A Probabilistic Model for the Deployment of Human-Enabled Edge Computing in Massive Sensing Scenarios

Published Version:
A Probabilistic Model for the Deployment of Human-Enabled Edge Computing in Massive Sensing Scenarios / Belli D.; Chessa S.; Foschini L.; Girolami M.. - In: IEEE INTERNET OF THINGS JOURNAL. - ISSN 2327-4662. - ELETTRONICO. - 7:3(2020), pp. 8924663.2421-8924663.2431. [10.1109/JIOT.2019.2957835]

Availability:
This version is available at: https://hdl.handle.net/11585/764221 since: 2020-07-04

Published:
DOI: http://doi.org/10.1109/JIOT.2019.2957835

A Probabilistic Model for the Deployment of Human-enabled Edge Computing in Massive Sensing Scenarios

Dimitri Belli[1] and Stefano Chessa[1,2]
[1]Department of Computer Science
University of Pisa, Italy
[2]CNR-ISTI, Pisa, Italy
E-mail: dimitri.belli@di.unipi.it, stefano.chessa@unipi.it

Luca Foschini[3] and Michele Girolami[4]
[3]Dipartimento di Informatica: Scienza e Ingegneria
University of Bologna, Italy
[4]CNR-ISTI, Pisa, Italy
E-mail: luca.foschini@unibo.it, michele.girolami@isti.cnr.it

***Abstract*— Human-enabled Edge Computing (HEC) is a recent smart city technology designed to combine the advantages of massive Mobile CrowdSensing (MCS) techniques with the potential of Multi-access Edge Computing (MEC). In this context, the architectural hierarchy of the network shifts the management of sensing information close to terminal nodes through the use of intermediate entities (edges) bridging the direct Cloud-Device communication channel. Recent proposals suggest the implementation of those edges, not only employing fixed MEC nodes, but also opportunistically using as edge nodes mobile devices selected among the terminal ones. However, inappropriate selection techniques may lead to an overestimation or an underestimation of the number of nodes to be used in such a layer. In this work, we propose a probabilistic model for the estimation of the number of mobile nodes to be selected as substitutes of fixed ones. The effectiveness of our model is verified with tests performed on real-world mobility traces.**

Keywords—Mobile CrowdSensing, Multi-access Edge Computing, Human-enabled Edge Computing, Social Mobility

I. Introduction

Multi-access Edge Computing (MEC) is a 5G key enabling technology introduced by European Telecommunications Standards Institute (ETSI) as an evolution of the Cloud Computing (CC) paradigm. MEC is an architecture that interposes between the direct cloud-device communication channel a middleware layer made up of edge nodes with data aggregation, storage, processing, and analytics capabilities. MEC edges bring the computation closer to the network terminal nodes as mobile and wearables; that enables low-latency, high-bandwidth, and real-time access to network analytics. For this reason, they are synergic with Internet of Things (IoT) applications since they allow an efficient partitioning of the functions. That saves IoT devices from performing complex functions that may affect their performance or energy efficiency [1]. We are interested in the use of MEC technologies in a massive sensing scenario, where many sensors produce data that are collected in the cloud for their subsequent processing with data analytics engines. A notable example of such scenario is Mobile CrowdSensing (MCS), that exploits sensors embedded in the personal smartphones of the MCS users to collect large masses of data covering wide regions.

In a typical MEC architecture supporting MCS, a limited number of edges, called in this paper Fixed MEC nodes (FMECs), are placed at fixed, strategic places, where they can easily intercept most of the data [2]. In this work, however, we consider a possible extension of this architecture, called Human-driven Edge Computing (HEC) where the edges can also be mobile and operated by the personal devices of the users themselves, called Mobile MEC nodes (M^2ECs) [3]. In particular, M^2ECs have the same mobility of their users and can collect data opportunistically from other devices by using short-range communication interfaces (like Bluetooth of direct Wi-Fi interfaces). The advantages of this approach are manifold: it reduces the costs for the maintenance of the infrastructure; it can be easily dynamically reconfigured; and the users may use free-cost short-range communications. The main drawback is in terms of higher data collection latencies, but they are typically acceptable for off-line and long-term data processing.

In this context, our work concerns the algorithms for the identification of the mobile devices that can act as M^2EC. Since the opportunity to collect data of the M^2EC depends on the frequency with which devices meet, we consider a selection strategy that operates according to the social relationships among the users of the MCS platform [4, 5], and we provide a probabilistic model and a close expression that can be used to estimate the expected contribution that a M^2EC can give to the data collection capability of the MCS platform. To the best of our knowledge, in the literature there are no theoretical models able to estimate such contribution. In particular, this paper presents a probabilistic model that, with a closed-form expression, can determine the number of M^2ECs to be selected in the plethora of MCS devices. The fine tuning of the model includes few parameters that can be easily distilled from the history of the users' behaviour considering the centrality of the potential M^2ECs within their respective communities. What makes our model unique is the possibility of estimating the contribution of the umpteenth M^2EC to be used in support of or in place of MEC middleware proxies. To the best of our knowledge in literature there are no theoretical models to estimate such contribution. The proposed probabilistic model has been validated through tests performed over the ParticipAct living lab dataset, a real-world MCS experiment realized with the contribution of approximately 170 university students of the Emilia Romagna region (Italy). The evidence demonstrates the effectiveness of our model in estimating the number of M^{2}ECs able to guarantee the same performances of a deployment realized with FMECs as in a standard MEC deployment.

The rest of the article is structured as follows. Section II presents an overview of the main studies in the areas of MCS, MEC, and HEC technologies. Section III introduces our MEC-based MCS architecture and the M^{2}EC's selection strategy. Section IV formalises the M^{2}EC selection method and the adopted probabilistic model. Section V describes the dataset adopted in the trial and the metrics used to validate our model. Section VI shows the experimental results obtained when the theoretical model is compared with the real-world one. The paper ends with Section VII in which are drawn conclusions and possible future developments of our research.

II. Related Work

Without claiming exhaustiveness, we survey in this section the MEC paradigm and the most recent massive application scenarios. More precisely, in Section II-A we review the main characteristics and limitations of the MEC paradigm. Section II-B describes the relationship between MEC and Mobile CrowdSensing (MCS), with particular attention to efficient device selection strategies for increasing the amount of data that can be gathered.

A. *The MEC paradigm*

MEC provides a new ecosystem based on radio access network edges with computational and storage capabilities. The MEC model aims at supporting peripheral nodes of the network by reducing latency for mobile users, optimizing mobile backhaul and middleware layer nodes performance. Such decentralized cloud technology makes the MEC one of the cornerstones of the new 5G systems [6], easing the convergence between telecommunication and information technology services [7-8]. Traditional MEC implementations are made up of software platforms which provide local services without considering the user mobility. Differently, advanced MEC implementations also consider aspects like traffic [9, 10], mobility [11] and account, introducing a heterogeneous type of networking to support both commercial and non-commercial applications, both in indoor and outdoor [12, 13, 14, 15]. Concerning the use of MEC with the Internet of Things (IoT) applications, recent studies focused on the development of platforms that ensure management and interoperability between large-scale devices without loss of bandwidth or increased latency [16]. With a plethora of devices interacting with intermediate MEC nodes, computation offloading techniques became valuable methods to save energy, battery lifetime and calculations. As the production of computation offloading is vast, we limit to mention a comprehensive study on these techniques over 5G heterogeneous networks [17]. Other major use cases of the MEC paradigm include optimization techniques for distributed content discovery and delivery [18, 19] and caching [20], big data storage and computation [21] and, notably, services for smart cities as localization [22], emergency and public safety [23, 24, 25]. It worth to notice that in MEC scenarios the problem of resource constraints cannot be avoided. In particular, the battery drain is the main concern for mobile users. In recent years, however, user recruitment policies leveraging profiles and current battery level of devices for task assignments have been devised, allowing a reasonable energy saving for those devices with limited resources [26, 27]. Alongside cost-saving energy solutions, there is also bandwidth wastage. Some propositions based on local data mining by mobile devices, with production of intermediate results, refined, from raw sensing data, have shown a real energy and bandwidth saving in data transmission to remote servers [28].

B. *Device Selection Strategies*

We are interested in combining the MEC paradigm with the Mobile CrowdSensing (MCS) [29], an approach for massive sensing in mobile scenarios. Nowadays, almost all mobile and wearable devices are equipped with multi-modal sensing capabilities which are able to collect data in a seamless way. The idea of MCS is to exploit the ubiquity of such devices in order to increase the amount and quality of data that can be gathered from the crowd [30]. Examples of data that can be collected (also referred to as tasks) include: media contents, recordings of audio tracks, sensor readings such as noise/light intensity etc. Strategies for task assignment include the study of recruitment areas considering the population density at different time of the day and the user's activity as the willingness in performing tasks and the time spent in city hotspots [31]. Under this respect, several studies also address privacy issue while collecting personal information [32, 33]. The MCS paradigm offers the possibility of gathering data from the crowd, while the MEC paradigm provides a scalable and dynamic architecture enabling such a massive collection of information. Recently, in the context of smart cities [34], hybrid solutions of MEC and MCS include models for optimizing content sharing by leveraging the sociality and mobility of the MCS nodes [4], MEC-based architectures for massive scale MCS services and privacy preserving citizens' data control flow [35, 36], and social-driven edge computing architectures based on incentives and centrality measures to reduce installation and maintenance costs for MEC intermediate layer nodes [3].

In this context, one of the main challenges is the selection of those devices acting as edge of the cloud, namely those devices collecting data from others and uploading data to the backend [37, 38, 39]. An efficient selection of the devices can result with high spatial and temporal coverage (as discussed in [40, 41]). Spatial coverage considers the number of areas of a region that can be monitored with a certain accuracy, such as the districts of a city. While temporal coverage refers to the time required in order to collect data from all the regions. In [42] authors propose a selection strategy based on the dimension of the region to cover. Authors consider a deterministic node mobility and propose several heuristics for the selection of devices, in order to minimize an objective function. Other recent works [43-45] consider incentive mechanisms for the users in order to increase the probability of collecting data from specific locations. In [46] authors adopt a selection strategy based on the quality of data the user can produce. The authors adopt a Compressive Sensing technique in order to predict the expected quality of data from the user and, in turn, they select appropriately the best candidates. Authors of [47] also adopt a method to select devices based on the quality of data they provide. Authors adopt a trust model called experience-reputation (ER) which evaluated the "trust relationships" between any two mobile device users. Authors test the ER model on several MCS testbed in which regular and malicious users are present.

A different approach for the selection of devices is based on the analysis of sociality of users. Since devices are mainly carried by humans, it is possible to adopt strategies based on the analysis of features of the human sociality and mobility. The main idea is to measure how much a user encounters or clusters with other users and, in turn, to select those with maximizing the contact probability. In fact, the more a user interacts with others, the more likely its device can collect data from neighbours through short range network interfaces (such as Bluetooth, Wi-Fi Direct or the upcoming LTE direct). The contact probability can be measured with metrics typical of graph theory such as clustering coefficient, betweenness or eigenvector centrality that all measure the centrality of a node in a graph. The device selection strategy presented in this work is also based on the analysis of sociality among users (see Section III-A). Our approach consists in first detecting the communities and then in selecting a representative for each of them. The selection of the representative is based by measuring a centrality score.

