechanism Design*) capaci di soddisfare simultaneamente vincoli critici quali la razionalità individuale (*Individual Rationality*), affinché la partecipazione sia vantaggiosa, e la compatibilità degli incentivi (*Incentive Compatibility*), per garantire la veridicità delle informazioni private dichiarate [5, 9]. In forma astratta, un ecosistema MCS può essere modellato come una tupla  $G = (N, S, u)$ , definita come segue:

- $N = \{1, 2, \dots, n\}$  è l’insieme finito degli agenti (o *giocatori*). In un tipico scenario MCS, questo insieme è partizionato in due sottoinsiemi funzionali: il *requester* (la piattaforma), che agisce come *mechanism designer*, e la popolazione di *workers* (gli utenti mobili), detentori delle risorse di sensing e delle informazioni private.
- $S = S_1 \times S_2 \times \dots \times S_n$  è lo spazio cartesiano dei profili di strategia. Qui  $S_i$  denota l’insieme delle azioni ammissibili per l’agente  $i$ . A seconda del modello,  $S_i$  può essere discreto (es. decisione binaria di accettare/rifiutare un task) o continuo (es. il valore dell’offerta  $b_i \in \mathbb{R}^+$  in un’asta inversa). Un profilo di strategia  $s = (s_1, \dots, s_n) \in S$  rappresenta l’esito congiunto delle scelte di tutti i partecipanti.
- $u = (u_1, u_2, \dots, u_n)$  è il vettore delle funzioni di utilità, dove  $u_i : S \rightarrow \mathbb{R}$  associa a ogni profilo di strategie  $s$  il *payoff* (guadagno netto) percepito dall’agente  $i$ . È fondamentale notare che, per la natura strategica del gioco, l’utilità  $u_i(s_i, s_{-i})$  dipende non solo dalla propria azione  $s_i$ , ma anche dalle strategie  $s_{-i}$  adottate dagli altri agenti.

Nel contesto MCS, per un utente  $i$ , l’utilità è tipicamente modellata come quasi-lineare:  $u_i = r_i - c_i$ , dove  $r_i$  è l’incentivo monetario ricevuto e  $c_i$  rappresenta il costo complessivo sostenuto (inclusivo di consumo energetico, tempo impiegato e potenziale perdita di privacy). Sulla base delle modalità di interazione tra gli agenti in  $N$  e della struttura delle strategie in  $S$ , è possibile operare una distinzione fondamentale tra i modelli di gioco, a seconda del livello di coordinamento consentito.

### 2.2.1 Giochi cooperativi e Non-Cooperativi

Nel contesto del Mobile Crowdsensing, la letteratura distingue due macro-categorie di modelli basandosi sulla capacità degli agenti di sottoscrivere accordi vincolanti (*binding agreements*).

**Giochi Cooperativi** In questo scenario, i partecipanti possono formare coalizioni stabili e negoziare la ripartizione di un valore aggregato. Il modello si fonda su una funzione caratteristica  $v(C)$  che assegna a ogni coalizione  $C \subseteq N$  il valore collettivo ottenibile [5].

L'obiettivo principale è garantire la stabilità della coalizione e l'equità nella distribuzione dei guadagni (tramite concetti come il *Core* o il *Valore di Shapley*). Sebbene concettualmente ideali per scenari collaborativi, nelle reti mobili reali questi modelli sono spesso penalizzati da elevati costi di comunicazione e dalla difficoltà di coordinare un gran numero di dispositivi eterogenei [3].

*Esempio applicativo:* In una campagna di monitoraggio ambientale, i tassisti potrebbero coalizzarsi per coprire l'intera area urbana, negoziando collettivamente un premio unico da spartire successivamente in base al contributo di ciascuno.

**Giochi Non-Cooperativi** Costituiscono il modello dominante per il MCS grazie alla loro scalabilità. Assumono che ogni agente  $i$  agisca in modo egoistico (*selfish*), massimizzando esclusivamente la propria funzione di utilità  $u_i$  in assenza di accordi vincolanti. Si assume inoltre che la razionalità degli agenti e le regole del gioco siano *common knowledge*.

Tale approccio porta a configurazioni di equilibrio (come l'Equilibrio di Nash) che spesso non sono Pareto-efficienti; la perdita di efficienza sociale dovuta alla mancanza di coordinamento centralizzato è quantificata dal *Prezzo dell'Anarchia* (PoA) [5, 3].

*Esempio applicativo:* Nel modello *user-centric*, ciascun taxista partecipa individualmente a un'asta inversa per aggiudicarsi i task di sensing. L'allocazione delle risorse emerge dalla pura competizione di prezzo (bidding) tra i singoli, senza alcuna cooperazione preventiva tra i conducenti.

## 2.2.2 Giochi Bayesiani

La struttura informativa del gioco rappresenta un elemento cruciale per il *Mechanism Design*. Mentre molti modelli teorici semplificati assumono l'ipotesi di **informazione completa** — ossia che tutti i giocatori conoscano le funzioni di utilità e i costi degli altri — nei sistemi MCS reali questa condizione è raramente plausibile. La piattaforma, infatti, non conosce i costi privati degli utenti, né la qualità istantanea dei loro sensori.

Il formalismo più appropriato per descrivere tale scenario è quello dei **Giochi Bayesiani**. In questo modello, ciascun agente  $i$  è caratterizzato da un *tipo* privato  $\theta_i$  (che può rappresentare il costo reale di sensing  $c_i$  o l'affidabilità del dispositivo), e gli altri giocatori non hanno accesso diretto a tale informazione [9, 12]. La piattaforma dispone unicamente di una distribuzione di probabilità a priori sui tipi, che ne sintetizza le aspettative (*beliefs*).

Questa asimmetria informativa solleva una questione centrale: come progettare meccanismi che inducano gli utenti a rivelare onestamente il proprio tipo privato? In assenza di incentivi adeguati, agenti razionali tenderanno inevitabilmente a manipolare le dichiarazioni (ad esempio sovrastimando i costi) per massimizzare il profitto, compromettendo l'efficienza complessiva del sistema [9].

## 2.2.3 Equilibrio di Nash e Strategie Dominanti

Per analizzare la stabilità delle interazioni strategiche, il concetto fondamentale è l'**Equilibrio di Nash (NE)**. Un profilo di strategie  $s^* = (s_1^*, \dots, s_n^*)$  costituisce un equilibrio di Nash se nessun giocatore ha incentivo a deviare unilateralmente dalla propria strategia, assunte come fisse le strategie degli altri:

$$u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*) \quad \forall i \in N, \forall s_i \in S_i.$$

In pratica, il NE descrive una situazione di stabilità in cui nessuno ha rimpianti rispetto alla propria scelta. Tuttavia, il concetto richiede che ogni agente sia in grado di prevedere correttamente le strategie altrui, un'ipotesi spesso troppo forte per utenti umani caratterizzati da razionalità limitata e incompleta conoscenza dell'ambiente.

Per ovviare a questa complessità decisionale, molti meccanismi di incentivazione nel MCS mirano a soluzioni in **strategie dominanti**. Una strategia si definisce dominante per l'agente  $i$  se massimizza la sua utilità a prescindere dalle azioni degli altri.

In questo quadro, un meccanismo d'asta è detto **truthful** (o *strategy-proof*) se la rivelazione veritiera del proprio costo  $c_i$  — ossia porre l'offerta  $b_i = c_i$  — è sempre una strategia dominante:

$$u_i(c_i, b_{-i}) \geq u_i(b'_i, b_{-i}) \quad \forall b'_i \neq c_i, \forall b_{-i}.$$

La proprietà di **Dominant Strategy Incentive Compatibility** (DSIC), supportata dal celebre *Revelation Principle* di Myerson, riduce drasticamente il carico cognitivo per l'utente: non è necessario speculare sulle mosse degli avversari o calcolare complessi equilibri; per massimizzare l'utilità attesa è sufficiente dichiarare il proprio costo reale [12].

## 2.2.4 Giochi di Stackelberg

Per modellare in modo più accurato l'interazione gerarchica tra piattaforma e utenti, è spesso utile ricorrere ai **Giochi di Stackelberg** (*Leader-Follower*). In questo schema sequenziale, la piattaforma agisce come *Leader*, fissando per prima i parametri chiave (es. budget totale, prezzi unitari o incentivi), mentre gli utenti (*Followers*) osservano tale mossa e reagiscono scegliendo la propria strategia migliore [14].

La soluzione del gioco, nota come **Equilibrio di Stackelberg** (SE), si ricava mediante induzione a ritroso (*backward induction*). Formalmente, il leader risolve un problema di ottimizzazione su due livelli (*bilevel optimization*):

$$\max_{s_L} u_L(s_L, s_F^*(s_L)),$$

dove  $s_L$  è la strategia del leader e  $s_F^*(s_L)$  rappresenta la risposta ottima collettiva dei follower a tale strategia. Nel contesto del MCS, questo modello descrive efficacemente i sistemi **platform-centric**, dove l'ente centrale controlla le tariffe, differenziandosi dai meccanismi *user-centric* basati su aste competitive.

### 2.2.5 Il meccanismo VCG

Per formalizzare la garanzia di veridicità nelle dichiarazioni, il punto di riferimento in letteratura è il meccanismo **Vickrey-Clarke-Groves (VCG)**. Tale modello estende l'intuizione dell'asta di Vickrey a scenari multi-agente complessi, promettendo simultaneamente l'efficienza allocativa (massimizzazione del valore globale) e la proprietà di *truthfulness* [9, 12]. Si consideri un'asta inversa per il MCS con un insieme di utenti  $N$ . Ogni utente  $i$  sostiene un costo privato  $c_i$  e sottomette un'offerta (*bid*)  $b_i$ . La piattaforma deve selezionare un sottoinsieme di vincitori  $S \subseteq N$  per massimizzare il **benessere sociale** (*Social Welfare*, SW), definito come la differenza tra il valore dei dati raccolti e i costi dichiarati:

$$SW(b) = \max_{S \subseteq N} \left( V(S) - \sum_{j \in S} b_j \right)$$

dove  $V(S)$  è la funzione di valutazione aggregata dei dati forniti dal gruppo  $S$ . Sia  $S^*$  l'insieme ottimale dei vincitori calcolato sulla base del vettore delle offerte  $b$ .

