Action full title:

Universal, mobile-centric and opportunistic communications architecture

Action acronym:

UMOBILE

Deliverable:

D4.5."Report on Data Collection and Inference Models"

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

Project Information:

Project Full Title	Universal, mobile-centric and opportunistic communications architecture
Project Acronym	UMOBILE
Grant agreement number	645124
Call identifier	H2020-ICT-2014-1
Topic	ICT-05-2014 Smart Networks and novel Internet Architectures
Programme	EU Framework Programme for Research and Innovation HORIZON 2020
Project Coordinator	Prof. Vassilis Tsaoussidis, Athena Research Center

Deliverable Information:

Deliverable Number-title	D4.5 Report on Data Collection and Inference Models
WP Number	WP4
WP Leader	Rute C. Sofia
Task Leader	Rute C. Sofia
Authors	**SENCEPTION**: Rute Sofia, Igor dos Santos, José Soares **ATHENA**: Sotiris Diamantopoulos, Christos-Alexandros Sarros, Dimitris Vardalis, Vassilis Tsaoussidis **AFA**: Angela d'Angelo
Contact	Senception2013 at gmail dot com
Due date	30.09.2017
Actual date of submission	30.09.2017

Dissemination Level:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

PU	Public	X
CO	Confidential, only for members of the consortium (including the Commission Services)	
Cl	Classified, as referred to in Commission Decision 2001/844/EC	

Document History:

Version	Date	Description
0.8	07/09/2017	Table of contents and first draft
0.9	20/09/2017	Feedback by partners
1.0	28/09/2017	Consortium release
1.1	17/07/2018	Revision of formulas in accordance to caveats

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

Executive Summary

Background: This report is written in the framework of the UMOBILE project WP4 “Services enablement”, task 4.2 “Data Collection and Contextual Inference”. The task dealt with data mining in the networking context. Specifically, the task has worked upon i) devising mechanisms and tools that assist in the collection of data from sensors and other sources; ii) working on contextual awareness aspects relevant to networking in general, and to UMOBILE in particular; iii) the dissemination of the processed data to other UMOBILE services and modules, to assist in a future optimization of the network operation, based on contextualization.

A relevant aspect worked in this task is the notion of usage contextualization and service personalization. A second relevant aspect concerns the capability to infer roaming behaviour in a way that keeps anonymity and privacy of the user as well as of the device. A third relevant aspect of the task was to understand how to assist UMOBILE, namely, which modules could benefit of such contextualization operation, and what/how to integrate this operation in the UMOBILE architecture. A fourth aspect worked upon in this task concerned deriving relevant guidelines for the community, based on data collection and experimentation.

For the purpose of contextualization, the task started with an analysis of UMOBILE requirements and how to best fit such requirements when considering data capture on the network. Operationally, the task started with a model derived from the Senception’s product *PerSense ™*. Such model has been incorporated into the *PerSense Mobile Light (PML)* middleware, a first tool to assist in data capture and understanding sensing limitations. Based on such product, and derived from an analysis of the UMOBILE use-cases, as well as derived from an analysis of project requirements, the task then proceeded to propose the UMOBILE *Contextual Manager* module, which is specified in this deliverable and which shall be available until the end of the project as open-source software, technology readiness level 6, in the context of the UMOBILE proof-of-concept being defined in WP5 (month 36).

This deliverable has the following goals: i) to introduce related literature concerning contextualization aspects in networking; ii) to describe the relevancy of such contextualization in networking in general as well as in UMOBILE; iii) to provide the specification of the contextual manager as our proposal for a flexible networking framework for contextual awareness in NDN/ICN; iv) to describe experiments carried out, detailing results obtained and the relevancy in the context of networking.