III. Extending A Mobile CrowdSensing Architecture with Mobile Edges

A. System model for Mobile CrowdSensing

In our system model, the mobile CrowdSensing architecture comprises three tiers: a client-tier, an edge-tier, and a server cloud-tier. The client-tier includes mobile terminal nodes. The edge-tier hosts MEC edges, namely, M²EC operating as conventional edges to serve other terminal devices. The server cloud-tier deals with aggregation, storage, remote analytics, and processing of sensed data. We assume that the CrowdSensing platform is administered by an entity (typically allocated in the cloud tier) that coordinates the clients by injecting the sensing tasks and that implements the storage and analytics on the received data. Differently than conventional architectures, in which the MEC are fixed and part of the infrastructure, in our model we assume that also mobile devices in the client tier can take the role of M²EC and thus act as data aggregation points for other clients, on the base of localized data exchange based on short range wireless interfaces (like Bluetooth, Wi-Fi direct or LTE direct). To this purpose, the administration entity of the MCS platform periodically selects, with a suitable algorithm, the client devices more suitable to act as M²EC.

Algorithm 1 – an abstract M²EC selection algorithm

Let T'' be the current time and $[T', T'']$ be the past period
Let $G(t), \forall t \in [T', T'']$ be the dynamic graph representing the social network and let C be the set of communities $\{C_1, \dots, C_n\}$ identified in $G(t)$
w.l.g. let $|C_i| \geq |C_j| \iff i \leq j$ (i.e. the set of communities is sorted in decreasing order according to their size)
Let $k \in [1, n]$ such that:
- $|\bigcup_{i \leq k} C_i| / \eta \geq \phi$
- $|\bigcup_{i \leq k+1} C_i \setminus \bigcup_{i \leq k} C_i| \leq \delta$;

For all $C_i \in C$ with $i \leq k$ do
  select $u_j \in C_i$ according to some centrality measure to act as M²EC

Considering that a M²EC is a mobile device that uses its short-range communication interface(s) to collect data from other CrowdSensing devices, follows that each M²EC guarantees a limited spatial coverage that varies continuously over time according to its mobility. For this reason, it behaves opportunistically and receives collected data from the other devices that come close enough to it for a sufficient period of time. The M²EC can then aggregate all collected data and transfer them (even later) to the cloud servers, possibly using broadband links. On the other hand, a device that did not have the opportunity to upload its collected data to a M²EC may, after a time, use itself broadband links to transfer the data to the cloud servers, even if these links are usually more expensive in terms of battery and costs.

Since the opportunities of communication between the devices and the M²EC depend on the mobility of the devices (and thus of the mobility and sociality of the user that carry them), several recent works already suggested to select M²EC on the basis of "social" relationship among the devices [4, 36]. The idea here is to consider the dynamic network induced by the contacts among devices as a social network (which, in fact, represents the physical social network of their users) where ties between nodes model the contacts (and their quality like frequency, duration, inter-contact time etc.) between devices, and to identify in this network the devices that are better connected (that have frequent encounters) with a larger number of other devices. Following this idea, these algorithms operate by identifying the communities to which the devices belong and by choosing in these communities the devices that are central to them according to some metric (for example betweenness centrality or others). Being well connected to the respective communities the devices selected as M²EC are then expected to serve well as "hubs" for their communities.

B. Selection Strategy for M²EC

In this work we are not interested in assessing any specific algorithm for the selection of M²EC, but rather to present a framework and a thorough assessment the strengths and the limits of our HEC approach. For this reason, we consider an abstract M²EC selection algorithms that well represents those approaches. The algorithm operates at discrete intervals (periods). When it is executed, it analyses the social network resulting from the devices in the previous period and it identifies the devices that will act as M²EC for the next period. For the sake of simplicity, we assume that the set of nodes in a period remains stable (this is reasonable if the periods are short enough, e.g. one day or so, anyway we will remove this assumption in the simulations). However, the links between devices in the network will vary over time. Without loss of generality, let us consider the period $[T', T'']$ in which is active a set of N devices of cardinality η. In general, at a given time $t \in [T', T'']$ a device $u \in N$ is connected to a set N_u^t (possibly

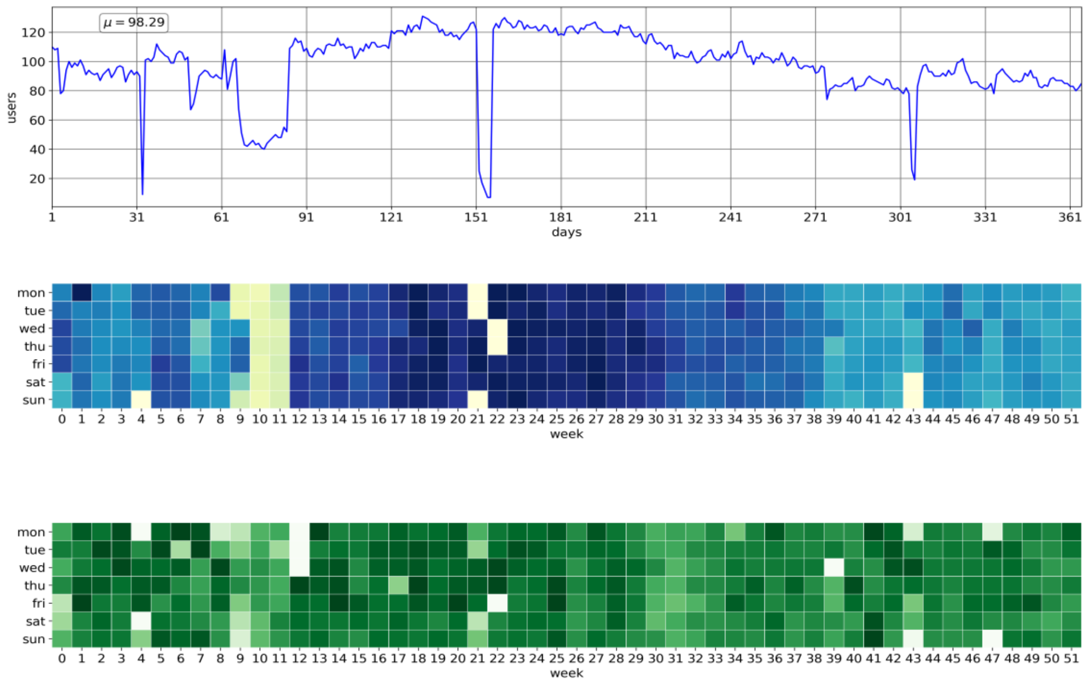

Fig. 1. Number of users and amount of traces available in 2014.

empty) of other devices, which are within the transmission range of u. This transmission range depends on the short-range radio interface used for the communications between devices, can be around 100 meters if Wi-Fi is used, for example. Hence the "social" network of the devices in this period can be represented by the dynamic graph $G(t) = (N, E(t))$ where the set of edges $E(t)$ is time varying and $(u, v) \in E(t) \iff u, v \in N \wedge v \in N_u^t \wedge u \in N_v^t$. At time T'' (at the end of the period) the system runs the algorithm for the selection of M²EC that will be active in the next period. To this purpose, the algorithm identifies first the set of communities $C_1, \dots, C_n$ in $G(t)$ (w.l.g. the communities are sorted in decreasing order of size) and then selects a representative for each of these communities starting from the largest one. The rationale for this behavior is that it is more convenient to select a node well connected in a large community because it will presumably behave as hub for a larger number of devices, thus requiring a smaller number of nodes selected as M²EC. However, not all representatives of all communities will be selected as M²EC. The reason is twofold:

- Some communities may be so small to make the M²EC selection poorly effective.
- Communities generally are not disjoint, hence there may be the case in which a community is already covered by the representatives of another one.

We model this by using two parameters, that is, ϕ which represents the fraction of nodes that are covered (i.e. that belong to a community for which we select a M²EC), and δ that represents the individual contribution given by the smallest community for which a M²EC is selected to the union of the communities selected. The abstract algorithm the selection of M²ECs is summarized in Algorithm 1 (as presented in [4]).

IV. The Analytical Model

M²ECs are devices acting as edges of the MEC architecture that can be selected with the algorithm described in Section III-B. However, selecting the number of M²EC for a HEC architecture is not easy task. In relation to this, we want to answer to the following question: given k M²EC already selected, which is contribution on each of them in terms of number of new nodes that each M²EC can connect with?

To this end, we describe in this section a probabilistic model which estimates the contribution of a M²EC. More specifically, given k different M²EC selected, we measure the number of unique nodes connected only to the k–*th* M²EC with respect to the previous $k-1$ M²ECs. We assume that the devices forms n communities denoted with $C_i, i \in [1 \dots n]$ and each community can contribute with at most one single M²EC. In other words, we assume that in each community we can assign one and only one M²EC as representative. The cardinality of the population is given by $\eta = |C_1 \cup \dots \cup C_n|$.

For the sake of simplicity, we assume that $|C_i| = g, \forall\, i \in [1, n]$, and for any pair of communities C_i and C_j, $i \neq j$, we define the conditional probability that, given a device c joins C_i, it also belongs to C_j:

$$p = P(c \in C_j | c \in C_i) \quad (1)$$

Note that, given the events $P(c \in C_j | c \in C_i)$ and $Q(c \in C_k | c \in C_i), \forall\, i, j, k \in [1 \dots n]$, we consider P and Q as independent events.

The probability γ_0 that $c \in C_n$ belongs only to C_n is given by the inverse of p. More specifically, this is the probability that, since c joins C_i it does not belong to any other of the $n-1$ communities:

$$\gamma_0 = (1-p)^{n-1}$$

The probability γ_h that since $c \in C_n$ it also belongs exactly to h other communities among $C_1, \dots, C_{n-1}$(and thus to $h+1$ communities in total, including C_n) is:

$$\gamma_h = \binom{n-1}{h} p^h (1-p)^{n-h-1} \quad (2)$$

Where p^h is the probability that c belongs to exactly h other communities, and $(1-p)^{n-h-1}$ is the probability that c does not belong to $n-h-1$ communities.

From equation (1), we can now derive the probability that since $c \in C_n$, it also belongs to $C_1 \cup \dots \cup C_{n-1}$ is:

$$\gamma = \sum_{h=1}^{n-1} \gamma_h$$

$$\gamma = \frac{1-(1-p)^n - p}{1-p}$$

Note that the probability that $c \in C_n$ belongs only to C_n, namely γ_0 is the inverse of γ.

We now assess the contribution of the k-*th* M²EC, which is the number of devices in C_k but not in $C_1, \dots, C_{k-1}$ that are connected only to the *k-th* M²EC. The contribution η_k is given by:

$$\eta_k = \sum_{i=0}^{g} i \binom{g}{i} \gamma_0^i \gamma^{g-i}$$

$$\eta_k = g(1-p)^{k-1} \quad (3)$$

V. The Experimental Dataset

The model presented in Section IV is validated with a real-world MCS dataset named ParticipAct [48]. We first describe the quality of the dataset, by showing the number of users and the traces collected during the experiment. Then, we present a strategy for detecting the communities, specifically those devices sharing the same position with a specific time frame. To this purpose, we present several settings used the community detection.

ParticipAct provides a mobility dataset from students of the University of Bologna, Italy. The dataset has been collected from December 2013 to February 2015. The volunteer students were equipped with an Android smartphone provisioned with an MCS application able to track the location through the Google location APIs. For each participant, the location is obtained by merging information from Wi-Fi Hot Spot coordinates, GPS and cell phone base station. For the purpose of our analysis, we extract from ParticipAct a subset of mobility traces, from January 2014 to December 2014.