Il cuore del meccanismo risiede nella **regola pivot di Clarke**, che determina il pagamento  $p_i$  per ogni vincitore  $i \in S^*$ . L'idea è remunerare l'utente per il "valore aggiunto" che porta al sistema, calcolato come l'esternalità che la sua assenza causerebbe:

$$p_i = \underbrace{\max_{S' \subseteq N \setminus \{i\}} \left( V(S') - \sum_{j \in S'} b_j \right)}_{\text{SW ottimo senza l'agente } i} - \underbrace{\left( V(S^*) - \sum_{j \in S^*, j \neq i} b_j \right)}_{\text{SW degli altri agenti con } i \text{ presente}}$$

Analizzando l'utilità netta dell'utente  $i$  ( $u_i = p_i - c_i$ ) e isolando i termini che non dipendono dalla sua strategia  $b_i$  in una funzione  $h(b_{-i})$ , si ottiene:

$$u_i(b_i, b_{-i}) = \left( V(S^*) - \sum_{j \in S^*, j \neq i} b_j - c_i \right) + h(b_{-i})$$

dove il contenuto centrale rappresenta il benessere sociale valutato al costo reale  $c_i$ ; pertanto, l'agente massimizza la propria utilità solo favorendo *l'ottimo globale*, rendendo la dichiarazione veritiera ( $b_i = c_i$ ) una strategia dominante. Purtroppo, ciò non è applicabile all'attacco pratico, poiché determinare l'insieme ottimale  $S^*$  implica spesso la risoluzione di problemi NP-hard, rendendo indispensabile il ricorso a euristiche efficienti.

## 2.3 Sfide in contesti reali per un sistema MCS

L'applicazione pratica dei modelli di Teoria dei Giochi e di *Mechanism Design*, discussi nelle sezioni precedenti, deve confrontarsi con la complessità intrinseca degli scenari di *deployment* reale. Mentre i modelli matematici tendono ad assumere condizioni idealizzate — sensori perfetti, disponibilità costante di utenti, razionalità illimitata — un sistema MCS operativo deve affrontare problematiche stocastiche legate alla qualità del dato, alla sicurezza, ai vincoli energetici e alle dinamiche spazio-temporali della folla.

### 2.3.1 Qualità dell'Informazione (QoI)

La **Quality of Information (QoI)** rappresenta la metrica fondamentale per valutare l'utilità di una campagna di sensing. In un ambiente partecipativo non controllato, la QoI è esposta a molteplici fattori di degrado. I sensori *consumer-grade* integrati negli smartphone non sono calibrati professionalmente e risultano soggetti a *drift* delle misurazioni e a rumore termico; a ciò si aggiunge il fattore umano, poiché l'inesperienza degli utenti può produrre errori macroscopici di acquisizione (fotografie sfocate, microfoni ostruiti, orientamento errato del magnetometro). Anche condizioni ambientali avverse, come la scarsa luminosità o l'elevato rumore di fondo, possono compromettere la qualità delle rilevazioni pur in presenza di hardware funzionante e operatori attenti.

Un aspetto controintuitivo del MCS è che il semplice aumento del numero di partecipanti non garantisce un miglioramento lineare della QoI (*"more is not always better"*). In assenza di meccanismi di filtraggio, l'iniezione di dati rumorosi può degradare l'aggregato complessivo invece di arricchirlo. Per mitigare questi rischi, la letteratura propone due approcci complementari: l'adozione di algoritmi di **Truth Discovery** (ad esempio il *Bayesian Truth Serum*), che stimano la "verità" pesando i contributi in base alla loro convergenza statistica senza richiedere un *ground truth* noto a priori, e l'introduzione di sistemi di **Reputation Scoring**, che storicizzano l'affidabilità dei singoli utenti e ne modulano il peso nelle aggregazioni future.

### 2.3.2 Trustworthiness e Privacy

La fiducia (*trust*) in un sistema MCS è un concetto composito che riguarda sia l'affidabilità tecnica del dispositivo (*trust by reliability*) sia l'onestà comportamentale dell'agente (*trust by decision*). Sul piano della sicurezza, la piattaforma è vulnerabile a diverse tipologie di attacco, tra cui i **Sybil Attacks**, in cui un singolo utente crea molteplici identità fintizie per alterare l'aggregazione dei dati o ottenere indebitamente le ricompense, e il **Data Poisoning**, basato sull'invio intenzionale di dati falsi con l'obiettivo di distorcere le mappe di sensing o le decisioni derivate.

Parallelamente, la raccolta e la conservazione di tracce spazio-temporali solleva criticità significative in ambito **privacy**, esponendo gli utenti a rischi di re-identificazione e profilazione (*inference attacks*). Le contromisure architetturali più robuste includono la **Differential Privacy**, che inietta rumore matematico controllato (ad esempio di tipo laplaciano o gaussiano) per rendere indistinguibile il contributo del singolo individuo, e protocolli di **Secure Multi-party Computation (SMP)**, che consentono di calcolare statistiche aggregate (come la media del rumore in un quartiere) senza che la piattaforma abbia mai accesso ai singoli input in chiaro.

### 2.3.3 Vincoli Energetici

Dal punto di vista dell'utente, l'energia è la risorsa critica per eccellenza. Il *costo energetico* della partecipazione si distribuisce su tre componenti principali: l'attivazione dei sensori (particolarmente onerosa per GPS e giroscopio), l'elaborazione locale dei dati (ad esempio per crittografia o compressione) e, soprattutto, la trasmissione dei dati verso la rete. In molti casi, l'*upload* attraverso interfacce radio cellulari (LTE/5G) costituisce la voce dominante di consumo.

Per garantire la sostenibilità a lungo termine, i meccanismi di selezione dovrebbero essere **battery-aware**, ovvero tenere esplicitamente conto dello stato di carica residua e della situazione d'uso del dispositivo, privilegiando ad esempio utenti in fase di ricarica o con ampie riserve energetiche. Inoltre, l'adozione di strategie di **piggybacking** consente di ridurre il consumo marginale accodando la trasmissione dei dati di sensing a sessioni di comunicazione già attive (come chiamate vocali, sessioni di navigazione web o sincronizzazioni periodiche di altre applicazioni).

### 2.3.4 Copertura Spaziale e Mobilità

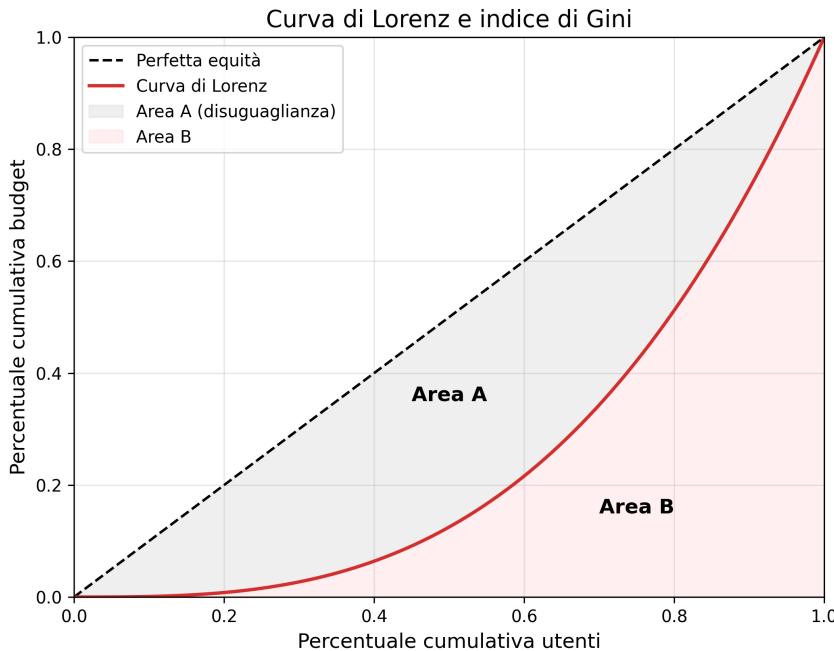
La natura dinamica della folla introduce sfide non banali nell'assegnazione dei task. La distribuzione spaziale degli utenti segue spesso leggi di potenza (*power laws*), con forti eterogeneità: densità molto elevate nelle aree centrali e scarsità marcata nelle periferie. I meccanismi di incentivazione devono quindi guidare attivamente lo spostamento verso le cosiddette aree " fredde", per evitare la formazione di buchi informativi persistenti.

Per i task *location-dependent*, la piattaforma necessita di modelli predittivi della disponibilità spaziale degli utenti. Mentre nelle simulazioni teoriche si ricorre frequentemente a modelli sintetici di mobilità (come il *Random Waypoint*), nei sistemi reali è preferibile sfruttare tracce GPS storiche per addestrare modelli probabilistici (ad esempio catene di Markov) in grado di stimare la probabilità che un determinato utente sia disponibile in una certa cella spaziale in un dato intervallo temporale.

### 2.3.5 Equità Distributiva: Indice di Gini

Oltre alla pura efficienza economica, un sistema MCS deve tenere conto dell'**equità distributiva** (*fairness*) nella ripartizione dei compensi. Meccanismi che premiano sistematicamente solo gli utenti più efficienti (ovvero quelli con costi marginali minimi) producono, nel medio periodo, fenomeni di *starvation* per la maggioranza dei partecipanti, che tendono ad abbandonare la piattaforma riducendone resilienza e capacità di copertura.