Table of contents

List of Definitions 6

1. Introduction 8

2. Contributions Towards Related Work 9

2.1. Network Contextualization 10

2.1. Network Contextualization in NDN 11

2.2. Interdisciplinary Approaches to Context-Awareness 11

2.3. Contextualization in UMOBILE 13

3. Developed Tools 14

3.1. PML 14

3.2. The Contextual Manager Module 15

3.2.1. CM Data Capture 16

3.2.2. Storage Module 18

3.2.3. Data Inference Module 19

3.2.4. Inference module, utility functions C, U, and I 21

3.2.5. Interfaces 23

4. Experimentation 24

4.1. Experiment I – Human Wireless Roaming Habits Study 25

4.1.1. Scenario A: Users with Strong Daily Routine Similarity 25

4.2. Scenario C: Variable Daily Similarity 30

4.2.1. Time Characterization 30

4.3. Experiment II – Social Interaction with Children 32

5. Conclusions 35

6. References 36

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

List of Definitions

Term	Meaning
AP	Networking device, software and/or hardware based, that resides in the Customer Premises. An access point connects users to other users within the network and also can serve as the point of interconnection between a WLAN and a fixed network.
BT	Bluetooth is a wireless for short-range communication.
CM	Contextual manager, the module in UMOBILE responsible for providing measures of availability, popularity of nodes, as well as similarity of nodes, to other UMOBILE modules.
Data packet	In NDN, once the Interest reaches a node that has the requested data, the node will return a Data packet that contains both the name and the content, together with a signature by the producer's key which binds the two. This Data packet follows in reverse the path taken by the Interest to get back to the requesting consumer.
DTN	Delay Tolerant Networking (DTN) supports interoperability of other networks by accommodating long disruptions and delays between and within those networks. DTN operates in a store-and-forward fashion where intermediate node can temporarily keep the messages and opportunistically forward them to the next hop. This inherently deals with temporary disruptions and allows connecting nodes that would otherwise be disconnected in space at any point in time by exploiting time-space paths.
EC	European Commission
E2E	End-to-end principle of the internet, which states that the edges of the system perform the complex operations, while the network simply transports information.
ICN	*Information-Centric Networking (ICN)* supports efficient delivery of both content and services by identifying information by name rather than the actual location. This decoupling of the information from its actual location breaks the need for end-to-end connectivity thus enabling much wider flexibility for efficient content and service retrieval. ICN also inherently supports caching thus enabling much better localized communications.
Interest packet	In NDN, a consumer puts the name of a desired piece of data into an Interest packet and sends it to the network. Routers use this name to forward the Interest toward the data producer(s).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

NDN	Named Data Networking (NDN) is a Future Internet architecture that aims to transition today's host-centric network architecture into a data-centric network architecture. In particular, users will no longer need to retrieve data from a specific physical location; instead, users will be able to search for content, independent of the location where the content is stored.
Node	A wireless or wired capable device.
OS	An operating system (OS) is system software that manages computer hardware and software resources and provides common services for computer programs. The operating system is a component of the system software in a computer system. Application programs usually require an operating system to function.
Service	Service refers to a computational operation or application running on the network which can fulfil an end user's request. The services can be hosted and computed in some specific nodes such as servers or gateways. Specifically, services are normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services. For the purposes of this definition; "at a distance" means that the service is provided without the parties being simultaneously present; "by electronic means" means that the service is sent initially and received at its destination by means of electronic equipment for the processing (including digital compression) and storage of data, and entirely transmitted, conveyed and received by wire, by radio, by optical means or by other electromagnetic means; "at the individual request of a recipient of services" means that the service is provided through the transmission of data on individual request. Refer to D2.2 for further details.
Context	Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves [1].
IoT	Internet of Things, a network of cyber-physical devices as well as the communication between such devices.
OPEX	Operational Expenditure. Resources that an entity spends on a day-to-day basis to maintain a service.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

1. Introduction

UMOBILE established the main goal of developing a mobile-centric, service oriented architecture that efficiently delivers content and services to end-users. By efficiently it is meant that content/services are reliably available with the expected quality of service and despite any impairments of the communication infrastructure. UMOBILE decouples services from their origin locations, shifting the host-centric paradigm to a new paradigm, one that incorporates aspects from both information-centric and opportunistic networking with the ultimate purpose of delivering an architecture focused in: i) improving aspects of the existing infrastructure (e.g., keeping traffic local to lower delays and OPEX); ii) improving the social routine of Internet users via technology-mediated approaches; iii) extending the reach of services to areas with little or no infrastructure (e.g., remote areas, emergency situations).