One of the main problems with MCS campaigns is related to the process of recruiting users. Recruitment means involving volunteers in data acquisition campaign. The difficulty in the recruitment process is mostly due to the scepticism of users towards the technology and their reluctance to share sensible information through their devices [49]. Besides, small afflictions such as fast battery consumption or the excessive use of computational resources due to intensive sensing activity are further aspects that limit the wide acceptance of this paradigm.

The recruitment process of the ParticipAct users aimed at involving interested students. To this end, the organizers offered economic incentives including free smartphones and flat data plans. The recruitment process never stopped for the whole duration of the project. Therefore, the number of users varies along with the time. We show in Fig. 1, the number of users providing useful data in the considered sub-period (January 2014 to December 2014). The number of users varies along the time and this depends on two main factors. Firstly, students were free to switch off their smartphone, to disable the localization features or to uninstall the mobile app used to gather the location. All of these events prevent from the collection of data, and therefore the number of users decreases as well. Secondly, the ParticipAct architecture encountered few technical issues, such as shutdown and system crash of the backend. During the whole 2014, we observed 4 main interruptions: early February, March, early June and early November. They had a limited duration of approximately 1 to 7 days. The March event last about 18 days. On the low part of Fig. 1, we show a heatmap with the number of users providing data, the heatmap is aggregated on a daily basis for every week of the year (52 weeks in total). The heatmap reflects the interruption events previously described. In particular, the event on early February happed the Sunday of the 4th week, the March event during weeks 9th to 10th, the early June event between the weeks 21st and 22nd, and the early November on week 23rd.

The number of users affects the amount of GPS traces collected. Fig. 1 also shows the percentage of the traces collected. We estimate that each user can provide at maximum 576 points every 24 hours (one each 2.5 minutes). The percentage is therefore given by dividing the amount of traces collected in a day, with respect to the maximum amount of traces that could be gathered. We measure an average of 75% of traces collected during the whole 2014, reflecting the good quality of the dataset that we analyze. The ratio varies according to the number of available users, as expected the fewer the users, the fewer the traces collected.

We now present the way we detect communities in ParticipAct. The goal is to identify groups of users sharing similar locations, so that to measure the contribution of each of the M²ECs as explained in Section III. For the purpose of this work, a community is defined as a set of distinct users with similar locations during a time period. Communities are not static in ParticipAct, rather they evolve over time. For example, communities detected on day *x* might differ remarkably from communities detected the day after. In order to capture such behavior, we run at regular intervals a spatial algorithm for detecting communities. We refer to each interval as a *layer*, whose output is a set of communities. As a representative example, we show in Fig. 2 the process we follow for detecting the communities. We first build the aggregate network by selecting a starting and an ending period (i.e. January 2014 to October 2014). Then, we detect the communities by running the DBSCAN [50] algorithm periodically, the time distance among layer is defined by Δ. DBSCAN analyzes all the GPS points available from the previous to the current layer, as for example

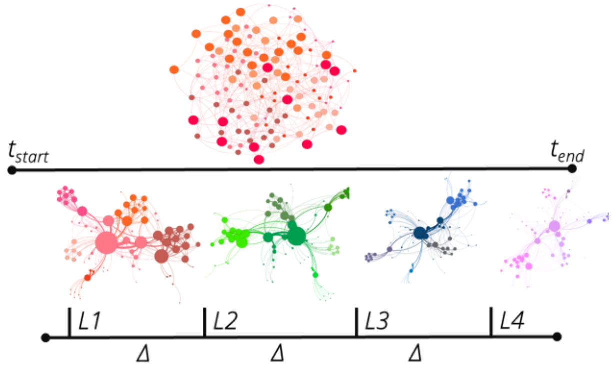

Fig. 2. Clustering with one layer every 2 days.

from L_1 to L_2, from L_2 to L_3 and from L_3 to L_4 (i.e. in this case $\Delta = 2$ days). The output after each run of the algorithm is a set of communities in the form of a list of user's IDs. DBSCAN works by clustering together GPS points with similar positions. More specifically, every cluster is composed by GPS points whose distance is at most ε and with at least ω points. We set a

distance $\varepsilon = 100$ meters in order to cluster those users laying in a range typical of ad-hoc connections among devices (Bluetooth or Wi-Fi Direct interfaces). For what concerns the minimum number of points, we set $\omega = u \times t$. It is given by $u = 3$, the minimum cardinality of a community (at least 3 people inside a community), and t which is the 50% of the expected GPS points that each user can provide. More specifically, the sampling rate of the locations is set to 2.5 minutes, therefore every user can provide at most 576 points in 24 hours. For the purpose of the community detection, we require that a

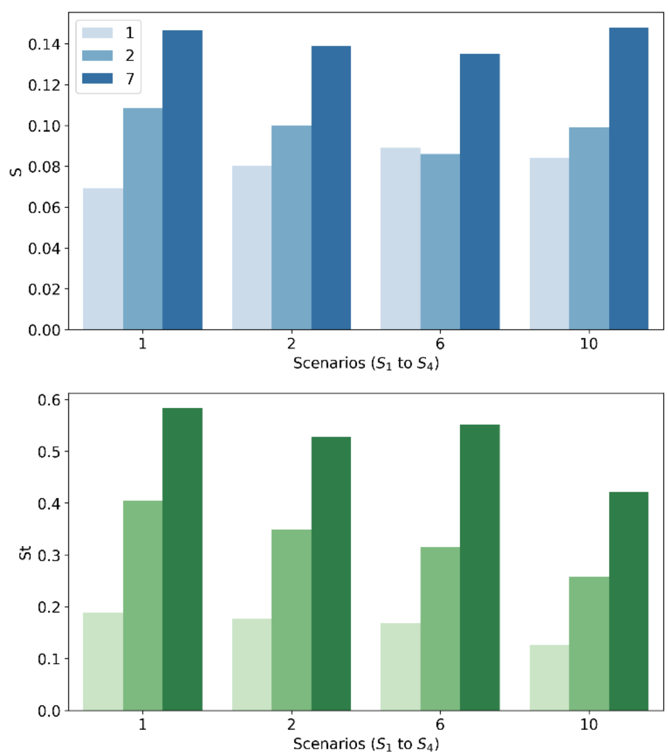

Fig. 3 Similarity (*S*) and Stability (*St*) of communities in 4 scenarios.

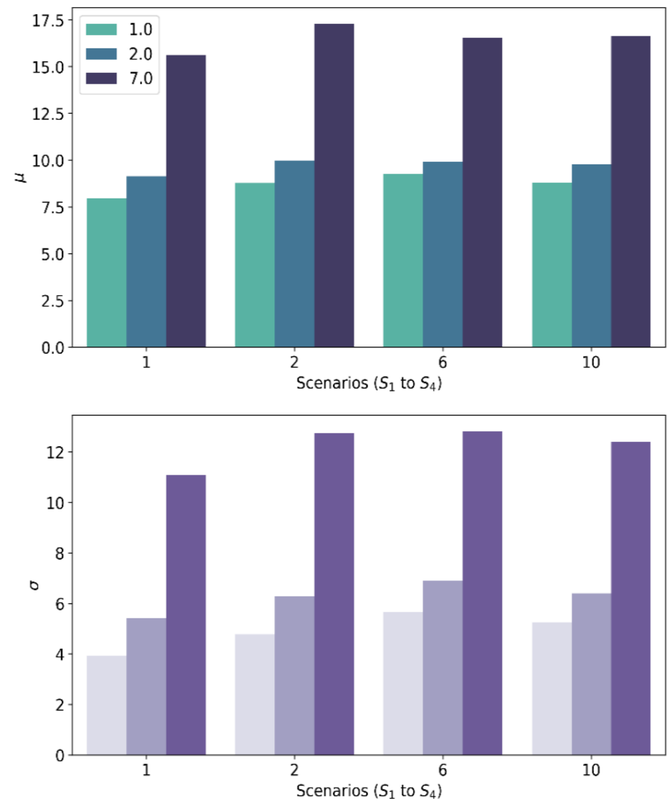

Fig. 4 Mean and std. deviation of the cardinality of the communities in 4 Scenarios.

community is composed by 3 distinct nodes each of which must provide at least $\omega = 3 \times 288$ points.

The parameter Δ determines the temporal distance between layers, therefore how frequently DBSCAN runs. Δ affects the communities detected, in fact by lowering Δ we increase the frequency of DBSCAN (e.g. daily), but every run is completed by using only the GPS points available during the layer. As for example, with Δ = 1-day DBSCAN is executed once per day with all points available from all the users in 1 day. Differently, setting Δ = 7 days, DBSCAN is executed once per week by using all the GPS points provided in 7 days. As results, communities detected with few GPS points might result not representative of the clusters of users during the layer.

We now present our strategy for selecting the value Δ. To this purpose we define the features that we expect from the communities, the *similarity (S)* and *stability (St)*. The selection strategy described in Section III-B is designed to assign to each community a M²EC in charge of collecting/sharing data from/to the members of its community. The more communities are found, the higher the number of M²EC that should be selected. Therefore, a first objective is to detect non-overlapping communities, they are communities that don't share users. In this way, each of the M²ECs selected can interact with distinct users, by avoiding replication of information collected/shared, this property is referred to as similarity. Similarity is computed by averaging the Jaccard index computed between all the communities detected in layers $L_1 \dots L_h$, as follows:

TABLE I
SCENARIOS SELECTED FOR EVALUATING THE MODEL

Scenario	Period	Duration	Similarity/ Stability	Mean/std dev	Δ value
S_1	Jan-Oct	10 months	0.15/0.42	16.62/12.38	7
S_2	Jan-Jun	6 months	0.09/0.32	9.9/6.9	2
S_3	Jan-Oct (2-by-2)	2 months	0.09/0.3	10.4/6.93	2
S_4	Jan-Oct (1-by-1)	1 month	0.12/0.41	10.2/5.9	2

$$S = \frac{1}{k} \sum_{\forall (C_x, C_y) \in L_i} J(C_x, C_y), \qquad \forall\, L_i \ldots L_h$$

Where $k = \binom{n}{2}$ is the number of combinations of n distinct communities found at layer L_i and J is the Jaccard index computed between C_x and C_y. The similarity S is therefore bound between [0,1], the higher S the more the communities are similar inside each layer. Conversely the lower S, the more the communities differ inside each layer.

Moreover, our selection strategy should be run rarely so that to select only once the M^2ECs. Our second goal is therefore to detect communities stable over the time. In other words, the members of the communities found at layer L_1 should remain similar over the time. This second property is referred to as stability. Stability (*St*) is computed by averaging the Jaccard index among the distinct users found in each layer, as follows:

$$St = \frac{1}{m} \sum_{\forall (i,j)} J(U_i, U_j)$$

where $m = \binom{h}{2}$ is the number of combinations of h layers, J is the Jaccard index between U_i and U_j, the list of users found at layers L_i and L_j and (i, j) are all the pairwise combinations of layers (e.g. $[L_1, L_2]$, $[L_2, L_3]$, … ,$[L_{h-1}, L_h]$). The higher *St* the more the communities are stable over the time. We analyze *S* and *St* by varying three key parameters:

1) *the value of Δ*
2) *the duration of the dataset*
3) *the time period used to compute similarity and stability.*

Δ ranges from: 1, 2 and 7 days. The duration of the dataset varies in the range: 1 month, 2 months, 6 months to 10 months. The time period we used depends on the duration. In particular, for 1 month duration, we analyzed the communities during the months: January 2014 to October 2014; for duration 2 months we considered the month pairs: January–February, February–March, March–April, April–May, May–June, June –July, July–August, August–September and September–October. Differently, for duration 6 months we considered the single period January–June, and for the duration 10 months the single period January–October.