Per quantificare la disuguaglianza nella distribuzione dei ricavi, si utilizzano la **Curva di Lorenz** e il corrispondente **Coefficiente di Gini**. La Figura 2.3 illustra la rappresentazione grafica standard: l'asse delle ascisse riporta la percentuale cumulativa degli utenti (ordinati per guadagno crescente), mentre l'asse delle ordinate mostra la percentuale cumulativa del budget distribuito. La diagonale a  $45^\circ$  rappresenta la perfetta equità ( $G = 0$ ), mentre la curva di Lorenz evidenzia la distribuzione reale.



**Figura 2.3:** Curva di Lorenz e coefficiente di Gini. La linea tratteggiata a  $45^\circ$  rappresenta la perfetta equità, mentre la curva continua descrive la distribuzione reale dei compensi. Il coefficiente di Gini è proporzionale all'area compresa tra le due curve: valori elevati indicano una forte concentrazione delle ricompense nelle mani di pochi utenti.

Il coefficiente di Gini  $G$  è definito come rapporto tra l'area  $A$  compresa tra la diagonale di equità e la curva di Lorenz e l'area totale  $A + B$ :

$$G = \frac{A}{A + B}.$$

Il valore di  $G$  è compreso nell'intervallo  $[0, 1]$ :  $G = 0$  corrisponde a una distribuzione perfettamente equa (tutti ricevono lo stesso compenso), mentre  $G = 1$  rappresenta la massima disuguaglianza (un solo utente assorbe l'intero budget). Nel MCS, l'analisi del Gini rende evidente il trade-off tra **efficienza** ed **equità**: meccanismi estremamente efficienti tendono a produrre valori di  $G$  elevati. Mantenere questo indicatore entro soglie accettabili è fondamentale per la sostenibilità sociale della piattaforma.

### 2.3.6 Analisi dei Pattern Temporali e Segmentazione dei Regimi

I sistemi di Mobile Crowdsensing lavorano in scenari urbani che cambiano continuamente, influenzati dall’alternanza giorno/notte, dai giorni festivi e dalla stagionalità. Per creare algoritmi solidi, capaci di funzionare in condizioni così diverse, è fondamentale gestire questa complessità temporale seguendo un metodo strutturato.

**Discretizzazione e Feature Extraction** Il primo passo consiste nel suddividere l’orizzonte temporale in *slot* di durata fissa (es. finestre orarie). Per ogni intervallo, vengono calcolate le metriche che descrivono lo stato del sistema, costruendo un vettore delle feature che include variabili chiave quali la densità di utenti attivi, il volume di task generati e la dispersione geografica della flotta. Questa trasformazione converte il flusso continuo di dati grezzi in una serie temporale discreta, pronta per l’analisi statistica.

**Clustering dei Regimi Operativi** Per identificare pattern ricorrenti senza imporre soglie arbitrarie, si adottano tecniche di apprendimento non supervisionato (*Unsupervised Learning*). L’algoritmo **K-Means** viene utilizzato per raggruppare gli slot temporali in  $k$  cluster distinti, basandosi sulla similarità delle caratteristiche osservate. La qualità del raggruppamento e la scelta del numero ottimale di cluster sono guidate da metriche quantitative come il **Silhouette Score**, che valuta la coerenza interna di ogni gruppo.

**Caratterizzazione On-Peak vs Off-Peak** L’analisi dei centroidi dei cluster permette di interpretare semanticamente i regimi operativi. Tipicamente emergono due macro-scenari contrapposti:

- **Regime On-Peak (Alta Domanda):** Caratterizzato da un’elevata densità di utenti e abbondanza di risorse. In questo scenario, la piattaforma può sfruttare la competizione tra i partecipanti per ottimizzare i costi.
- **Regime Off-Peak (Bassa Domanda):** Caratterizzato da scarsità di offerta e potenziali vuoti di copertura. Qui il rischio principale è l’impossibilità di servire i task, costringendo il sistema a incrementare gli incentivi per garantire la partecipazione.

Riconoscere questi scenari è fondamentale per superare i limiti delle politiche statiche: invece di applicare una logica unica per tutte le situazioni, il sistema può così adattare gli incentivi in tempo reale, rispondendo in modo intelligente alle variazioni della domanda e dell’offerta.

## 2.4 Razionalità Limitata (Bounded Rationality)

La modellazione classica dei meccanismi d'asta e della Teoria dei Giochi algoritmica poggia, tradizionalmente, sull'assunto neoclassico della **razionalità perfetta** (spesso personificata nell'*Homo Economicus*). All'interno di questo paradigma, gli agenti sono idealizzati come decisori infallibili: si presume dispongano di una capacità computazionale illimitata, abbiano accesso a una conoscenza completa delle regole del gioco e agiscano secondo preferenze perfettamente coerenti e stabili nel tempo. Tuttavia, quando spostiamo l'attenzione dai modelli teorici ai sistemi di *Mobile Crowdsensing* (MCS) reali, questa impalcatura scricchiola.

Nei contesti urbani, gli agenti sono esseri umani che interagiscono tramite smartphone in ambienti dinamici e spesso caotici. L'ipotesi di onniscienza e infallibilità risulta non solo eccessivamente semplificatrice, ma potenzialmente dannosa per la capacità predittiva dei modelli: ignorare la natura umana degli utenti rischia di portare alla progettazione di meccanismi che funzionano sulla carta, ma falliscono sul campo.

### 2.4.1 Bias Cognitivi e Limitazione: il modello di Herbert Simon

Fu il premio Nobel Herbert Simon, già nel 1955, a introdurre il concetto rivoluzionario di **razionalità limitata** (*bounded rationality*) per colmare il divario tra la teoria economica e l'osservazione empirica del comportamento umano [16]. Secondo la prospettiva di Simon, la razionalità degli individui non è assoluta, ma è vincolata da tre fattori strutturali ineludibili:

1. **Limiti cognitivi:** la mente umana possiede risorse finite in termini di attenzione, memoria di lavoro e capacità di calcolo.
2. **Imperfezione informativa:** raramente un decisore dispone di tutte le informazioni necessarie; spesso opera in condizioni di incertezza o ambiguità.
3. **Vincoli temporali:** la pressione del tempo impedisce di valutare esaustivamente tutte le alternative possibili prima di agire.

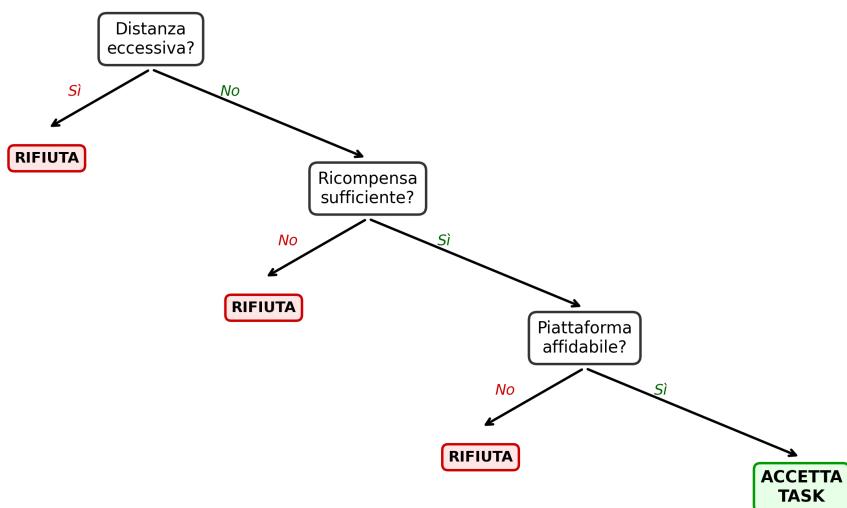
Di conseguenza, gli agenti reali abbandonano la pretesa di raggiungere l'ottimo globale in favore del criterio pragmatico del **satisficing** (crasi di *satisfy* e *suffice*). L'obiettivo dell'utente non è massimizzare matematicamente una funzione di utilità complessa, ma individuare una soluzione che sia "sufficientemente buona" da superare una soglia minima di accettabilità e benessere personale. Per navigare la complessità delle decisioni quotidiane senza rimanerne paralizzati, gli individui si affidano a scorciatoie mentali note come **euristiche** (*fast-and-frugal heuristics*). Sebbene queste strategie siano efficienti ed evolutivamente vantaggiose, introducono deviazioni sistematiche dalla logica razionale, note come **bias cognitivi** [7].

Nel contesto specifico del Crowdsensing, dove le micro-decisioni devono essere prese rapidamente, i bias che influenzano maggiormente il comportamento degli utenti sono:

- **Anchoring bias (Effetto Ancoraggio):** È la tendenza a fare eccessivo affidamento sulla prima informazione offerta ("ancora") quando si deve prendere una decisione. Ad esempio, se l'interfaccia della piattaforma suggerisce un prezzo iniziale basso per un task, l'utente potrebbe percepire come "vantaggiosa" un'offerta di poco superiore, anche se questa rimane oggettivamente sottostimata rispetto ai costi reali.
- **Loss aversion (Avversione alla perdita):** In psicologia economica, il dispiacere causato da una perdita è percepito con un'intensità molto maggiore rispetto al piacere derivante da un guadagno di pari entità. Nel MCS, questo significa che la paura di "sprecare" batteria o benzina inutilmente pesa più della prospettiva di guadagnare pochi euro, portando gli utenti ad adottare strategie di offerta eccessivamente prudenti o difensive.
- **Overconfidence (Eccesso di fiducia):** Al polo opposto, si osserva spesso una sovrastima delle proprie capacità o una sottostima dei rischi. Un utente potrebbe credere erroneamente di poter completare un task in meno tempo o con meno sforzo di quanto realmente necessario, portandolo a fare offerte troppo basse che si riveleranno non profittevoli.

### 2.4.2 Modelli Fast-and-Frugal Trees (FFTs)

Per tradurre queste osservazioni psicologiche in un modello operativo e simulabile, la letteratura suggerisce l'adozione dei **Fast-and-Frugal Trees (FFTs)** [7]. Un FFT è un modello decisionale estremamente snello, caratterizzato da una struttura lessicografica e *non compensatoria*. A differenza dei modelli complessi in cui un difetto su un attributo può essere compensato da un pregio su un altro, nell'FFT l'agente esamina una sequenza di indizi (*cues*) in un ordine di priorità rigido. A ogni passo (nodo), l'agente può prendere una decisione definitiva.