A relevant aspect to be considered when pushing network services closer to the user (e.g., mobility management, local communication) concerns a better understanding over time and space of the devices behaviour (positioning; availability of resources). For instance, users in some areas may suffer from intermittent and unstable Internet connectivity not just over space, as well as over time (e.g., specific days of the week). Or, the network may experience loads at regular periods, or simply due to an unexpected event.

Contextual awareness can assist several aspects of networking, and has been the focus debate in related literature for long, and its relevancy in the context of challenged networking scenarios, such as in the *Internet of Things* (IoT) is well covered by several surveys [1]. The most recent evolution of contextualization in IoT concerns *Fog Computing* [2]; social interaction in IoT [3]; the use of contextualization to improve aspects of network operation such as routing [4].

Network contextualization is derived from network data mining and measurement. Being able to characterize roaming habits as well as to capture/measure internal usage in a way that does not endanger anonymity and data privacy is therefore one of the goals that have been set in the context of the work of task 4.2. Such characterization goes beyond the integration of movement prediction and/or anticipation mechanisms in the network operation, e.g., in routing or mobility management. In fact, the human movement and roaming behavior is becoming more relevant, and today, due to an extensive effort derived from several interdisciplinary initiatives as well as from extensive and wide traces collections, it is globally accepted that there is a connection between social behavior and a user's roaming behavior. It is the social behavior that assists in defining future user movement, both from an individual perspective, and from a group perspective.

Being capable of estimating such behavior is therefore relevant to optimize the network operation, be it from a mobility management perspective - e.g., handover optimization -; from a resource management perspective - e.g., performing a more intelligent load-balancing based on potential future moves of specific devices -; from a routing perspective - e.g., creating more robust routing mechanisms by selecting a priori paths that have a chance to be more stable in the presence of node movement.

In this document, we cover the work concerning contextualization, including data collection and inference aspects that has been developed in WP4, task 4.2. In operational terms, the focus is on the following aspects:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

- A better understanding on how data mining can assist the overall network operation.
- Assist in usage and network contextualization to develop new types of services/applications.
- Provide results of experiments carried out so far.
- Specify the contextual manager of UMOBILE, and its interfaces in the context of the UMOBILE architecture.

The remainder document is organized as follows. Section 2 covers background on network contextualization and how this task worked upon such background. Section 3 describes tools developed in the context of the task, namely, PML, as well as the specification of the open-source Contextual Manager module. The Contextual Manager is under development (WP5) and the code (technology readiness level 6) will be made available until the end of the project. Section 4 describes the experiments carried out so far. We conclude the deliverable in section 5.

2. Contributions Towards Related Work

Context-awareness has been a field of extensive work throughout the last decades, having been first described as a computing model for user interaction with multiple mobile and stationary devices with the capability of adapting to location of use, to a collection of nearby people and objects, as well as to changes that such objects may attain over time [6]. Enabling devices as well as applications and operating systems to adjust to surrounding conditions has been the main definition of context in regards to interaction between users and devices. Chen and Kotz examined context-aware systems and applications, types of context used and models of context information, systems that support collecting and disseminating context and applications that adapt to changing context.

In this project, the initial context notion remains: context is *"any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves"*. Nevertheless, due to the advent of pervasive wireless/mobile technology and to the increasing number of sensors around, context-awareness should be integrated into both services and network, as we shall explain in this section.

In the realm of applications, context-awareness has been dealt with in an *active* manner (context that is immediately presented to the user) or *passive* manner (context is stored and later retrieved). Active context-awareness required more infrastructure support.

With the technology evolution as well as with new computing and networking paradigms such as IoT, research has been extensively dealing with context-aware computing solutions derived from a (limited) number of sensors until the most recent advent of Big Data, where context-awareness becomes more critical, in assisting the decision on which categories of data to process and when, among other features.

In pervasive wireless computing, more relevant that Big Data requirements is to understand if context derived from "small data" can assist in making better decisions in what concerns the network operation;

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

applications' adaptability; data dissemination, thus resulting in better *Quality of Experience (QoE)*. For a concrete survey on the evolution of context-awareness, Perera et al. provide a thorough overview on the evolution of context-awareness for IoT, including projects until 2011 [8].