Fig. 3 reports the results for *S* and *St*. As a general observation, we note that by increasing the duration of the layer Δ (1 to 7 days), similarity and stability increase as well. In fact, the duration of the layer has the effect of increasing/ decreasing the amount of points used by DBSCAN to detect clusters. Clusters detected with few GPS points, i.e. Δ = 1 day (24 hours) result too weak to capture any routinely movement of users of ParticipAct. Differently, with Δ = 7 days, clusters are more robust remaining more stable week after week. Moreover, we observe that *S* and *St* do not significantly change by setting a specific value of Δ across different durations (1 month to 10 months). As for example, fixing duration 1 month and varying Δ = 1 to 7 days, the standard deviation of S ranges from 0.03 to 0.02. Similarly, fixing duration 2 months and ranging Δ from 1

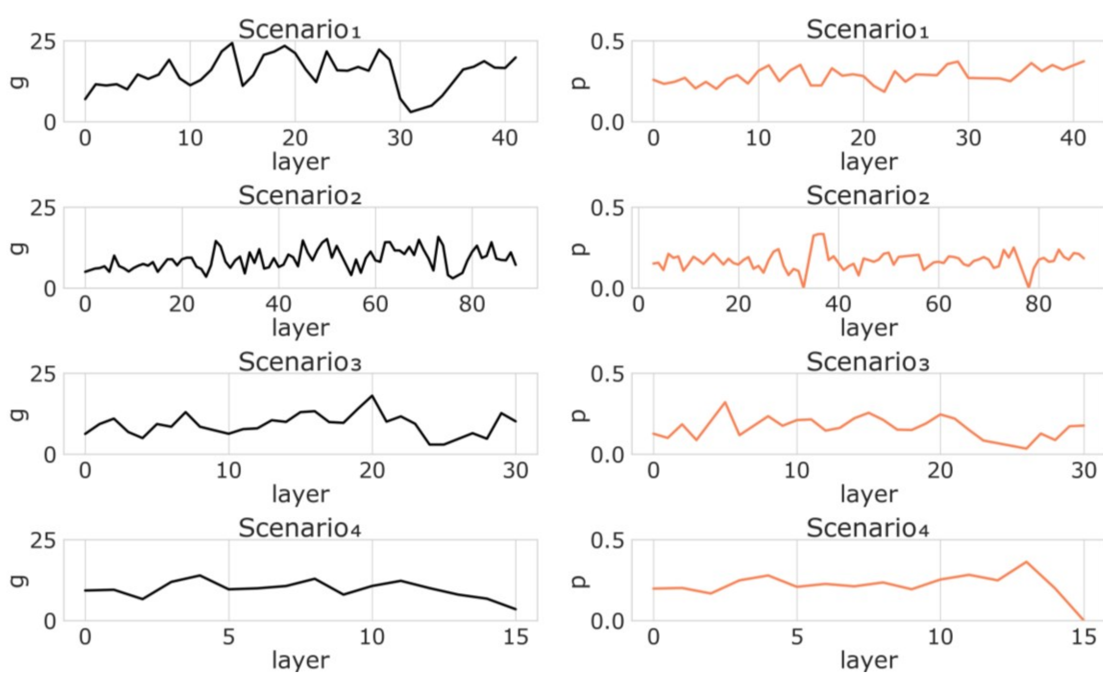

Fig. 5 Trends of *g* and *p* in scenarios S_1 to S_4.

to 7 days, σ_S ranges from 0.02 to 0.009. The same considerations also apply for the values of stability.

Finally, we measure mean μ and standard deviation σ of the cardinality of the communities detected, as shown in Fig. 4. As expected by increasing Δ, communities also increase since more GPS points from different users are considered in order to detect the clusters. We observe also higher values of σ when Δ varies from 1 day to 7 days.

In order to evaluate the model presented in Section IV, we select 4 representative scenarios. The idea is to select very different conditions in which test our model. To this purpose, we consider very long, long, medium and a short-range duration, and we chose a value of Δ accordingly. We consider that such settings well reflect the different conditions of the user mobility in ParticipAct. The settings of such scenarios are summarized in Table I.

VI. Experimental Results

The goal of our experimentation is to configure the model parameter and to assess how much it fits the results obtained from the ParticipAct. To this purpose, we adopt the following methodology:

1) *calibration of the model*
2) *contribution of M^2EC in ParticipAct*
3) *comparison of the models.*

A. Model Calibration

Recall that, from Section III, η_k defined in Equation (3) models the expected contribution of each of the M^2ECs selected, and it depends on parameters g and p that are the average cardinality of the communities detected and the probability defined in Equation (1), respectively. The number k of the selected M^2EC is computed with the strategy defined in Section III-B. For example, with $k = 5$ we select 5 M^2EC each of which covers a distinct community. We calibrate parameters g and p for each of the scenarios in Table I. In particular, g is calibrated by averaging the cardinality of the communities detected with DBSCAN, layer per layer, while p is calibrated by averaging the probability of all communities' nodes, without repetitions, in each layer. Fig. 5 shows how g and p vary in the 4 scenarios. From the figure, we observe that for both values long-term tests (S_1 and S_2) show a more fluctuating trends than medium-term tests (S_3 and S_4). The period in which g and p have the more stable trend corresponds to the setting S_4. Although on different scales, the range in which both values fluctuate is comparable: g ranges from 0 up to 25 community's members while the probability p is always bound between 0 and 0.25.

$$
\begin{array}{l}
L_1 = [\; v_1^1 \;\; v_2^1 \ldots v_n^1 \;] \\
L_2 = [\; v_1^2 \;\; v_2^2 \ldots v_n^2 \;] \\
\qquad\quad \ldots \\
L_h = [\; v_1^h \;\; v_2^h \ldots v_n^h \;] \\
\qquad\;\; \overline{C_1} \;\; \overline{C_2} \ldots \overline{C_n}
\end{array}
$$

Fig. 6. Communities contribution.

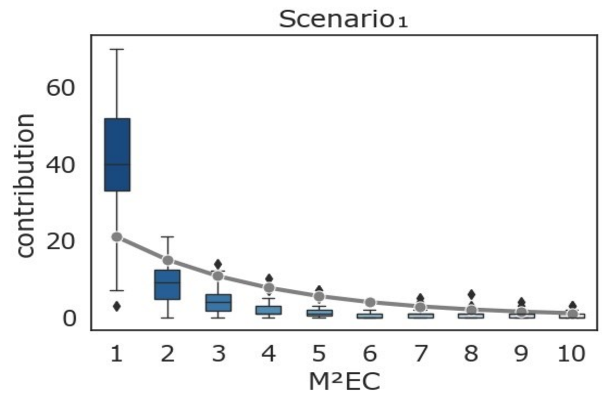

Fig. 7. Comparison of the probabilistic model in Scenario 1.

B. Computing the Contribution of M^2EC in ParticipAct

For every scenario described Table I, we detect the communities at each layer by using DBSCAN properly configured. Each community detected is represented by a M^2EC according to the selection strategy defined in Section III-B. For example, given the layer L_j, DBSCAN identifies a set $\{C_1, \ldots, C_n\}$ of n communities (ordered for decreasing size), and the algorithm selects, for each community, a representative M^2EC according to some centrality measure. Let v_i^j be the contribution of the M^2EC for C_i at L_j. It is immediate that $v_1^1 = |C_1|$ and, in general, $v_i^j = C_i \setminus \cup_{l<i} C_l$. As a result, we obtain h vectors of contribution values, one for each layer $L_1, \ldots, L_h$, as shown in Fig. 6. We are interested in measuring the distribution of the contribution values over the h layers. In other words, we study how the vector of contributions varies from the widest community computed over h layers ($\overline{C}_1$), to the vector of contribution of the smallest community computed over h layers ($\overline{C}_n$), as shown in Fig. 6.

C. Comparison of the model prediction with experimental results

We present here the comparison of Equation (3) with data obtained from the ParticipAct real-world dataset. Specifically, we show the distribution of the contribution values as boxplots, one for each dimension community (from the widest to the smallest), the trend of the model is shown with a super-imposed solid line. The boxplots show median, 25th and 75th percentile in the body, and minimum and maximum points of the

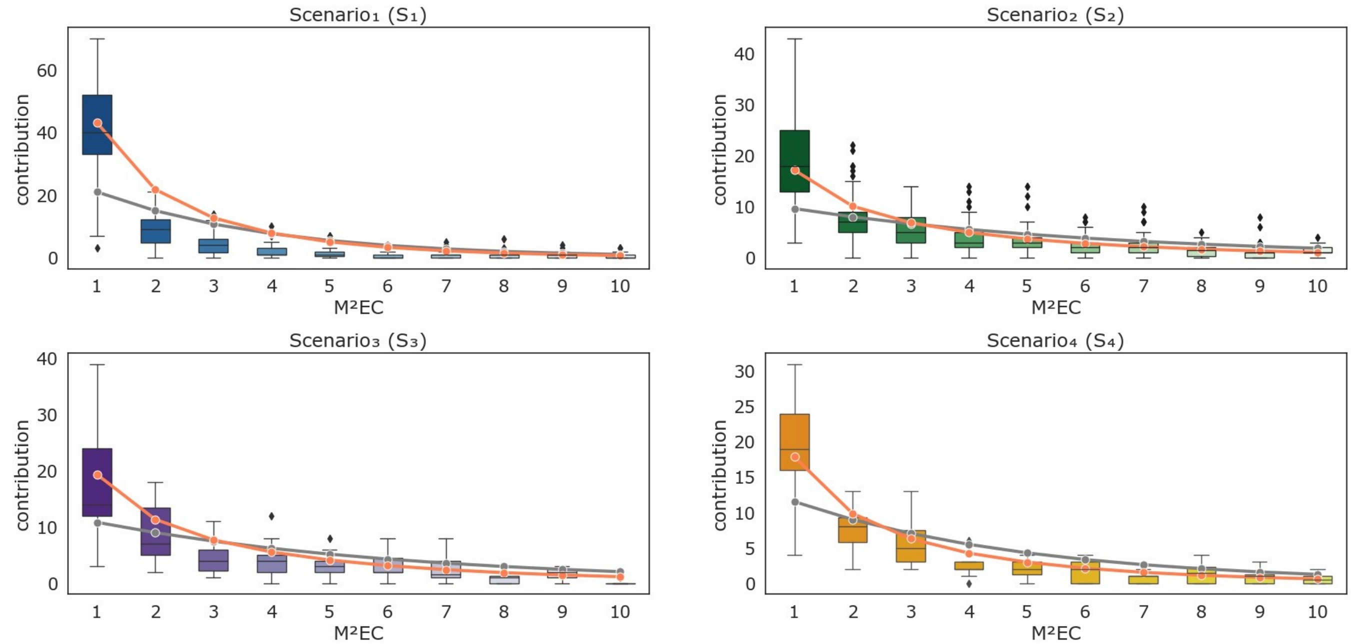

Fig. 8. Comparison of models with the 4 Scenarios.

contribution values in the whiskers. The outliers are shown as black diamonds. Fig. 7 shows a comparison of our model with respect to results we obtain from ParticipAct only in Scenario 1 (see Table I). We observe that our model underestimates the contribution of the first M^{2}EC, while it overestimates the contribution of the subsequent ones. In particular, the grey line is always below the median value of the boxplots for M^{2}EC #1 and above the median for all the others. This behavior depends on the fact that our model assumes that the communities have all the same cardinality g (see Section IV). Such assumption is acceptable since the number of users of a CrowdSensing platform is higher than the maximum size of the user's communities (which are necessarily bounded) and becomes too restrictive when the number of users considered is limited. Consequently, the variation in the size of the communities becomes non-negligible (as discussed in Fig. 4).