**Figura 2.4:** Struttura decisionale Fast-and-Frugal Tree (FFT) a cascata applicata al MCS. Il processo inizia dal nodo radice in alto a sinistra e procede diagonalmente. A ogni bivio, il mancato soddisfacimento di un criterio critico (distanza, ricompensa) comporta l'immediata uscita dal sistema (rifiuto), senza che l'utente debba elaborare tutte le informazioni disponibili.

Come illustrato in Figura 2.4, questo approccio modella fedelmente la natura "sbrigativa" dell'interazione umana su dispositivi mobili. L'utente non esegue un calcolo complesso di utilità attesa, ma applica una serie di filtri rapidi: se un task è troppo lontano o paga troppo poco, viene scartato immediatamente, rendendo irrilevanti eventuali altre caratteristiche positive.

L'integrazione di questi modelli decisionali nelle simulazioni è cruciale: permette di superare l'astrazione dell'agente perfettamente razionale e di comprendere come le soglie di attenzione individuali influenzino, a livello aggregato, la partecipazione alla campagna di crowdsensing. Questo passaggio dal comportamento micro-individuale alle dinamiche macro-sistemiche ci porta ad analizzare come tali euristiche impattino concretamente il meccanismo d'asta.

### 2.4.3 Impatto sui Meccanismi d'Asta

L'introduzione della razionalità limitata nello scenario di simulazione non è un mero esercizio accademico, ma ha ripercussioni profonde sulle dinamiche dell'asta. Le deviazioni dal comportamento teorico generano fenomeni che, se ignorati, renderebbero le previsioni inaffidabili. Tra gli effetti più rilevanti osservati si segnalano:

1. **Incertezza e rumore nella stima dei costi:** Nel mondo reale, un utente non conosce il valore esatto di  $c_i$  (il costo per eseguire il task). Agisce sulla base di una stima approssimativa, influenzata dal traffico, dallo stato della batteria o semplicemente dall'umore. Questo introduce un "rumore" stocastico nel sistema che i modelli deterministici non prevedono.
2. **Overbidding strategico (Prudenza):** Anche in presenza di meccanismi *truthful* (che incentivano la verità), l'avversione al rischio spinge gli utenti a chiedere un compenso superiore ai propri costi stimati ( $b_i > c_i$ ). Questo "marginе di sicurezza" serve a proteggersi da imprevisti, ma ha l'effetto collaterale di alzare i costi complessivi per la piattaforma.
3. **Underbidding opportunistico (Winner's Curse):** Utenti inesperti o troppo sicuri di sé (overconfidence) potrebbero offrire meno dei propri costi reali ( $b_i < c_i$ ) nella speranza di vincere l'asta. Sebbene possano vincere nel breve termine, operano in perdita; questo genera frustrazione e porta inevitabilmente all'abbandono della piattaforma (*churn*), danneggiando la sostenibilità del sistema a lungo termine.

Tali evidenze dimostrano che le garanzie matematiche di *truthfulness* non sono sufficienti, da sole, a governare il fattore umano. Diventa imperativo non solo progettare algoritmi robusti, ma testarli empiricamente contro popolazioni simulate dotate di razionalità limitata, approccio che costituisce il cuore della validazione sperimentale di questa tesi.

## 2.5 Submodularità e Ottimizzazione

Al termine della fase di raccolta delle offerte (*bidding*), la piattaforma affronta il problema decisionale critico: la determinazione dei vincitori (*Winner Determination*). L'obiettivo è selezionare il sottoinsieme di utenti che massimizzi il valore complessivo dei dati raccolti (ad esempio, ottimizzando la copertura spaziale o la qualità informativa), nel rigoroso rispetto del vincolo di budget prefissato.

Si tratta di un classico problema di **ottimizzazione combinatoria**: teoricamente, la piattaforma dovrebbe valutare ogni possibile combinazione di utenti per individuare la coalizione ottimale. Tuttavia, al crescere del numero di partecipanti, lo spazio delle soluzioni esplode esponenzialmente, rendendo il calcolo della soluzione esatta matematicamente intrattabile (un problema noto come *NP-hard*). Per gestire tale complessità computazionale senza rinunciare alla qualità della soluzione, si ricorre alla teoria delle **funzioni submodulari**, che consente di ottenere allocazioni efficienti in tempi ridotti.

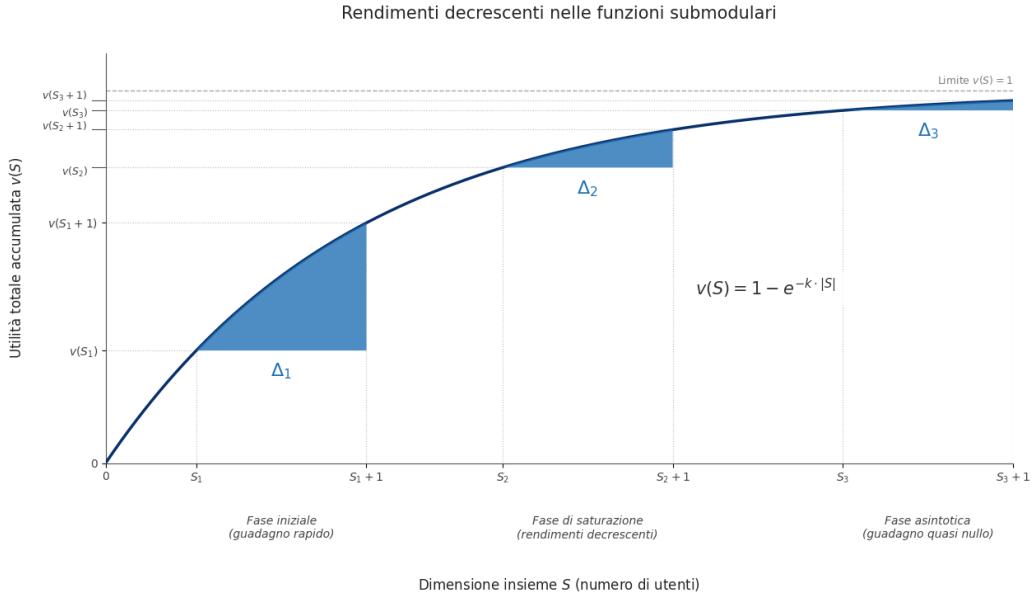
### 2.5.1 Proprietà di Submodularità

Nel contesto del Mobile Crowdsensing (MCS), la funzione di valutazione  $v(S)$  — che quantifica l'utilità aggregata di un insieme di utenti  $S$  — gode frequentemente della proprietà di **submodularità**.

**Definizione 2.2** (Funzione submodulare). Sia  $f : 2^{\mathcal{U}} \rightarrow \mathbb{R}$  una funzione definita su tutti i possibili sottoinsiemi di utenti  $\mathcal{U}$ . La funzione  $f$  è detta *submodulare* se, per ogni coppia di insiemi  $A \subseteq B \subseteq \mathcal{U}$  e per ogni nuovo elemento  $x \in \mathcal{U} \setminus B$ , vale la diseguaglianza:

$$f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B).$$

In termini intuitivi, questa definizione formalizza il principio economico dei **rendimenti marginali decrescenti**. Il contributo marginale di un singolo utente  $x$  non è un valore assoluto, ma dipende dal contesto: è massimo quando l'insieme di utenti già selezionati  $A$  è ridotto (o vuoto), mentre diminuisce progressivamente quando l'insieme  $B$  diventa ampio e ricco di informazioni (Figura 2.5).



**Figura 2.5:** Visualizzazione dei rendimenti marginali decrescenti. L'area colorata rappresenta l'utilità totale  $v(S)$  accumulata. Le barre verticali indicano il valore aggiunto da un nuovo utente in fasi diverse: il contributo  $\Delta_1$  (fase iniziale) è nettamente superiore a  $\Delta_2$  (fase di saturazione). Questo andamento riflette la naturale ridondanza dei dati spaziali nel crowdsensing.

Si consideri l'esempio pratico di un task di copertura fotografica urbana:

- Il contributo del **primo utente** è elevato, in quanto copre un'area o un evento precedentemente non osservato, colmando una lacuna informativa totale.
- Il contributo di un utente successivo, che osserva la stessa scena o un'area adiacente già coperta, è marginale a causa della ridondanza informativa.

Assumendo che la funzione sia anche **monotona** (l'aggiunta di risorse non decrementa mai l'utilità totale, ovvero  $A \subseteq B \Rightarrow f(A) \leq f(B)$ ), la submodularità rappresenta il modello matematico di riferimento per descrivere l'aggregazione di informazioni in sistemi distribuiti.

### 2.5.2 Algoritmi Greedy

La ricerca della soluzione ottima globale  $S^*$  tramite metodi di forza bruta ("Brute Force") è impraticabile su larga scala. Fortunatamente, la struttura submodulare del problema legittima l'utilizzo di **algoritmi greedy** (golosi). Tale approccio costruisce la soluzione in modo iterativo e miope (*myopic*), selezionando a ogni passo l'utente che offre il miglior incremento locale della funzione obiettivo, senza valutare le conseguenze future.

Sebbene semplice, questa strategia è supportata da solide garanzie teoriche, come dimostrato dal celebre teorema di Nemhauser, Wolsey e Fisher:

**Teorema 2.1** (Nemhauser-Wolsey-Fisher, 1978). *Se la funzione obiettivo  $f$  è monotona, submodulare e non negativa, un algoritmo greedy che seleziona iterativamente l'elemento con il massimo guadagno marginale produce una soluzione  $S_{\text{greedy}}$  tale che:*

$$f(S_{\text{greedy}}) \geq \left(1 - \frac{1}{e}\right) f(S^*) \approx 0.632 f(S^*),$$

dove  $S^*$  è la soluzione ottima globale ed  $e$  è la base dei logaritmi naturali [13].