In IoT, context-awareness is applied for multiple purposes. For instance, it may assist in a more intelligent selection of sensors [10] [13]. Or, it may be relevant to assist in service selection [14].

More relevant, however, is the research that is being developed and that assists in pushing the network operation closer to the end-user, in an attempt to simplify the network operation and to improve QoE/QoS [12], as explained next.

2.1. Network Contextualization

New paradigms in networking have been trying to overcome issues concerning challenged network environments, e.g., due to delays (DTNs) as well as due to topological variability derived from devices being carried around by people relying on different notions of context-awareness to improve the network operation.

The HAGGLE project [15] focused on overcoming issues faced by opportunistic wireless networks (*Pocket Switched Networks*) due to the end-to-end internet design. HAGGLE relied on contextual awareness, among other new features, to assist in bringing networking functions closer to the user, thus resulting in better traffic locality. Context in HAGGLE has been identified with social habits and encounters between individual users, having been applied in naming (identification of data instead of host reachability); local data dissemination derived from frequency of encounters [16].

The ULOOP project [18] focused on improving conditions for *user-centric networks* to emerge faster. User-centric networks are in essence opportunistic wireless networks where internet end-users own (in addition to carrying) personal devices that become networking devices (that perform networking functions, such as mobility management, or routing). In ULOOP, network contextualization has been applied to assist mobility management in terms of handover optimization [19][20]. Here, context concerned data that assisted in understanding roaming habits and preferences of individual users in wireless networks, towards visited networks, and that allowed the network to make a decision i) on how to perform a handover, based on ranking of such visited networks [21]; ii) on which mobility management anchor point to delegate the function of handing over.

Mobility modeling, and trajectory prediction is another networking area where context-awareness has become increasingly relevant [22][23], being context often denoted with a time characterization (e.g., patterns of encounters). Simplistic modeling of social behavior lead *to social mobility models* [23][24], of which the *Community Based Model* (CMM) [27] or the *Sociological Interaction Mobility for Population Simulation (SIMPS) [28]* model are the most relevant models in terms of integrating a *social attractiveness* perspective into the modeling of movement. Properties integrated into these models were derived from traces' observations and therefore, context here concerns roaming habits of users over time and space.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

2.1. Network Contextualization in NDN

A fundamental difference between information-centric networks such as NDN and IP networks is that in NDN forwarding and routing are decoupled. While forwarding can detect and recover from link failures independent from the routing information, the existence of a routing plane helps bootstrapping adaptive forwarding and handling link recovery. On the other hand, NDN routing protocols may benefit from the existence of an adaptive forwarding plane due to the relaxed requirement on timely detection of failures and convergence delay which may allow the usage of routing in more dynamic networking scenarios. Hence, in NDN, network contextualization is relevant not just to assist in interface selection; it is relevant to assist in the development of forwarding strategies, as well as routing better suited for information-centric environments where there is high topological variability (such as in opportunistic scenarios).

A second area where context-awareness becomes more relevant is in data dissemination. In this context, context-awareness integrating social features has been applied by several authors in the development of social-aware opportunistic mechanisms for data dissemination in the context of opportunistic routing, as in the case of the PodNet architecture [30], where users advertise the data objects that they have interest in. When two nodes meet they decide whether or not to exchange data based on the information gathered in terms of categories of interests. Contentplace [29] builds upon this notion, adding the novelty of exchanging short summaries for the data objects they are carrying, thus contributing a decentralized dissemination solution.

Hence, up until now, context-awareness in NDN has been applied to assist in local, direct data dissemination. However, there are aspects where it is essential, such as in naming, or routing. For that purpose, from an information-centric perspective it is relevant to consider new, interdisciplinary approaches to context-awareness, that go beyond the usual time-variant notion of context, relevant to challenge environments due to the dimension of mobility of devices.