In order to overcome this limitation, we determine empirically a correction factor, that we apply to Equation (3). Such factor considers the variable size of the communities considered by introducing a correction related to the standard deviation σ of the size of the communities and the value of k. According to this correction, the model in (3) becomes:

$$\eta_k = \sqrt{\frac{\sigma}{\alpha * k}}\, g(1-p)^{k-1} \qquad (4)$$

where α is a smoothing factor for a fine-grained tuning.

We are now able to compare the model described in (4) with respect to the 4 scenarios S_1 to S_4. The scenarios reproduce different settings on the way we detect community and on the duration of the dataset. We span from 10 months (S_1) to 1 month (S_4). Fig. 8 reports the performance of the models proposed in (3) and (4) with respect to the contribution of the M^{2}EC computed in ParticipAct as described in Section VI-B. The figure shows in grey the original model without the correction factor, and in orange the model with the correction factor. We restrict the comparison to the first 10 M^{2}EC, after which we observe that their contribution is negligible.

We observe that the model in (4) well reproduces the trend of the M^{2}EC contribution in all the scenarios, both in the long ones (S_1 and S_2) and in the short ones (S_3 and S_4). As a general trend, the model provides high contribution for the first M^{2}ECs, after which the contribution rapidly decreases. In particular, the model in (4) estimates correctly the contribution of M^{2}EC #1 which is generally higher with respect to the others, this can be observed in Fig. 8, since the orange dots always fit close to the median values of the boxplots.

D. *Quantitative Measurement of the Statistical Data Fitting Technique*

In order to assess the adaptability of the adjusted model to the real community data set, we applied the Kolmogorov-Smirnov (K-S) test on both distributions. The K-S test is well suited to our case study because it allows to compare the shape of two sample distributions by assessing the adaptability of each other. Specifically, the result of the K-S test accepts or rejects the hypothesis of the adaptability of the real data distribution with the adjusted theoretical sample distribution. Tests have been performed on the four proposed scenarios with a critical value of $K_\alpha = 1.36$ ($i.e,\ \alpha = 0.05$).

TABLE II
RESULTS OF K-S TESTS FOR THE DIFFERENT SCENARIOS.

S_1	S_2	S_3	S_4
0.110	0.974	0.974	0.974

Table II shows the *p-values* returned for each of the 4 scenarios, e.g. S_1 to S_4. We observe that the hypothesis of the adaptability of the theoretical model to the real data distribution can be accepted for three out of four scenarios. As reported in Table I, S2, S3, and S4 last respectively 6, 2, and 1 months. In such scenarios, mobility and sociality of users tend to remain more stable with minor fluctuations. Differently, scenario S1 corresponds to a very long period (10 months, from January to October 2014). During such long period, users cross very different conditions: lessons, test period, lessons, Eastern break, summer holidays, test period, and back to lessons. Therefore, mobility and sociality are also affected by such variation. Moreover, during that period we observed some technical issues along the data collection that reduced drastically the amount of data collected. Figure 1 reports a graphical representation of users and traces collected during the whole 2014. As a result, the K-S test on scenario S_1 returns a *p-value* lower than the others, below the level of acceptability of our hypothesis. We still include scenario S_1 for the purpose of completeness; however, our model has been designed to exploit routine of user mobility in order to select M^2EC. Therefore, it is a case not particularly relevant for our algorithm that can be configured to select the M^2EC devices more frequently.

VII. Conclusions and Future Work

The HEC model implements MEC technologies in massive sensing application scenarios such as in MCS [51]. Commonly, these architectures are characterized by edges supporting data collection operations of terminal nodes. Some recent proposals explored the possibility of implementing such edges with mobile devices belonging to the users of the MCS platform itself. This enables a dynamic selection of edge nodes (called M^2EC), that should, however, consider the social relationships among the plethora of MCS users, to reduce platform costs and ensure a good spatial coverage. In this work, we addressed the problem of selecting the number of mobile nodes to promote as M^2EC by proposing a probabilistic model.

In order to devise the selection algorithm and to reduce our model to a closed-form expression, we conducted a qualitative analysis of real-world mobility traces dataset, that have been also used for the validation of the model itself. Specifically, based on the analysis of the traces in the dataset we designed the algorithm in a way to select first the representative nodes of the larger communities. Secondly, the limited difference in size between the communities and the fact that the intersection among communities is not empty suggested to approximate the size of the communities with a constant and to model with Equation 1 the probability that the same device belongs to different communities. Finally, the analysis of the dataset proved also useful to determine the correction factor for the calibration of the model to achieve a good fitting (also confirmed by the K-S test) of the model with the experimental results.

These encouraging results are leading us to reflect on some future research opportunities. Since the proposed probabilistic model focuses only on M^2EC, a first step forward could be the implementation of such model into an architecture made up of both fixed and mobile MEC. Moreover, the model is sensitive to community size, and based on the community detection algorithm in use the result may vary significantly. For this reason, as a further advancement of the present work we are considering a longitudinal comparison of the main community detection algorithms designed to capture dynamics of the human mobility in MEC-based MCS systems.

References

[1] A. Obinikpo, Y. Zhang, H. Song, T. H. Luan, and B. Kantarci, “Queuing algorithm for effective target coverage in mobile crowd sensing”, IEEE Internet of Things Journal, vol. 4, no. 4, pp. 1046–1055, Aug 2017.

[2] D. Belli, S. Chessa, A. Corradi, G. Di Paolo, L. Foschini, M. Girolami, “Selection of mobile edges for a hybrid crowdsensing architecture”, IEEE Symposium on Computers and Communications (ISCC), 2019.

[3] P. Bellavista, D. Belli, S. Chessa, L. Foschini, “A social-driven edge computing architecture for mobile crowd sensing management”, IEEE Communications Magazine, vol. 57, pp. 68-73, 2019.

[4] D. Belli, S. Chessa, L. Foschini, M. Girolami, “A social-based approach to mobile edge computing”, IEEE Symposium on Computers and Communications (ISCC), pp. 292-297, 2018.

[5] V. I. Nistorica, C. Chilipirea, and C. Dobre, “How many people areneeded for a crowdsensing campaign?”, in IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP), Sep 2016, pp. 353–358.

[6] P. Hu, S. Dhelim, H. Ning, and T. Qiu, “Survey on fog computing: Architecture, key technologies, applications and open issues”, Journal of Network and Computer Applications, vol. 98, pp. 27 – 42, 2017.

[7] T. Taleb, k. Samdanis. B. Mada, H. Flinck, S. Dutta, D Sabella, “On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration”, IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1657-1681, 2017.

[8] A. Ahmed and E. Ahmed, “A survey on mobile edge computing”, Proc. 10th International Conference on Intelligent Systems and Control (ISCO), pp. 1-8, 2016.

[9] K. Farkas, G. Feher, A. Benczur, and C. Sidlo, “Crowdsending-based public transport information service in smart cities”, IEEE Communications Magazine, vol. 53, no. 8, pp. 158–165, Aug 2015.

[10] M. Moreno, A. Skarmeta, and A. Jara, “How to intelligently make sense of real data of smart cities”, in International Conference on Recent Advances in Internet of Things (RIoT), pp. 1–6, Apr 2015.

[11] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang, and X. Zhou, “Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm”, ACM Comput. Surv., vol. 48, no. 1, pp. 1–31, Aug. 2015.

[12] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, A. Neal, “Mobile-edge computing introductory technical white paper”, White Paper, Mobile-Edge Computing (MEC) Industry Initiative, pp. 1089-7801, 2014.

[13] L. Summers and S. D. Johnson, “Does the configuration of the street network influence where outdoor serious violence takes place? Using space syntax to test crime pattern theory”, Journal of Quantitative Criminology, vol. 33, no. 2, pp. 397–420, Jun 2017.

[14] S. Rosen, S.-J. Lee, J. Lee, P. Congdon, Z. M. Mao, and K. Burden,“MCNet: Crowdsourcing wireless performance measurements throughthe eyes of mobile devices”, IEEE Communications Magazine, vol. 52, no. 10, pp. 86–91, 2014.

[15] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, “Urban computing: Concepts, methodologies, and applications”, ACM Transactions on Intelligent Systems and Technology, vol. 5, no. 3, pp. 1–55, Sep 2014.

[16] H. C. Hsieh, J. L. Chen, A. Benslimane, “5G virtualized multi-access edge computing platform for IoT applications,” Journal of Network and Computer Applications, vol. 115, pp. 94-102, 2018.

[17] S. Barbarossa, S. Sardellitti, P. Di Lorenzo, “Communicating while computing: Distributed mobile cloud computing over 5G heterogeneous networks”, IEEE Signal Processing Magazine, pp. 45-55, 2014.

[18] M. Girolami, P. Barsocchi, S. Chessa and F. Furfari, “A social-based service discovery protocol for mobile Ad Hoc networks”, 12th Annual Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET), pp. 103-110, 2013.

[19] Y. Han, Y. Zhu, and J. Yu, “A distributed utility-maximizing algo-rithm for data collection in mobile crowd sensing”, in IEEE Global Communications Conference (GLOBECOM), pp. 277–282, Dec 2014.

[20] S. Retal, M. Bagaa, T. Taleb, H. Flinck, “Content delivery network slicing: QoE and cost awareness”, IEEE International Conference on Communications (ICC), pp. 1-6, 2017.

[21] B. Tang, Z. Chen, G. Hefferman, S. Pei, T. Wei, H. He, “Incorporating intelligence in fog computing for big data analysis in smart cities”, IEEE Transactions on Industrial Informatics, pp. 2140-2150, 2017.

[22] P. Cassarà, F. Potortì, P. Barsocchi and M. Girolami, "Choosing an RSS device-free localization algorithm for Ambient Assisted Living", International Conference on Indoor Positioning and Indoor Navigation (IPIN), Banff, AB, pp. 1-8, 2015.

[23] N. H. Motlagh, M. Bagaa, T. Taleb, “UAV-based IoT platform: A crowd surveillance use case”, IEEE Communications Magazine, pp. 128-134, 2017.

[24] E. K. Markakis, I. Politis, A. Lykourgiotis, Y. Rebahi, G. Mastorakis, C. X. Mavromoustakis, E. Pallis, “Efficient next generation emergency communications over multi-access edge computing”, IEEE Communications Magazine, vol. 55, n. 11, pp. 92-97, 2017.

[25] E. Barka, C. A. Kerrache, N. Lagraa, and A. Lakas, “Behavior-aware UAV-assisted crowd sensing technique for urban vehicularenvironments,” in 15th IEEE Annual Consumer CommunicationsNetworking Conference (CCNC), Jan 2018, pp. 1–7.

[26] Fazel, B. Kantarci, “SOBER-MCS: Sociability oriented and battery efficient recruitment for mobile crowd-sensing”, Sensors, 1593, May 2018.