Questo risultato fornisce un *lower bound* fondamentale: nel caso peggiore, l'algoritmo greedy garantisce almeno il **63%** dell'efficienza della soluzione ottima (spesso irraggiungibile). Nelle applicazioni reali di MCS, le prestazioni osservate sono tipicamente molto superiori a questo limite teorico, approssimando l'ottimo con un costo computazionale polinomiale anziché esponenziale.

### 2.5.3 Selezione dei Vincitori (Winner Determination)

Nel meccanismo d'asta IMCU proposto in questa tesi, l'approccio greedy è impiegato per risolvere il problema di **Winner Determination**. L'algoritmo parte da un insieme vuoto  $S_0 = \emptyset$  e aggiunge iterativamente l'utente  $i$  che massimizza il **guadagno marginale netto**:

$$\delta_i(S_{t-1}) = v(S_{t-1} \cup \{i\}) - v(S_{t-1}) - b_i.$$

In ogni iterazione  $t$ , il sistema valuta il bilancio tra l'incremento di utilità portato dall'utente e il prezzo  $b_i$  richiesto. Il processo termina all'esaurimento del budget o quando nessun utente disponibile fornisce un contributo netto positivo.

Per ottimizzare l'esecuzione su dataset estesi e garantire la scalabilità, si implementa la tecnica della **Lazy Evaluation** (o *Accelerated Greedy*). Tale metodo sfrutta la proprietà di submodularità per cui i guadagni marginali sono monotoni non-crescenti: il valore aggiunto di un utente non può aumentare man mano che l'insieme dei vincitori cresce. Di conseguenza, non è necessario ricalcolare il valore attuale di tutti i candidati a ogni passo; se il valore "vecchio" (calcolato in iterazioni precedenti) di un utente è già inferiore al valore attuale del miglior candidato corrente, quell'utente può essere scartato senza ulteriori calcoli.

# Capitolo 3

## Modello Matematico del Sistema e degli Agenti

### 3.1 Introduzione al Modello Computazionale

La valutazione rigorosa delle performance di un algoritmo di incentivazione non può prescindere da una definizione formale e precisa del contesto operativo in cui esso viene applicato. Prima di poter discutere di aste, strategie di bidding o allocazione delle risorse, è necessario definire il "terreno di gioco": un framework matematico capace di descrivere le interazioni tra gli attori coinvolti, le loro motivazioni economiche e i vincoli fisici che ne limitano l'azione.

L'obiettivo di questo capitolo è quindi duplice:

- In primo luogo, si intende tradurre lo scenario operativo di una campagna di *Mobile Crowdsensing* (MCS) in un modello computabile, definendo le variabili di stato e le funzioni obiettivo che governano il sistema;
- In secondo luogo, si vuole colmare il divario tra i modelli teorici astratti — spesso basati su assunzioni semplificatorie — e la complessità di uno scenario reale, caratterizzato da eterogeneità spaziale e temporale.

Come riferimento teorico fondamentale, in questa tesi si adotta il modello **User-Centric** proposto in letteratura da Yang et al. [18]. A differenza dei modelli *platform-centric*, dove l'ente centrale impone prezzi e task, l'approccio user-centric delega agli utenti la libertà di dichiarare i propri costi e le proprie preferenze, introducendo una dinamica di mercato competitiva che richiede l'uso della Teoria dei Giochi per essere analizzata; tuttavia, l'adozione pura del modello teorico non è sufficiente per una validazione che abbia rilevanza pratica. Per evitare di lavorare su uno scenario puramente sintetico, il modello matematico è stato adattato e calibrato per rispecchiare una situazione reale di mobilità urbana. A tal fine, il framework integra i dati effettivi dei flussi di taxi nella città di Roma, estratti dal dataset CRAWDAD [2]. Questa scelta metodologica permette di ancorare la simulazione a vincoli topologici e pattern di mobilità realistici, rendendo i risultati dell'analisi economica direttamente applicabili a contesti urbani concreti.

Possiamo schematizzare il sistema risultante come un mercato elettronico (*e-market*) dinamico, in cui la domanda di informazioni geolocalizzate incontra l'offerta di risorse di sensing distribuite. Le entità che popolano e animano questo ecosistema sono tre:

1. **La Piattaforma (Crowdsourcer):** È l'ente centrale (ad esempio, una municipalità, un'agenzia per il monitoraggio ambientale o un centro di ricerca) che ha la necessità di raccogliere dati distribuiti sul territorio. Il suo ruolo è quello di *market maker*: definisce i task da svolgere, gestisce l'asta inversa per selezionare gli utenti più idonei e remunera i partecipanti, operando sempre sotto un vincolo di budget prefissato che ne limita la capacità di spesa.
2. **I Task ( $\mathcal{T}$ ):** Rappresentano l'unità elementare della domanda di dati. Non sono richieste generiche, ma obiettivi precisi ancorati allo spazio e al tempo (es. "misurare l'inquinamento acustico in Piazza Venezia tra le 08:00 e le 09:00"). A ogni task la piattaforma associa un valore economico ( $v$ ), che riflette l'utilità marginale o l'importanza strategica di quel dato specifico per il servizio finale.
3. **Gli Utenti ( $\mathcal{U}$ ):** Costituiscono la controparte operativa dell'offerta. Nel nostro caso di studio, questi corrispondono ai tassisti romani: agenti mobili che, muovendosi per le proprie attività lavorative primarie, possono offrire servizi di sensing in modalità opportunistica o partecipativa. Essi sono modellati come agenti razionali che sostengono costi operativi reali (carburante, tempo, usura del veicolo) e richiedono una compensazione adeguata per deviare dai loro percorsi o impegnare le proprie risorse.

## 3.2 L'Entità Task: Formalizzazione della Domanda

All'interno del sistema, il **task di sensing** costituisce l'unità elementare della domanda. Invece di utilizzare grafi sintetici o distribuzioni spaziali casuali, che spesso semplificano eccessivamente la realtà urbana, in questo lavoro ogni task è stato ancorato a coordinate geospaziali reali. Nello specifico, la posizione dei task è derivata direttamente dalla densità dei punti di interesse estratti dal dataset dei taxi di Roma, garantendo così che la domanda di dati sia coerente con le aree effettivamente frequentate e accessibili dalla flotta.

Possiamo formalizzare questa entità come segue:

**Definizione 3.1** (Task di Sensing). Ogni task  $\tau_j$  è identificato univocamente da una tupla:

$$\tau_j = \langle id_j, pos_j, v_j \rangle \quad (3.1)$$

dove:

- $id_j \in \mathbb{N}^+$  è l'identificativo numerico del task.
- $pos_j = (\phi_j, \lambda_j)$  indica la posizione geografica precisa (latitudine e longitudine) nel sistema di riferimento WGS84 (EPSG:4326), che corrisponde al punto in cui deve essere effettuata la rilevazione.
- $v_j \in \mathbb{R}_0^+$  rappresenta il **valore** (*valuation*) del task. Questa variabile è cruciale perché quantifica economicamente l'utilità marginale che la piattaforma attribuisce al completamento di quel particolare task.

Per quanto riguarda la determinazione del valore  $v_j$ , è necessario adottare un modello probabilistico che rifletta l'eterogeneità dell'importanza dei dati (alcuni dati valgono più di altri). In linea con la letteratura di riferimento, e per mantenere la comparabilità metodologica con il lavoro di Yang et al. [18], si è scelto di modellare il valore come una variabile casuale.

**Definizione 3.2** (Distribuzione del Valore dei Task). Il valore  $v_j$  di ciascun task è estratto da una distribuzione **uniforme** continua definita nell'intervallo  $[v_{\min}, v_{\max}]$ . La relativa funzione di densità di probabilità (PDF) è:

$$f(v) = \begin{cases} \frac{1}{v_{\max}-v_{\min}} & \text{se } v \in [v_{\min}, v_{\max}] \\ 0 & \text{altrimenti} \end{cases} \quad (3.2)$$

Un aspetto fondamentale di questa fase è stata la **calibrazione dei parametri economici**. Per la Fase 1, non si sono scelti valori arbitrari, ma si è proceduto a una stima empirica volta a garantire la sostenibilità del sistema. I parametri sono stati fissati come segue:

$$\begin{aligned} v_{\min} &= 1.8 \text{ €} \\ v_{\max} &= 15.0 \text{ €} \end{aligned} \tag{3.3}$$

La logica dietro questa scelta è duplice. Da un lato, si vuole evitare che i task abbiano valori irrisori che verrebbero sistematicamente ignorati; dall'altro, si vuole limitare l'esborso massimo per la piattaforma. Con questi estremi, il valore medio atteso di un task risulta essere:

$$\mathbb{E}[V] = \frac{1.8 \text{ €} + 15.0 \text{ €}}{2} = 8.4 \text{ €}$$

Confrontando questo valore con i costi operativi stimati per i tassisti (che, come vedremo, si aggirano intorno a una media di 5.75 € per servire un pacchetto tipico di 5 task entro 2 km, assumendo un costo chilometrico  $\kappa \approx 0.575 \text{ €}/\text{km}$ ), otteniamo un rapporto di profitabilità:

$$\frac{\mathbb{E}[V]}{\text{costo medio}} \approx 1.46$$

Questo margine di circa il 46% è stato introdotto intenzionalmente per assicurare che la partecipazione sia, in media, *economicamente vantaggiosa* per gli utenti. Senza questo incentivo strutturale, il vincolo di *Razionalità Individuale* rischierebbe di non essere soddisfatto per una larga parte della popolazione, compromettendo la validità della simulazione.

### 3.3 L'Entità Utente: L'Agente Razionale

Dopo aver definito la domanda (i task), passiamo ad analizzare l'offerta, ovvero gli utenti che eseguono le rilevazioni. In questo modello, gli utenti non sono semplici nodi passivi di una rete, ma vengono modellati come **agenti economici autonomi**.