2.2. Interdisciplinary Approaches to Context-Awareness

Interdisciplinary efforts focusing on bringing in social-aware context notions into computer science are in the rise. Their main focus is on considering the different sociological, psychological and computational factors that affect/define interaction between people. Contextual aspects such as distance, orientation towards others; density of people around us, as well as noise levels are features that can assist in better understanding the physical proximity between people, the social proximity between people, and the relation of people towards the use of spaces. Such context can today easily be collected by regular sensorial devices, carried or controlled by people. Hence, personal technology can be relied upon to capture indicators that can measure both *physical* and *social proximity*. By measuring proximity, one can better defined personal spaces; develop tools that can stimulate social cohesion, as well as stimulate a better relation towards spaces around us.

The discussion on guidelines to assist the detection of social interaction via sensing technologies has been the focus of work by Alvarez-Garcia et al. [31], who debated on a number of sociological markers like co-activity, proximity, speech activity, and similarity of locations visited. Then, in the context of human-computer interaction, there is an extensive line of work focusing on improving aspects such as video delivery; nearness of remote staff; direction detecting. This line of work attempts to assist in better usability as well as in better addressing the design of technological solutions [32].

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

Quantification of social interaction via pervasive wireless sensing systems is a recent line of research, for which some aspects have already been addressed in related literature. Most work has been focused on distance estimation via sensing technology (Wi-Fi or Bluetooth) as well as the relation of distance and capacity to grasp real-world social interaction patterns [33][34][35]. In this line of work DARSIS [36] is a system that allows on-the-fly exchange of facing direction information between users and facilitates the interpersonal distance recognition process by sharing RSSI values among devices. DARSIS has been shown to have an accuracy level of 81.4% in terms of detection of interactions in a real-world environment.

In terms of available middleware that can be used to carry experiences in realistic conditions, Sociometer [35] focuses on the notion of social engagement given by proximity and conversational activities to understand how users interact. Being a mark in terms of better understanding cues concerning social context and the fact that activity recognition can be correlated with social engagement, Sociometer falls into the category of intrusive tools, requiring line-of-sight as it is infrared based. Social engagement, on the other hand, can take place independently of whether or not there are obstacles between people, and whether or not users are facing each other. Similarly to Sociometer, SociableSense [36] aims at inferring individual behavior in the context of office environments. SocialSense is based on a smartphone platform, thus being less intrusive. SociableSense adaptively controls the sampling rate of accelerometer, Bluetooth, and microphone sensors in order to estimate the user's sociability, and strength of relationship with colleagues. Neverthless, SociableSense brings in some privacy issues and increases the dependency over third party systems, by implementing a computation distribution scheme that dynamically decides where to perform the computation of classification tasks, i.e., locally or on remote servers including cloud services.

NSense [37] is open-source non-intrusive middleware that characterizes the user's personal space and social context, by capturing different physical and social proximity indicators. Aiming to reduce outside dependencies and to protect the user's privacy, NSense performs local inference only (in the smartphone). In order to balance energy and accuracy, NSense makes use of adaptive schemes while controlling the sampling rate of accelerometer, bluetooth, Wi-Fi direct, and microphone sensors. NSense is currently being used in studies concerning social proximity.

The MTracker [21] is middleware developed in the context of the ULOOP project that performs passive ranking of visited wireless networks, based on the user's social habits, taking into consideration aspects such as the frequency of visits to a network; duration of such visits; etc, with the main goal of improving the network operation in terms of mobility management.

In the context of pervasive sensing, indicators of proximity such as distance or location, as well as individual motion are usually considered to characterize physical proximity. Other relevant aspects that can be considered are measures of the cost between different devices based on their interaction over time and space (social strength); surrounding sound level; similarities in motion or in mobility.

The context-awareness work developed in UMOBILE follows the line of interdisciplinary work, considering that the availability of pervasive sensing technology brings in the possibility to further explore interaction among devices not only in sociological terms or in terms of human-computer interaction, but in fact in the context of opportunistic and direct data transmission between devices. By exploring such interaction and by better classifying/correlating user behavior not only in terms of roaming (movement) but also in terms of interaction (e.g. crowd density in time and space) the overall network operation can be better suited to support challenged environments, such as the ones that UMOBILE is being devised to tackle.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

2.3. Contextualization in UMOBILE

In the UMOBILE project context-awareness follows the line of interdisciplinary work described on section 2.2. , with the main goal of providing the network operation as well as service controllers with measures of: i) node and link availability; ii) node and link popularity (betweeness); iii) node and link similarity.