[27] N. D. Lane, Y. Chon, L. Zhou, Y. Zhang, F. Li, D. Kim, G. Ding, F. Zhao, and H. Cha, “Piggyback crowdsensing (PCS): energy efficient crowdsourcing of mobile sensor data by exploiting smartphone app opportunities,” in Proc. of the ACM Conference on Embedded Networked Sensor Systems, ser. SenSys, 2013.

[28] W. Sherchan, P. P. Jayaraman, S. Krishnaswamy, A. Zaslawsky, S. Loke, A. Sinha, “Using on-the-move mining for mobile crowdsensing”, IEEE 13th International Conference on Mobile Data Management, pp. 115-124, 2012.

[29] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovic, P. Bouvry, “A survey on mobile crowdsensing systems: Challenges, solutions and opportunities,” IEEE Communications Surveys & Tutorials, 2019.

[30] R. K. Ganti, F. Ye, and H. Lei, “Mobile crowdsensing: current state and future challenges”, IEEE Communications Magazine, vol. 49, n. 11, pp. 32-39, 2011.

[31] D. Wu, H. Li, R. Wang, “User characteristic aware participant selection for mobile crowdsensing”, Sensors, 3959, Nov. 2018.

[32] M. A. Alsheikh, Y. Jiao, D. Niyato, P. Wang, D. Leong, Z. Han, “The accuracy-privacy trade-off of mobile crowdsensing,” IEEE Communications Magazine, pp. 132-139, 2017.

[33] H. Xiong, D. Zhang, L. Wang, J. P. Gibson, and J. Zhu, “EEMC: Enabling energy-efficient mobile crowdsensing with anonymous participants”, ACM Transactions on Intelligent Systems and Technology (TIST), vol. 6, no. 3, pp. 1–26, Apr. 2015.

[34] F. Montori, L. Bedogni, A. D. Chiappari, and L. Bononi, “SenSquare: A mobile crowdsensing architecture for smart cities”, in 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT), pp. 536–541, Dec. 2016.

[35] M. Marjanović, A. Antonić, I. P. Žarko, “Edge computing architecture for mobile crowdsensing,” IEEE Access, pp. 10662-10674, 2018.

[36] L. Ma, X. Liu, Q. Pei, Y. Xiang, “Privacy-preserving reputation management for edge computing enhanced mobile crowdsensing,” IEEE Transactions on Services Computing, 2018.

[37] F. Restuccia, S. K. Das & J. Payton, “Incentive mechanisms for participatory sensing: Survey and research challenges”, ACM Transactions on Sensor Networks (TOSN), vol. 12, no. 2, 2016.

[38] Y. Chen, H. Chen, S. Yang, X. Gao, F. Wu, “Jump-start crowdsensing: A three layer incentive framework for mobile crowdsensing,” IEEE/ACM 25th International Symposium on Quality of Services, pp. 1-6, 2017.

[39] J. Lin, M. Li, D. Yang, and G. Xue, “Sybil-proof online incentive mechanisms for crowdsensing”, in IEEE Conference on Computer Communications (INFOCOM), Apr 2018.

[40] C. Fiandrino, B. Kantarci, F. Anjomshoa, D. Kliazovic, P. Bouvry, J. Matthews, “Sociability-driven user recruitment in mobile crowdsensing internet of things platforms”, IEEE Global Communications Conference GLOBECOM, pp. 1-6, 2016.

[41] Z. He, J. Cao, and X. Liu, “High quality participant recruitment in vehicle-based crowdsourcing using predictable mobility,” inProc.IEEE Conf. Comput. Commun. (INFOCOM), Hong Kong, Apr. 2015,pp. 2542–2550.

[42] H. Li, T. Li, and Y. Wang, “Dynamic participant recruitment of mobilecrowd sensing for heterogeneous sensing tasks,” inProc. IEEE 12thInt. Conf. Mobile Ad Hoc Sensor Syst., Oct. 2015, pp. 136–144.

[43] M. Karaliopoulos, O. Telelis, and I. Koutsopoulos, “User recruit-ment for mobile crowdsensing over opportunistic networks,” inProc.IEEE Conf. Comput. Commun. (INFOCOM), Hong Kong, Apr. 2015,pp. 2254–2262.

[44] Z. Feng, Y. Zhu, Q. Zhang, L. M. Ni, and A. V. Vasilakos,“TRAC: Truthful auction for location-aware collaborative sensing inmobile crowdsourcing,” inProc. IEEE Int. Conf. Comput. Commun.(INFOCOM), 2014, pp. 1231–1239.

[45] D. Zhang, H. Xiong, L. Wang, and G. Chen, “CrowdRecruiter:Selecting participants for piggyback crowdsensing under probabilis-tic coverage constraint,” inProc. ACM Int. Joint Conf. PervasiveUbiquitous Comput. (UbiComp), 2014, pp. 703–714

[46] Y. Lin, F. Wu, L. Kong and G. Chen, "Quality-Based User Recruitment in Mobile CrowdSensing," 2018 14th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN), Shenyang, China, 2018, pp. 74-80.

[47] N. B. Truong, G. M. Lee, T. Um and M. Mackay, "Trust Evaluation Mechanism for User Recruitment in Mobile Crowd-Sensing in the Internet of Things," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 10, pp. 2705-2719, Oct. 2019.

[48] G. Cardone, A. Cirri, A. Corradi, and L. Foschini, “The ParticipAct Mobile CrowdSensing Living Lab: The Testbed for Smart Cities”, IEEE Communications Magazine, vol. 52, n. 10, pp. 78-85, 2014.

[49] M. Karliopoulos, O. telelis, I. koutsopoulos, “User recruitment for mobile crowdsensing over opportunistic networks”, IEEE Conference on Computer Communications (INFOCOM), pp. 2254-2262, 2015.

[50] M. Ester, H. P. Kriegel, J. Sander, X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise”, Kdd, vol. 96, n. 34, pp. 226-231, 1996.

[51] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, “Mobile edge computing—A key technology towards 5G”, ETSI White Paper, vol. 11, no. 11, pp. 1-16, 2015.

Revision Letter

Submission to IEEE IoT Journal
"A Probabilistic Model for the Deployment of Human-enabled Edge Computing in Massive Sensing Scenarios"

Pisa, October 15th, 2019

Dear Editor in Chief,

we have carefully read and considered comments from the Reviewers and, along the suggested guidelines, we have revised our paper entitled:

"A Probabilistic Model for the Deployment of Human-enabled Edge Computing in Massive Sensing Scenarios" by Stefano Chessa, Michele Girolami, Luca Foschini and myself, to be considered for possible publication to IEEE Internet of Things Journal.

For your convenience, this accompanying letter details how we have answered to the received revision requests and to all the points raised by the Reviewers.
Please feel free to contact me for any possible problem about the revised paper submission or for any clarification needed about the revisions made.

Best Regards,

Dimitri Belli

Dimitri Belli
University of Pisa
Largo Bruno Pontecorvo 3, 50126 Pisa, ITALY
Phone: +39-050-22123179; E-Mail: dimitri.belli@di.unipi.it

AE: Associated Editor

Comment 1

The contributions with respect to the literature have to be better highlighted.

Answer 1

In order to better highlight the contributions of our work with respect to the present literature, we have re-structured Section II (Related Work). Section II is now organized in two subsections: II-A and II-B. Section II-A covers the MEC paradigm and the main application scenarios. We report in such section state-of-the-art solutions that exploit edge nodes in a cloud environment. Section II-B describes the MEC paradigm combined with the Mobile CrowdSensing (MCS) and describes those strategies for selecting devices in an efficient way. We describe some selection strategies based on the amount of data and the quality of data that the devices can provide, as well as strategies based on the sociality of users. Under this respect, we clarify the approach followed by the solution presented in this work. We report below the changes applied to the original paper:

“

II Related Work

Without claiming exhaustiveness, we survey in this section the MEC paradigm and the most recent massive application scenarios. More precisely, in Section II-A we review the main characteristics and limitations of the MEC paradigm. Section II-B describes the relationship between MEC and Mobile CrowdSensing (MCS), with particular attention to efficient device selection strategies for increasing the amount of data that can be gathered.

A The MEC paradigm

MEC provides a new ecosystem based on radio access network edges with computational and storage capabilities. The MEC model aims at supporting peripheral nodes of the network by reducing latency for mobile users, optimizing mobile backhaul and middleware layer nodes performance. Such decentralized cloud technology makes the MEC one of the cornerstones of the new 5G systems [6], easing the convergence between telecommunication and information technology services [7-8]. Traditional MEC implementations are made up of software platforms which provide local services without considering the user mobility. Differently, advanced MEC implementations also consider aspects like traffic [9, 10], mobility [31] and account, introducing a heterogeneous type of networking to support both commercial and non-commercial applications, both in indoor and outdoor [12, 13, 14, 15]. Concerning the use of MEC with the Internet of Things (IoT) applications, recent studies focused on the development of platforms that ensure management and interoperability between large-scale devices without loss of bandwidth or increased latency [16]. With a plethora of devices interacting with intermediate MEC nodes, computation offloading techniques became valuable methods to save energy, battery lifetime and calculations. As the production of computation offloading is vast, we limit to mention a comprehensive study on these techniques over 5G heterogeneous networks [17]. Other major use cases of the MEC paradigm include optimization techniques for distributed content discovery and delivery [18, 19] and caching [20], big data storage and computation [21] and, notably, services for smart cities as localization [22], emergency and public safety [23, 24, 25]. It worth to notice that in MEC scenarios the problem of resource constraints cannot be avoided. In particular, the battery drain is the main concern for mobile users. In recent years, however, user recruitment policies leveraging profiles and current battery level of devices for task assignments have been devised, allowing a reasonable energy saving for those devices with limited resources [26, 27]. Alongside cost-saving energy solutions, there is also bandwidth wastage. Some propositions based on local data mining by mobile devices, with production of intermediate results, refined, from raw sensing data, have shown a real energy and bandwidth saving in data transmission to remote servers [28].

B Device Selection Strategies

We are interested in combining the MEC paradigm with the Mobile CrowdSensing (MCS) [29], an approach for massive sensing in mobile scenarios. Nowadays, almost all mobile and wearable devices are equipped with multi-modal sensing capabilities which are able to collect data in a seamless way. The idea of MCS is to exploit the ubiquity of such devices in order to increase the amount and quality of data that can be gathered from the crowd [30]. Examples of data that can be collected (also referred to as tasks) include: media contents, recordings of audio tracks, sensor readings such as noise/light intensity etc. Strategies for task assignment include the study of recruitment areas considering the population density at different time of the day and the user's activity as the willingness in performing tasks and the time spent in city hotspots [31]. Under this respect, several studies also address privacy issue while collecting personal information [32, 33]. The MCS paradigm offers the possibility of gathering data from the crowd, while the MEC paradigm provides a scalable and dynamic architecture enabling such a massive collection of information. Recently, in the context of smart cities [34], hybrid solutions of MEC and MCS include models for optimizing content sharing by leveraging the sociality and mobility of the MCS nodes [4], MEC-based architectures for massive scale MCS services and privacy preserving citizens' data control flow [35,36], and social-driven edge computing architectures based on incentives and centrality measures to reduce installation and maintenance costs for MEC intermediate layer nodes [3].