Per questa prima fase sperimentale, assumiamo che tali agenti operino in regime di **razionalità perfetta**: ciò significa che sono in grado di calcolare esattamente i propri costi e agiscono sempre per massimizzare il proprio profitto (un'assunzione forte che verrà poi rilassata nelle fasi successive per introdurre maggior realismo comportamentale).

Nel contesto specifico del nostro caso di studio, questi agenti corrispondono ai tassisti operanti nell'area metropolitana di Roma, le cui tracce di mobilità sono state estratte dal dataset CRAWDAD [2]. Formalmente, possiamo definire ogni agente attraverso la seguente tupla:

**Definizione 3.3** (Utente/Agente). Un utente  $u_i$  è caratterizzato da quattro elementi fondamentali:

$$u_i = \langle id_i, pos_i, \kappa_i, \Gamma_i \rangle \quad (3.4)$$

dove:

- $id_i \in \mathbb{N}^+$  è l'identificativo univoco che ci permette di tracciare l'agente nel sistema.
- $pos_i \in \mathbb{R}^2$  rappresenta la posizione iniziale dell'utente (latitudine e longitudine proiettate), corrispondente all'ultima posizione nota registrata nel dataset prima dell'inizio dell'asta.
- $\kappa_i \in \mathbb{R}^+$  è il **costo operativo chilometrico** ( $\text{€}/\text{km}$ ). È un parametro privato fondamentale, poiché rappresenta quanto "costa" all'utente percorrere un chilometro aggiuntivo per servire la piattaforma.
- $\Gamma_i \subseteq \mathcal{T}$  è il **pacchetto di task** (*bundle*) ammissibili. Non rappresenta tutti i task del sistema, ma solo quel sottoinsieme che l'utente  $i$  è effettivamente in grado di raggiungere e considerare.

Un'attenzione particolare merita il parametro  $\kappa_i$ , che serve a catturare l'eterogeneità della flotta reale. Non tutti i veicoli hanno gli stessi costi: il consumo di carburante, l'usura, i costi di manutenzione e l'efficienza del motore variano da auto ad auto. Per rendere la simulazione verosimile, abbiamo analizzato i costi operativi del trasporto pubblico non di linea a Roma (con riferimento all'anno del dataset, il 2014). Sulla base di questa analisi, abbiamo modellato  $\kappa_i$  come una variabile aleatoria uniforme:

$$\kappa_i \sim \mathcal{U}[0.45, 0.70] \text{ €}/\text{km} \quad (3.5)$$

Questo intervallo non è casuale: il limite inferiore (0.45  $\text{€}/\text{km}$ ) riflette i costi di veicoli moderni e performanti (es. motorizzazioni diesel efficienti o ibride), mentre il limite superiore (0.70  $\text{€}/\text{km}$ ) rappresenta veicoli più datati o con consumi maggiori. In questo modo, la popolazione simulata riflette la varietà di un vero parco auto urbano.

*Nota 3.1.* È importante notare che l’insieme  $\Gamma_i$  (i task considerati) non è una caratteristica statica dell’utente, ma viene **determinato dinamicamente**. Un tassista a Roma Nord non considererà mai un task a Roma Sud a causa della distanza proibitiva. Pertanto,  $\Gamma_i$  viene costruito calcolando la prossimità geografica tra l’utente e i task disponibili: come vedremo nel dettaglio nella Sezione 5.7 (Capitolo 5), introduciamo un vincolo di **raggio di copertura**  $r$  che filtra automaticamente i task troppo lontani, rendendo il problema computazionalmente trattabile e realisticamente sensato.

## 3.4 Modello di Costo e Strategie di Bidding

Per descrivere il comportamento degli agenti è necessario specificare come essi stimano il costo per l’esecuzione di un insieme di task e come trasformano tale informazione in un’offerta (*bid*) da inviare alla piattaforma. In questa sezione vengono quindi introdotti il modello di costo, le ipotesi sulle strategie di bidding in Fase 1 e la corrispondente funzione di utilità.

### 3.4.1 Funzione di Costo Privato ( $c_i$ )

Si assume che ogni utente  $i$  conosca il proprio costo operativo e che il costo complessivo per completare un insieme di task  $\Gamma_i$  dipenda in modo lineare dalla distanza che deve percorrere per servirli.

**Definizione 3.4** (Funzione di Costo Privato). Il costo  $c_i$  sostenuto dall’utente  $u_i$  per servire l’insieme di task  $\Gamma_i$  è definito come:

$$c_i(\Gamma_i) = \kappa_i \cdot D_i(\Gamma_i) \quad (3.6)$$

dove  $D_i(\Gamma_i)$  è la distanza totale di servizio stimata.

Per stimare  $D_i(\Gamma_i)$  si adotta, in questa prima fase, un modello di routing volutamente semplice, indicato come **Star Routing**. L’idea è che l’utente parta dalla propria posizione base  $pos_i$  e raggiunga ogni task in  $\Gamma_i$  singolarmente, senza ottimizzare l’ordine di visita.

**Definizione 3.5** (Distanza di Servizio con Correzione Urbana). La distanza totale  $D_i(\Gamma_i)$  è calcolata come somma delle distanze geodetiche tra la posizione dell'utente e ciascun task, corrette da un fattore che tiene conto della tortuosità della rete stradale urbana:

$$D_i(\Gamma_i) = \eta_{\text{urban}} \cdot \sum_{\tau_j \in \Gamma_i} d_H(\text{pos}_i, \text{pos}_j) \quad (3.7)$$

dove:

- $d_H(\cdot, \cdot)$  è la **distanza geodetica di Haversine**, che approssima la distanza ortodromica tra due punti su una sfera (raggio della Terra  $R = 6,371$  km):

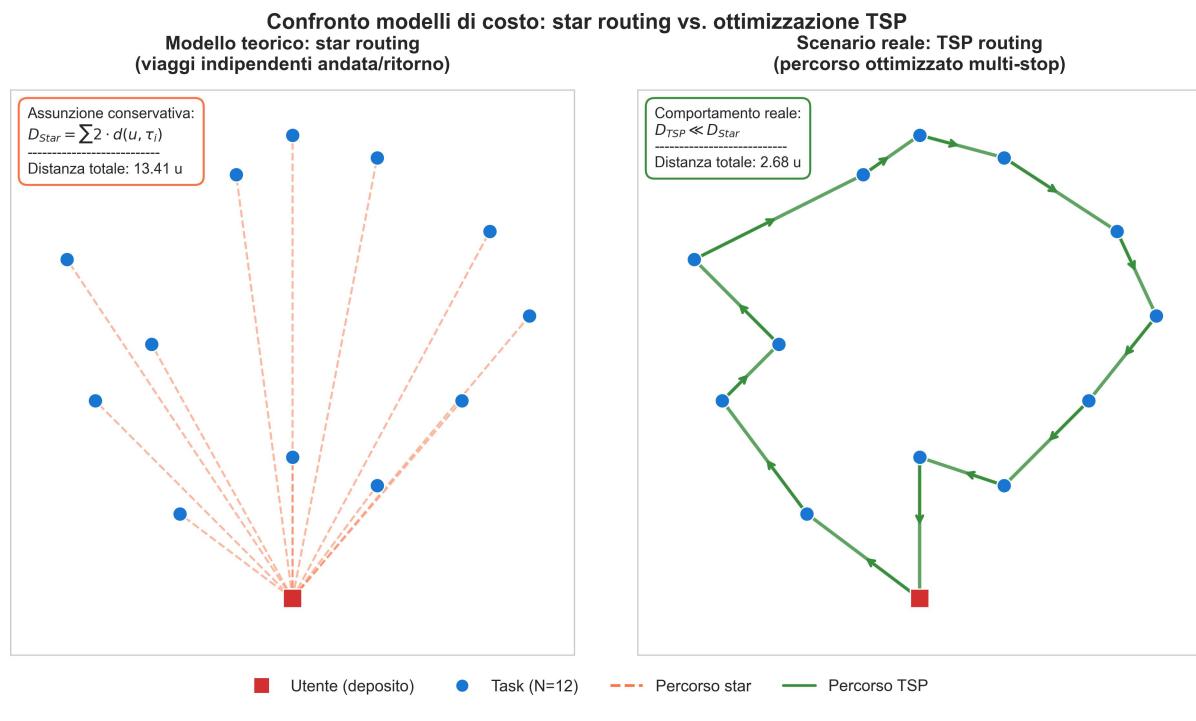
$$a = \sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_i)\cos(\phi_j)\sin^2\left(\frac{\Delta\lambda}{2}\right) \quad (3.8)$$

$$d_H = 2R \cdot \text{atan2}(\sqrt{a}, \sqrt{1-a}) \quad (3.9)$$

- $\eta_{\text{urban}}$  è il **fattore di correzione urbano**. È introdotto per tener conto del fatto che, in città, la distanza effettivamente percorsa lungo la rete stradale è in media superiore a quella geodetica a causa della presenza di incroci, sensi unici, vincoli di direzione e irregolarità del tessuto urbano. Nel caso di Roma, città con un centro storico particolarmente complesso, si è scelto un valore conservativo  $\eta_{\text{urban}} = 1.30$ , in linea con i range tipici (1.2–1.4) riportati per reti stradali europee con topologia irregolare [1].

La Figura 3.1 mostra come il modello star differisce da un routing ottimizzato di tipo TSP:

- **Modello Star (A):** l'utente parte dalla posizione base (deposito, quadrato rosso) per raggiungere ogni task (punti blu) individualmente, effettuando viaggi separati di andata e ritorno. Questo approccio sovrastima significativamente la distanza totale (13.41 u nell'esempio), rappresentando un limite superiore conservativo (*upper bound*) che preserva la proprietà di Individual Rationality anche in presenza di routing subottimale nella pratica.
- **Routing TSP (B):** l'utente ottimizza il percorso visitando tutti i task in un unico tour continuo, minimizzando la distanza totale (2.68 u nell'esempio). Questo rappresenta lo scenario reale in cui un conducente razionale pianifica un percorso multi-stop efficiente, ottenendo costi operativi significativamente inferiori rispetto alla stima conservativa del modello star.