For that, we consider a specific context (control) plane, where context follows the original definition [6] presented in section 2., and integrates three different categories. The context can be related to the usage, user or the network context. In *usage context*, the context plane considers time and space characterization of device and services (e.g., resources such as CPU or energy; categories of apps). In *user context*, the context plane integrates a time and space characterization of individual user roaming behavior (habits). While in network context, the context plane considers a time and space characterization of local networking conditions, i.e., a device's neighborhood and its relations towards that neighborhood, over time and space.

The user is seen as a carrier of a mobile object. Its context is captured non-intrusively via local connectivity (external) as well as on device usage (internal). By non-intrusive it is meant that this service takes advantage of the natural networking footprint that is overhead by devices, be it via Wi-Fi, Bluetooth, as well as any other means (e.g. LTE Direct). Our current implementation efforts are focused on short-range wireless in the form of Wi-Fi and Wi-Fi Direct.

The context plane takes care of the collection, storage, and resolution of the context data. Data collection (capture) is performed seamlessly and directly via the usual wireless and mobile interfaces as well as via native applications for which the user configures interests or other type of personal indicator preferences. For instance, an application can request a one-time configuration of categories of interests such as *music*, *food*, etc. Such meta-data is passed to the contextual manager, associated to the device identifier (e.g., UUID). Metrics derived from such contextualization are then passed, upon demand or periodically, to other planes, such as the routing plane. Storage is provided locally, on the device only. Resolution concerns utility functions that the CM integrates, to compute weights that can characterize the node and link measures mentioned above: betweeness, availability, and similarity.

The context plane in UMOBILE is being implemented as a specific software module named *Contextual Manager, CM,* which in essence is a customer premises' (background) service. The CM seamlessly captures wireless data to characterize a device's affinity network (roaming patterns and peers over time and space) as well as to characterize the device's usage habits and interests (internal device information). For the sake of the proof-of-concept under development in UMOBILE in Work Package 5 (WP5), the CM resides on an end-user device, as illustrated in Figure 1. Despite being an autonomous service, it is being devised to integrate the UMOBILE End-User Service (EUS). The CM interacts with other UMOBILE modules (e.g. routing, naming, or services) via the provisioning of specific utility functions that provide indicators (e.g. routing costs and/or routing utility functions) of the social behavior of users to assist in more efficient data dissemination. Detailed information on the CM is provided in section 3.2 (Developed Tools).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

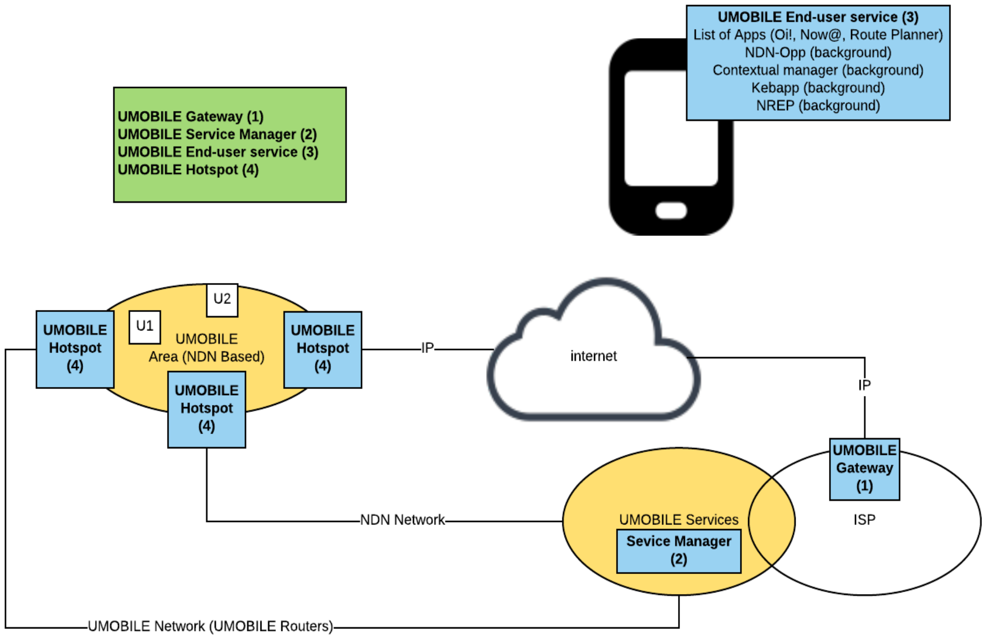


Figure 1: UMOBILE elements and whereabouts of the Contextual Manager.