In this context, one of the main challenges is the selection of those devices acting as edge of the cloud, namely those devices collecting data from others and uploading data to the backend [37, 38, 39]. An efficient selection of the devices can result with high spatial and temporal coverage (as discussed in [40, 41]). Spatial coverage considers the number of areas of a region that can be monitored with a certain accuracy, such as the districts of a city. While temporal coverage refers to the time required in order to collect data from all the regions. In [42] authors propose a selection strategy based on the dimension of the region to cover. Authors consider a deterministic node mobility and propose several heuristics for the selection of devices, in order to minimize an objective function. Other recent works [43-45] consider incentive mechanisms for the users in order to increase the probability of collecting data from specific locations. In [46] authors adopt a selection strategy based on the quality of data the user can produce. The authors adopt a Compressive Sensing technique in order to predict the expected quality of data from the user and, in turn, they select appropriately the best candidates. Authors of [47] also adopt a method to select devices based on the quality of data they provide. Authors adopt a trust model called experience-reputation (ER) which evaluated the "trust relationships" between any two mobile device users. Authors test the ER model on several MCS testbed in which regular and malicious users are present.

A different approach for the selection of devices is based on the analysis of sociality of users. Since devices are mainly carried by humans, it is possible to adopt strategies based on the analysis of features of the human sociality and mobility. The main idea is to measure how much a user encounters or clusters with other users and, in turn, to select those with maximizing the contact probability. In fact, the more a user interacts with others, the more likely its device can collect data from neighbours through short range network interfaces (such as Bluetooth, Wi-Fi Direct or the upcoming LTE direct). The contact probability can be measured with metrics typical of graph theory such as clustering coefficient, betweenness or eigenvector centrality that all measure the centrality of a node in a graph. The device selection strategy presented in this work is also based on the analysis of sociality among users (see Section III-A). Our approach consists in first detecting the communities and then in selecting a representative for each of them. The selection of the representative is based by measuring a centrality score."

Comment 2

The authors should explain in detail the used statistical data fitting technique. How the qualitative analysis of real-world datasets is used to derive the model's close form, should be explained in detail and motivated.

Answer 2

This point was also raised by Reviewer 2, who also suggested to use the Kolmogorov-Smirnov test (K-S test). We appreciated the comment from the reviewers and we decided to measure how our model fits with the observation from the ParticipAct dataset. To this purpose, we measure the K-S in order to accept or reject the *null* hypothesis that our model fits with the observation.

More specifically, we have carried out a measurement of how much our theoretical model presented in Section IV adapts to the observations of the ParticipAct dataset. To this end, we used the Kolmogorov-Smirnov test. All the tests have been performed with a critical value of $K_\alpha = 1.36$ ($i.e. \alpha = 0.05$). The results led to acceptance of the *null* hypothesis (i.e. the adaptability of the theoretical model to the real distribution) on three out of the four scenarios proposed. Let us note that the scenario in which the fitting is worse is the one less relevant for our algorithm and can be avoided by configuring the algorithm for a more frequent selection of the M^2EC. The presentation of the K-S test and its evaluation are presented as a new subsection of Section VI:

“

D. Quantitative Measurement of the Statistical Data Fitting Technique

In order to assess the adaptability of the adjusted model to the real community data set, we applied the Kolmogorov-Smirnov (K-S) test on both distributions. The K-S test is well suited to our case study because it allows to compare the shape of two sample distributions by assessing the adaptability of each other. Specifically, the result of the K-S test accepts or rejects the hypothesis of the adaptability of the real data distribution with the adjusted theoretical sample distribution. Tests have been performed on the four proposed scenarios with a critical value of $K_\alpha = 1.36$ ($i.e,\ \alpha = 0.05$). Table II shows the *p-values* returned for each of the 4 scenarios, e.g. S1 to S4. We observe that the hypothesis of the adaptability of the theoretical model to the real data distribution can be accepted for three out of four scenarios. As reported in Table I, S2, S3, and S4 last respectively 6, 2, and 1 months. In such scenarios, mobility and sociality of users tend to remain more stable with minor fluctuations. Differently, scenario S_1 corresponds to a very long period (10 months, from January to October 2014). During such long period, users cross very different conditions: lessons, test period, lessons, Eastern break, summer holidays, test period, and back to lessons. Therefore, mobility and sociality are also affected by such variation. Moreover, during that period we observed some technical issues along the data collection that reduced drastically the amount of data collected. Figure 1 reports a graphical representation of users and traces collected during the whole 2014. As a result, the K-S test on scenario S1 returns a *p-value* lower than the others, below the level of acceptability of our hypothesis. We still include scenario S_1 for the purpose of completeness; however, our model has been designed to exploit routine of user mobility in order to select M^2EC. Therefore, it is a case not particularly relevant for our algorithm that can be configured to select the M^2EC devices more frequently.

TABLE II
RESULTS OF K-S TESTS FOR THE DIFFERENT SCENARIOS.

S_1	S_2	S_3	S_4
0.110	0.974	0.974	0.974

R1: Review 1

Comment 1

The idea of probabilistic estimation of the number of MMEC seems to be relatively interesting. However, the paper is oddly written and difficult to follow.

Answer 1

We have revised the whole manuscript, simplifying the most problematic part and streamlining the sentences difficult to understand, making them more concise and direct. We also revised the English form; we are now confident that such corrections have considerably improved its readability.

Comment 2

There is no clear contributions or story of the paper i.e., what the authors carried out as compared to the literature

Answer 2

In Section I (Introduction), we better clarified the contribution of our paper by guiding the reader across the story of our work. Besides, we completely re-arranged the related work session so to highlight the contribution of our work with respect to the literature in the area (see also the reply to the 1st comment of the Associate Editor). Concerning the story of the paper, we modified the introduction as follows:

"In this context, our work concerns the algorithms for the identification of the mobile devices that can act as M^2EC. Since the opportunity to collect data of the M^2EC depends on the frequency with which devices meet, we consider a selection strategy that operates according to the social relationships among the users of the MCS platform [4, 5], and we provide a probabilistic model and a close expression that can be used to estimate the expected contribution that a M^2EC can give to the data collection capability of the MCS platform. To the best of our knowledge, in the literature there are no theoretical models able to estimate such contribution. In particular, this paper presents a probabilistic model that, with a closed-form expression, can determine the number of M^2ECs to be selected in the plethora of MCS devices. The fine tuning of the model includes few parameters that can be easily distilled from the history of the users' behaviour considering the centrality of the potential M^2ECs within their respective communities. The proposed probabilistic model has been validated through tests performed over the ParticipAct living lab dataset, a real-world MCS experiment realized with the contribution of approximately 170 university students of the Emilia Romagna region (Italy). The evidence demonstrates the effectiveness of our model in estimating the number of M^2ECs able to guarantee the same performances of a deployment realized with FMECs as in a standard MEC deployment."

Comment 3

Moreover, the related work is introduced in a separate section and there is no clear system model

Answer 3

As introduced in the response to comment 1 of the Associate Editor and to the previous comment from Reviewer 1, we have reorganized the related work section into two subsections, II-A and II-B, also in accordance with the reply to comment 1 of the Associate Editor. Moreover, we have rephrased the entire introductive period of section III-A, so to clarify the system model we are dealing with. The new introduction is as follows:

"In our system model, the mobile CrowdSensing architecture comprises three tiers: a client-tier, an edge-tier, and a server cloud-tier. The client-tier includes mobile terminal nodes. The edge-tier hosts MEC edges, namely, M^2EC operating as conventional edges to serve other terminal devices. The server cloud-tier deals with aggregation, storage, remote analytics, and processing of sensed data. We assume that the CrowdSensing platform is administered by an entity (typically allocated in the cloud tier) that coordinates the clients by injecting the sensing tasks and that implements the storage and analytics on the received data. Differently than conventional architectures, in which the MEC are fixed and part of the infrastructure, in our model we assume that also mobile devices in the client tier can take the role of M^2EC and thus act as data aggregation points for other clients, on the base of localized data exchange based on short range wireless interfaces (like Bluetooth, Wi-Fi direct or LTE direct). To this purpose, the administration entity of the MCS platform periodically selects, with a suitable algorithm, the client devices more suitable to act as M^2EC."

Comment 4

The paper has very long and convoluted sentences. It is quite difficult to understand sentences that span 7-8 lines as frequently appeared in the paper.

Answer 4

We have carried out a thorough review of the text, making the longer periods more concise. As an example, below we report a couple of examples of the indicative periods that we have adjusted so to improve the understanding of the concepts presented.

In Section II the sentence "MEC provides a new ecosystem based on radio access network edges with computational and storage, aiming at supporting peripheral nodes of the network by reducing latency for mobile users, optimizing mobile backhaul and middleware layer nodes" has been rearranged as follows: "MEC provides a new ecosystem based on radio access network edges with computational and storage capabilities. The MEC model aims at supporting peripheral nodes of the network by reducing latency for mobile users, optimizing mobile backhaul and middleware layer nodes performance."

In Section III-B the sentence "This because of two reasons: (i) some communities may be too small and it may be not worth to select a M^2EC for them, and (ii) in general the communities are not disjoint, hence there may be the case in which a community is already covered by the representatives of other communities." has been rearranged in "The reason is twofold:

- Some communities may be so small to make the M^2EC selection poorly effective.
- Communities generally are not disjoint, hence there may be the case in which a community is already covered by the representatives of another one."

Comment 5

In Section IV, the title is not reasonable, and it is not clear what the authors meant by "contribution of Mobile MECs". Similarly, what is meant by "the recruitment process of the ParticipAct users" in Section V is unclear.

Answer 5

We agree with the reviewer's comment and we changed the title of Section IV in "The Analytical Model". We confidently believe that it is now more concise and understandable. Moreover, in order to clarify the meaning of what we intend for "user recruitment process", we added in Section V the following period:

"One of the main problems with MCS campaigns is related to the process of recruiting users. Recruitment means involving volunteers in data acquisition campaign. The difficulty in the recruitment process is mostly due to the scepticism of users towards the technology and their reluctance to share sensible information through their devices [49]. Besides, small afflictions such as fast battery consumption or the excessive use of computational resources due to intensive sensing activity are further aspects that limit the wide acceptance of this paradigm."

Comment 6

In Section V, what is the relevance or significance of mentioning "pre-payed data traffic plan so that to avoid students to use their personal smartphones and credit"

Answer 6

We rephrase the sentence "pre-payed data traffic plan so that to avoid students to use their personal smartphones and credit" with "economic incentives including free smartphones and flat data plans."

Comment 7

The paper has many weird expressions such as: "under this respect"

Answer 7

We have re-arranged all the unusual expressions to improve the readability of the paper, in line with the above comments 1 and 4.

Comment 8

The paper has many typos, e.g., "also sows", "dynamism"

Answer 8

We double-checked the entire manuscript fixing errors/inaccuracies such as plurals, misleading words, punctuation, and so on. We are confident that the paper writing style has been improved.

Comment 9

Most of the figures are without sub-captions or captions

Answer 9

We are concerned about a compatibility problem in the PDF, as we assure that all figures had the caption. If the problem persists, we may need technical support from the editor.

Comment 10

The way of illustrating the time period used to compute similarity and stability in point 3 of Section V is ridiculous

Answer 10

We understand the comment of the reviewer and we realized that we did not explain well enough what Figure 3 in Section V represents. The x axis does not represent the time, rather the 4 different scenarios that we considered in the evaluation. These scenarios correspond to subsets of the PartcipacAct datasets of duration 1, 2, 6 and 10 months, respectively. Table I provides all the information characterizing the 4 scenarios. We have clarified such aspect in the figure caption. Note also that we have re-arranged Fig. 3 in two figures Fig. 3 and Fig. 4 so that to improve its readability. We report below the new figures and their captions.