**Figura 3.1:** Confronto tra il modello Star Routing adottato (A) e un routing ottimale TSP (B). Il modello star sovrasta il costo reale di un fattore superiore a 4 in questo scenario, garantendo robustezza nelle stime dei costi.

**Nota sulle limitazioni dello Star Routing.** Il modello di star routing è, per costruzione, una **semplificazione conservativa** rispetto a un routing ottimale basato sul problema del commesso viaggiatore (Traveling Salesman Problem, TSP). In letteratura è noto che, per piccoli insiemi di punti (ad esempio cluster di 3–8 nodi), uno schema di tipo "stella" può sovrastimare la lunghezza del tour ottimale di un fattore tipicamente compreso tra 2 e 4, a seconda della disposizione spaziale dei task. Questa scelta va quindi letta come una forma di cautela: si preferisce assumere costi leggermente più alti rispetto al minimo teorico, anziché rischiare di sottostimarli.

*Nota 3.2.* L'adozione dello star routing ha alcune conseguenze dirette sul comportamento del meccanismo:

1. **Bid più alti:** i costi stimati sono maggiori, quindi le offerte risultano più elevate e la selezione dei vincitori tende a essere più prudente (a parità di budget, il numero di vincitori è in genere inferiore rispetto a uno scenario con routing ottimale).
2. **Pagamenti critici non sottostimati:** i pagamenti calcolati a partire da questi costi rimangono su un ordine di grandezza tale da garantire la proprietà di *Individual Rationality* anche se, nella realtà, l'utente scegliesse un percorso non ottimale.
3. **Efficienza leggermente ridotta, robustezza aumentata:** si sacrifica una parte dell'efficienza allocativa potenziale per ottenere un comportamento più robusto e "fail-safe" dal punto di vista economico.

**Motivazione della scelta.** Nella Fase 1 si è adottato esclusivamente il modello di star routing per tutti gli utenti, indipendentemente dal loro livello di razionalità. Questa scelta è motivata dalla necessità di isolare le proprietà teoriche del meccanismo in uno scenario controllato: utilizzando un'unica metrica di costo conservativa, si garantisce che le bid e i pagamenti rispettino le proprietà di Individual Rationality e Incentive Compatibility, senza introdurre variabilità dovuta a eterogeneità comportamentale nel routing.

Nelle Fasi 2 e 3, il modello è stato esteso per incorporare una maggiore fedeltà comportamentale: il metodo di routing adottato da ciascun utente dipende dal suo livello di razionalità. Gli utenti altamente razionali tendono a ottimizzare i percorsi multi-stop (riducendo i costi operativi), mentre gli utenti con razionalità limitata adottano strategie meno efficienti, dal semplice star routing fino a percorsi sostanzialmente casuali. Questa eterogeneità consente di verificare come la bounded rationality influenzi non solo le decisioni di partecipazione all'asta, ma anche i costi effettivi sostenuti dagli utenti nella fase operativa.

### 3.4.2 Strategia di Bidding in Regime di Razionalità Perfetta

Nella Fase 1, lo studio si fonda sull'assunzione di **Razionalità Perfetta**. Gli agenti sono modellati come decisori razionali che conoscono esattamente la propria funzione di costo, comprendono perfettamente le regole del meccanismo d'asta IMCU e possiedono capacità computazionali sufficienti per calcolare la strategia che massimizza la loro utilità attesa [16].

L'obiettivo teorico del meccanismo è garantire l'incentivo alla veridicità. In Teoria dei Giochi Algoritmica, questa proprietà è formalizzata come segue:

**Definizione 3.6** (Truthfulness o Incentive Compatibility). Un meccanismo di incentivazione si definisce **Truthful** (o *Incentive-Compatible*) se, per ogni utente  $u_i$ , la strategia di dichiarare il proprio costo reale  $c_i$  massimizza l'utilità attesa, indipendentemente dalle strategie adottate dagli altri partecipanti  $\mathbf{b}_{-i}$ . In altre parole, la dichiarazione veritiera  $b_i = c_i$  costituisce una **strategia dominante**.

Per dimostrare che l'algoritmo IMCU soddisfa questa proprietà, facciamo riferimento alla celebre caratterizzazione fornita da Myerson per i meccanismi a parametro singolo.

**Teorema 3.1** (Caratterizzazione di Myerson [12]). *Un meccanismo d'asta a offerta inversa è truthful se e solo se soddisfa le seguenti due condizioni necessarie e sufficienti:*

1. **Monotonicità della regola di selezione:** *Se un utente  $u_i$  viene selezionato come vincitore sottomettendo un'offerta  $b_i$ , deve continuare a essere selezionato sottomettendo una qualsiasi offerta  $b'_i < b_i$  (a parità di altre condizioni).*
2. **Pagamento Critico:** *Il pagamento corrisposto a ciascun vincitore è indipendente dalla sua offerta ed è pari al valore critico, ovvero l'estremo superiore delle offerte che l'utente avrebbe potuto sottomettere pur continuando a vincere.*

Sulla base di questo fondamento teorico, possiamo enunciare e dimostrare il teorema fondamentale relativo alla validità del meccanismo adottato.

**Teorema 3.2** (Veridicità del Meccanismo IMCU – da Yang et al. [18]). *Il meccanismo d'asta IMCU è **truthful**. Pertanto, per qualsiasi utente  $i$  con costo privato  $c_i(\Gamma_i)$ , la strategia  $b_i = c_i(\Gamma_i)$  è una strategia dominante.*

*Dimostrazione.* Sfruttando il Teorema 3.1, la dimostrazione si riduce alla verifica delle due condizioni di Myerson all'interno dell'algoritmo IMCU.

- **Analisi della Monotonicità.** La fase di selezione dei vincitori nell'algoritmo IMCU segue un approccio *greedy* basato sul valore marginale. Gli utenti vengono ordinati e selezionati iterativamente in base al contributo marginale netto  $v_i(\mathcal{S}) - b_i$ . Supponiamo che l'utente  $i$  vinca con un'offerta  $b_i$ . Se l'utente riduce la sua offerta a  $b'_i < b_i$ , la differenza  $v_i(\mathcal{S}) - b'_i$  aumenta. Questo miglioramento (o invarianza) nel ranking garantisce che l'utente venga considerato dall'algoritmo greedy in una fase precedente o uguale rispetto al caso con  $b_i$ , assicurando la sua inclusione nell'insieme dei vincitori  $S^*$ . La regola di selezione è dunque monotona [18].
- **Analisi del Pagamento Critico.** La fase di determinazione dei pagamenti dell'IMCU è esplicitamente costruita per calcolare la soglia critica. Il pagamento  $p_i$  non è funzione diretta del bid  $b_i$  del vincitore, ma è determinato dalle offerte dei concorrenti esclusi ("i perdenti") e dai valori marginali del sistema. Formalmente, il pagamento è calcolato come:

$$p_i = \max\{b'_i : i \in S^*(b'_i, \mathbf{b}_{-i})\} \quad (3.10)$$

- **Conclusione.** Poiché entrambe le condizioni di Myerson sono soddisfatte per costruzione, qualsiasi deviazione dalla strategia veritiera (sia essa *overbidding* o *underbidding*) porta a un'utilità attesa debolmente inferiore o nulla. Di conseguenza,  $b_i = c_i$  è una **strategia debolmente dominante** e, in assenza di casi patologici legati al tie-breaking, risulta essere una strategia dominante stretta [9].

□

**Corollario 3.2.1** (Comportamento di Bidding Assunto in Fase 1). *Data l'assunzione di razionalità perfetta e il Teorema 3.2, nella Fase 1 si postula che ogni utente  $u_i$  sottometta alla piattaforma un'offerta esattamente pari al proprio costo privato:*

$$b_i(\Gamma_i) = c_i(\Gamma_i) \quad (3.11)$$

*Dimostrazione.* Per ipotesi, un agente perfettamente razionale:

1. conosce la propria funzione di utilità  $u_i(b_i, \mathbf{b}_{-i})$ ;
2. conosce le regole del meccanismo e la proprietà di veridicità dimostrata nel Teorema 3.2;
3. possiede capacità di calcolo illimitata per individuare la strategia ottima.

Poiché  $b_i = c_i$  è una strategia dominante, essa rappresenta la *best response* dell'agente indipendentemente dalle azioni altrui. Un agente razionale non ha alcun incentivo a deviare da tale strategia.

Formalmente, indicando con  $U_i(b_i | \mathbf{b}_{-i})$  l'utilità attesa dell'utente  $i$ , dal teorema precedente discende che:

$$U_i(c_i | \mathbf{b}_{-i}) \geq U_i(b'_i | \mathbf{b}_{-i}) \quad \forall b'_i \neq c_i, \forall \mathbf{b}_{-i} \quad (3.12)$$

Questa uguaglianza costituisce la **Ground Truth** comportamentale della Fase 1 (come discusso nel dettaglio nel Capitolo 5): la simulazione non serve a "verificare" empiricamente che gli utenti *scelgano* di essere veritieri (proprietà già garantita a priori dalla teoria), ma piuttosto a simulare numericamente l'**Equilibrio di Nash** del gioco, assumendo che l'intera popolazione di agenti abbia già convergito verso il comportamento ottimale.  $\square$

### 3.4.3 Funzione di Utilità dell'Agente

Definito il modello di costo e la strategia di bidding, è necessario definire come quantificare il *guadagno netto* che un agente ottiene dalla partecipazione all'asta IMCU.

**Definizione 3.7** (Utilità dell'Utente). L'utilità  $u_i$  dell'utente  $u_i$  è modellata attraverso una funzione quasi-lineare standard:

$$u_i = \begin{cases} p_i - c_i(\Gamma_i) & \text{se } u_i \in S^* \text{ (vincitore)} \\ 0 & \text{se } u_i \notin S^* \text{ (non vincitore)} \end{cases} \quad (3.13)$$

dove  $p_i$  è il pagamento trasferito dalla piattaforma e  $S^*$  è l'insieme dei vincitori selezionati.