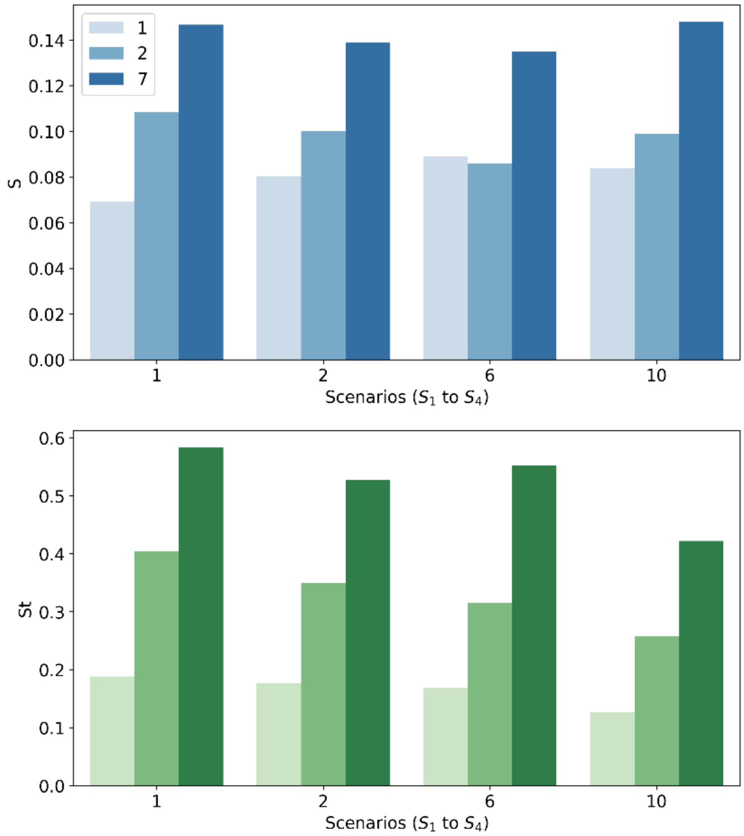

Fig. 3 Similarity (*S*) and Stability (*St*) of communities in 4 Scenarios.

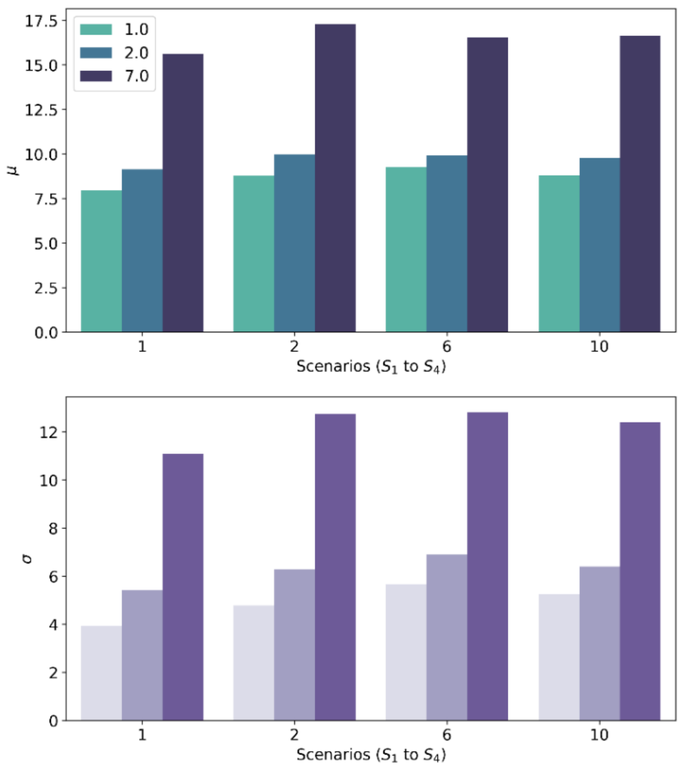

Fig. 4 Mean and std. deviation of the cardinality of the communities in 4 Scenarios.

Comment 11

In Section VII, the explanatory paragraphs are appropriate as a conclusion of the paper

Answer 11

According to the reviewer's comment, we have strongly modified the conclusive Section, adding a period summarizing the result obtained on the basis of the proposed model, and specifying in more detail the future work directions. Also, according to the comment 2 of Reviewer 2, we have condensed these observations in the conclusion section, as follows:

"In order to devise the selection algorithm and to reduce our model to a closed-form expression, we conducted a qualitative analysis of real-world mobility traces dataset, that have been also used for the validation of the model itself. Specifically, based on the analysis of the traces in the dataset we designed the algorithm in a way to select first the representative nodes of the larger communities. Secondly, the limited difference in size between the communities and the fact that the intersection among communities is not empty suggested to approximate the size of the communities with a constant and to model with Equation 1 the probability that the same device belongs to different communities. Finally, the analysis of the dataset proved also useful to determine the correction factor for the calibration of the model to achieve a good fitting (also confirmed by the K-S test) of the model with the experimental results.

These encouraging results are leading us to reflect on some future research opportunities. Since the proposed probabilistic model focuses only on M^2EC, a first step forward could be the implementation of such model into an architecture made up of both fixed and mobile MEC. Moreover, the model is sensitive to community size, and based on the community detection algorithm in use the result may vary significantly. For this reason, as a further advancement of the present work we are considering a longitudinal comparison of the main community detection algorithms designed to capture dynamics of the human mobility in MEC-based MCS systems."

R2: Review 2

Comment 1

The authors should explain clearly what kind of statistical data fitting technique they use. They should also examine the goodness of fit using methods like K-S test etc.

Answer 1

This point was also raised by the Associated Editor comment 2, who also suggested to use the Kolmogorov-Smirnov test (K-S test). We appreciated the comment from the both of the reviewers, and we decided to measure how our model fits with the observation from the ParticipAct dataset. To this purpose, we measure the K-S in order to accept or reject the *null* hypothesis that our model fits with the observation.

More specifically, we have carried out a measurement of how much our theoretical model presented in Section IV adapts to the observations of the ParticipAct dataset. To this end, we used the Kolmogorov-Smirnov test. All the tests have been performed with a critical value of $K_\alpha = 1.36$ ($i.e. \alpha = 0.05$). The results led to acceptance of the *null* hypothesis (i.e. the adaptability of the theoretical model to the real distribution) on three out of the four scenarios proposed. Let us note that the scenario in which the fitting is worse is the one less relevant for our algorithm and can be avoided by configuring the algorithm for a more frequent selection of the M^2EC. The presentation of the K-S test and its evaluation are presented as a new subsection of Section VI:

"

D. Quantitative Measurement of the Statistical Data Fitting Technique

In order to assess the adaptability of the adjusted model to the real community data set, we applied the Kolmogorov-Smirnov (K-S) test on both distributions. The K-S test is well suited to our case study because it allows to compare the shape of two sample distributions by assessing the adaptability of each other. Specifically, the result of the K-S test accepts or rejects the hypothesis of the adaptability of the real data distribution with the adjusted theoretical sample distribution. Tests have been performed on the four proposed scenarios with a critical value of $K_\alpha = 1.36$ $(i.e,\ \alpha = 0.05)$. Table II shows the *p-values* returned for each of the 4 scenarios, e.g. S1 to S4. We observe that the hypothesis of the adaptability of the theoretical model to the real data distribution can be accepted for three out of four scenarios. As reported in Table I, S2, S3, and S4 last respectively 6, 2, and 1 months. In such scenarios, mobility and sociality of users tend to remain more stable with minor fluctuations. Differently, scenario S_1 corresponds to a very long period (10 months, from January to October 2014). During such long period, users cross very different conditions: lessons, test period, lessons, Eastern break, summer holidays, test period, and back to lessons. Therefore, mobility and sociality are also affected by such variation. Moreover, during that period we observed some technical issues along the data collection that reduced drastically the amount of data collected. Figure 1 reports a graphical representation of users and traces collected during the whole 2014. As a result, the K-S test on scenario S1 returns a *p-value* lower than the others, below the level of acceptability of our hypothesis. We still include scenario S_1 for the purpose of completeness; however, our model has been designed to exploit routine of user mobility in order to select M^{2}EC. Therefore, it is a case not particularly relevant for our algorithm that can be configured to select the M^{2}EC devices more frequently.

TABLE II
RESULTS OF K-S TESTS FOR THE DIFFERENT SCENARIOS.

S_1	S_2	S_3	S_4
0.110	0.974	0.974	0.974

Comment 2

What is the basis of using qualitative analysis of real-world datasets is used to arrive at the close form for the model?

Answer 2

The analysis of the real-world dataset allowed us to identify and, to some extent, quantify four features that led us in the definition of the algorithm and of the model:

1) at least one community larger than others for each layer considered. This led us to model the algorithm to sort first the communities according to their size and to start selecting the representative M^{2}EC from the larger ones, in the effort of maximizing the coverage.
2) Although the communities identified in each layer have different sizes, the difference in size observed in the dataset is not high (at least respect to the whole of the dataset members). This led us to approximate the size of communities with the constant *g*.
3) The communities are usually not disjoint but have intersecting members. We modelled this fact with probability *p* in equation 1.
4) The assumptions at points 2 and 3 were key to obtain a close form for the model, that, however, is necessarily approximated and requires calibration. The analysis of the dataset proved then useful to determine a correction factor to achieve a good fitting (also confirmed by the K-S test) of the model with the experimental data.

We have summarized this all in the conclusion section, specifically we have rewritten the second paragraph of the conclusions as follows:

"In order to devise the selection algorithm and to reduce our model to a closed-form expression, we conducted a qualitative analysis of real-world mobility traces dataset, that have been also used for the validation of the model itself. Specifically, based on the analysis of the traces in the dataset we designed the algorithm in a way to select first the representative nodes of the larger communities.

Secondly, the limited difference in size between the communities and the fact that the intersection among communities is not empty suggested to approximate the size of the communities with a constant and to model with Equation 1 the probability that the same device belongs to different communities. Finally, the analysis of the dataset proved also useful to determine the correction factor for the calibration of the model to achieve a good fitting (also confirmed by the K-S test) of the model with the experimental results.

These encouraging results are leading us to reflect on some future research opportunities. Since the proposed probabilistic model focuses only on M^2EC, a first step forward could be the implementation of such model into an architecture made up of both fixed and mobile MEC. Moreover, the model is sensitive to community size, and based on the community detection algorithm in use the result may vary significantly. For this reason, as a further advancement of the present work we are considering a longitudinal comparison of the main community detection algorithms designed to capture dynamics of the human mobility in MEC-based MCS systems."

Comment 3

Is this method commonly used? Can the authors refer to literature where similar technique is followed for other applications?

Answer 3

To the best of our knowledge the proposed method based on an analytic model to select the M^2EC is unique. There are in literature user devices recruitment techniques for MCS campaigns in MEC architecture in support or to compensate for the limitations of the MEC proxies, but none of these focuses on the estimation of the number of third-party devices to be used. In order to emphasise this aspect, we have added some references on *similar* works and enriched the introduction with the following period:

"What makes our model unique is the possibility of estimating the contribution of the umpteenth M^2EC to be used in support of or in place of MEC middleware proxies. To the best of our knowledge in literature there are no theoretical models to estimate such contribution."