**Assunzione 3.1** (Utilità di Riserva Nulla). Si assume che l'utilità di riserva di ogni utente, ossia il payoff associato all'opzione di non partecipare all'asta, sia normalizzata a zero. Questa ipotesi è comune nei meccanismi di crowdsensing [18] e riflette il fatto che gli utenti non sostengono costi operativi se non vengono ingaggiati.

Affinché il meccanismo sia sostenibile in un contesto realistico, è necessario garantire la proprietà di **Razionalità Individuale** (Individual Rationality, IR): nessun utente razionale deve subire una perdita (utilità negativa) partecipando al meccanismo.

**Proposizione 3.1** (Razionalità Individuale del Meccanismo IMCU). *Un meccanismo d'asta è individualmente razionale (IR) se l'utilità ex-post di ogni partecipante è non negativa, ovvero  $u_i \geq 0$  per ogni utente  $i$ . Il meccanismo IMCU soddisfa questa proprietà per costruzione.*

*Dimostrazione.* È necessario verificare la condizione per i due possibili stati dell'utente: vincitore o non vincitore.

- **Caso 1: Utente non selezionato ( $i \notin S^*$ ).** Dalla definizione della funzione di utilità, se l'utente non vince, non sostiene costi e non riceve pagamenti, quindi:

$$u_i = 0 \geq 0 \quad (3.14)$$

- **Caso 2: Utente selezionato** ( $i \in S^*$ ). Se l'utente risulta vincitore, la sua utilità è data da  $u_i = p_i - c_i$ . Dalla proprietà del *pagamento critico* dimostrata nel Teorema 3.2 (in virtù del Teorema di Myerson), sappiamo che il pagamento ricevuto è sempre maggiore o uguale al bid dichiarato per vincere:

$$p_i \geq b_i \quad (3.15)$$

Nella Fase 1, in regime di razionalità perfetta, dal Corollario 3.2.1 sappiamo che l'utente dichiara il vero costo ( $b_i = c_i$ ). Sostituendo questa uguaglianza:

$$p_i \geq c_i \implies p_i - c_i \geq 0 \quad (3.16)$$

Pertanto,  $u_i \geq 0$  anche per ogni vincitore.

- **Conclusione.** In tutti gli scenari possibili, l'utilità dell'agente è non negativa. Il meccanismo IMCU rispetta rigorosamente la proprietà di **Individual Rationality**, assicurando che la partecipazione non comporti mai un rischio economico per l'utente [18, 9].

□

**Corollario 3.2.2** (Condizione di Partecipazione Volontaria). *Data la proprietà di Individual Rationality (Proposizione 3.1) e l'ipotesi di utilità di riserva nulla (Assunzione 3.1), ogni utente razionale ha un incentivo debole (o nullo nel caso limite) a partecipare attivamente all'asta IMCU.*

*Dimostrazione.* Siano  $U_i^{\text{partecipa}}$  l'utilità attesa derivante dalla partecipazione e  $U_i^{\text{non partecipa}} = 0$  l'utilità derivante dall'astensione. Dalla Proposizione 3.1, abbiamo dimostrato che il meccanismo garantisce sempre un risultato non negativo:

$$U_i^{\text{partecipa}} \geq 0 \quad (3.17)$$

Confrontando le due opzioni, appare evidente che:

$$U_i^{\text{partecipa}} \geq U_i^{\text{non partecipa}} \quad (3.18)$$

In termini pratici, ciò significa che per un agente razionale la partecipazione è un'operazione a **rischio zero**. Non esistono disincentivi economici:

- Nel caso peggiore (l'utente perde l'asta), il suo bilancio rimane invariato ( $u_i = 0$ ), esattamente come se avesse deciso di ignorare il sistema;
- Nel caso favorevole (l'utente vince), ottiene un profitto netto garantito ( $p_i > c_i$ ).

□

# Elenco delle figure

1.1 Architettura concettuale di un sistema di Mobile Crowdsensing. La piattaforma centrale coordina tre attori principali: (i) pubblica task geolocalizzati agli utenti mobili, (ii) raccoglie dati dai sensori dei dispositivi, (iii) aggrega le informazioni e fornisce servizi ai consumatori finali (amministrazioni pubbliche, enti di ricerca, aziende) [6]. . . . .	4
2.1 Schema delle principali architetture di raccolta dati mobili. La figura mostra la tassonomia fra WSN, paradigmi peer-to-peer e Mobile Crowdsensing, evidenziando per ciascuno le caratteristiche chiave di deployment e di funzionamento, in linea con le classificazioni proposte nelle principali survey di letteratura [10, 6, 3]. . . . .	15
2.2 Architettura logica a tre livelli per sistemi MCS evoluti. . . . .	18
2.3 Curva di Lorenz e coefficiente di Gini. La linea tratteggiata a 45° rappresenta la perfetta equità, mentre la curva continua descrive la distribuzione reale dei compensi. Il coefficiente di Gini è proporzionale all'area compresa tra le due curve: valori elevati indicano una forte concentrazione delle ricompense nelle mani di pochi utenti. . . . .	30
2.4 Struttura decisionale Fast-and-Frugal Tree (FFT) a cascata applicata al MCS. Il processo inizia dal nodo radice in alto a sinistra e procede diagonalmente. A ogni bivio, il mancato soddisfacimento di un criterio critico (distanza, ricompensa) comporta l'immediata uscita dal sistema (rifiuto), senza che l'utente debba elaborare tutte le informazioni disponibili. . . . .	34

2.5 Visualizzazione dei rendimenti marginali decrescenti. L'area colorata rappresenta l'utilità totale $v(S)$ accumulata. Le barre verticali indicano il valore aggiunto da un nuovo utente in fasi diverse: il contributo $\Delta_1$ (fase iniziale) è nettamente superiore a $\Delta_2$ (fase di saturazione). Questo andamento riflette la naturale ridondanza dei dati spaziali nel crowdsensing. . . . .	37
3.1 Confronto tra il modello Star Routing e TSP . . . . .	47

# Bibliografia

- [1] M. Barthelemy. «Spatial networks». In: *Physics Reports* 499.1-3 (2011), pp. 1–101.
- [2] L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici e A. Rabuffi. *CRAWDAD romataxi*. IEEE Dataport. Dic. 2022. URL: <https://ieee-dataport.org/open-access/crawdad-roma-taxi>.
- [3] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Klazovich e P. Bouvry. «A Survey on Mobile Crowdsensing Systems: Challenges, Solutions, and Opportunities». In: *IEEE Communications Surveys & Tutorials* 21.3 (2019), pp. 2419–2465.
- [4] D. Christin, A. Reinhardt, S. S. Kanhere e M. Hollick. «A survey on privacy in mobile participatory sensing applications». In: *Journal of Systems and Software* 84.11 (2011), pp. 1928–1946.
- [5] D. Fudenberg e J. Tirole. *Game Theory*. MIT press, 1991.
- [6] R. K. Ganti, F. Ye e H. Lei. «Mobile crowdsensing: current state and future challenges». In: *IEEE Communications Magazine* 49.11 (2011), pp. 32–39.
- [7] G. Gigerenzer e D. G. Goldstein. «Reasoning the fast and frugal way: models of bounded rationality». In: *Psychological review* 103.4 (1996), p. 650.
- [8] L. G. Jaimes, I. J. Vergara-Laurens e A. Raij. «A Survey of Incentive Techniques for Mobile Crowd Sensing». In: *IEEE Internet of Things Journal* 2.5 (2015), pp. 370–380.
- [9] V. Krishna. *Auction theory*. Academic press, 2009.
- [10] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury e A. T. Campbell. «A survey of mobile phone sensing». In: *IEEE Communications Magazine* 48.9 (2010), pp. 140–150.

- [11] N. Maisonneuve, M. Stevens, M. E. Niessen e L. Steels. «NoiseTube: Measuring and mapping noise pollution with mobile phones». In: *Information Technologies in Environmental Engineering*. Springer, 2009, pp. 215–228.
- [12] R. B. Myerson. «Optimal auction design». In: *Mathematics of operations research* 6.1 (1981), pp. 58–73.
- [13] G. L. Nemhauser, L. A. Wolsey e M. L. Fisher. «An analysis of approximations for maximizing submodular set functions—I». In: *Mathematical programming* 14.1 (1978), pp. 265–294.
- [14] J. Nie, J. Luo, Z. Xiong, D. Niyato e P. Wang. «A Stackelberg Game Approach Toward Socially-Aware Incentive Mechanisms for Mobile Crowdsensing». In: *IEEE Transactions on Wireless Communications* 18.1 (2019), pp. 724–738.
- [15] F. Restuccia, N. Ghosh, S. Bhattacharjee, S. K. Das e T. Melodia. «Quality of Information in Mobile Crowdsensing: Survey and Research Challenges». In: *ACM Transactions on Sensor Networks (TOSN)* 13.4 (2017), pp. 1–43.
- [16] H. A. Simon. «A Behavioral Model of Rational Choice». In: *The Quarterly Journal of Economics* 69.1 (1955), pp. 99–118.
- [17] J. White, C. Thompson, H. Turner, B. Dougherty e D. C. Schmidt. «Waze: a community-based traffic and navigation app». In: *Proceedings of the International Conference on Mobile Computing, Applications, and Services*. 2010.
- [18] D. Yang, G. Xue, X. Fang e J. Tang. «Incentive Mechanisms for Crowdsensing: Crowd-sourcing With Smartphones». In: *IEEE/ACM Transactions on Networking* 24.3 (2016). Published online in 2015, issue date 2016, pp. 1732–1744.
- [19] X. Zhang, Z. Yang, W. Sun, Y. Liu, S. Tang, K. Xing e X. Mao. «Incentives for mobile crowd sensing: A survey». In: *IEEE Communications Surveys & Tutorials* 18.1 (2015), pp. 54–67.