3. Developed Tools

This section describes the tools developed in task 4.2 in the context of data collection and inference measures. The two tools are PerSense Mobile Light (PML) [41] and the CM.

3.1. PML

PML is a light version tool of the PerSense ™ product line of Senception. This product line is a personal platform for interaction and communication with two main features: i) stimulation of interaction via learning and inference of daily routine context in a opportunistic way via wireless and mobile networks ("How was your day?") and notifications to circles about such context ("let's share!"); ii) secure communication anywhere, anytime (instant messaging, video calls).

This light version aims solely at assisting researchers in capturing the natural networking footprint left around by devices, to assist in inference of roaming habits, and thus to assist in the network operation. Released in the project in May 2015 under LGPLv3.0, PML captures information concerning a user's affinity network (contacts derived from Wi-Fi Direct and Bluetooth) as well as concerning roaming habits, over time and space (Wi-Fi).

The tool has been developed to assist the research community in gathering meaningful traces and develop scientific studies, by reusing the collected traces. Freely available for research purposes, PML can be extended upon request, and is one of the tools that shall be available via the UMOBILE Lab.

PML stores all data locally on an SQLite database, for the period of one week, running in background. Additionally, each day the tool generates three different traceset reports automatically (and statically) at

23h59 minutes. In June 2016, version 2.0 has been released, allowing, among others, for users to get three reports in csv format: i) roaming diary report (waypoints based on Access Points crossed); ii) affinity network report (peers); iii) visited networks of the device (Access Points to which the user connected to). The reports can be checked via File Manager (stored as a compressed zip file), as well as sent by e-mail via the PML menu.

Each row in the **roaming diary** report has the following fields: *id*, *bssid*, *dayoftheweek*, *ssid*, *attractiveness*, *dateTime*, *latitude*, *longitude*. id represents the sequential identifier of the AP waypoint crossed; ssid and bssid identify the AP; dayoftheweek is an integer corresponding to the day of the week, starting by Sunday as 1, and ending with Saturday (7). Attractiveness is a binary field stating whether or not the device connected to the respective AP: if connected, attractiveness is set to 1; 0 otherwise. dateTime provides the day and time when the device entered the range of the AP. Latitude and longitude provide the coordinates of the device.

The **visited networks' report** has the following row format: *id; ssid; bssid; timeon; timeout; dayoftheweek; hour*. id represents the sequential identifier of the AP waypoint crossed; ssid and bssid identify the AP; timeon and timeout correspond to timestamps when the device enters the range of an AP, and when it leaves such range. dayoftheweek is an integer corresponding to the day of the week, starting by Sunday as 1, and ending with Saturday (7). Hour corresponds to the 24-hour timeslot of the day.

The ***affinity network report*** provides a list of neighbors over time (affinity network). Each row has the following format: sequential identifier (id); identifier of the device (uuid); MAC address (MAC); date and time when the peer was last encountered (dateTime); GPS coordinates for the device.

PML is set to provide high accuracy in terms of location, but works well if location services are set to low accuracy in order to spare battery.

3.2. The Contextual Manager Module

The UMOBILE CM is a UMOBILE service that runs in background on end-user devices and that can be easily adapted to access points. Its ultimate purpose is to provide other UMOBILE modules with contextual awareness derived from: i) internal device usage; ii) external applications; iii) available network sensors. As described in deliverable D3.1, D3.3, as well as D5.1 and D5.3, the CM performs contextualization derived from data that is either directly captured via multiple sensors (currently, Bluetooth and Wi-Fi interfaces) as well as via external sensing applications, such as PML.

A high-level perspective of the contextual manager software architecture is illustrated in Figure 2, where the color code reflects the current CM implementation status[1]. The CM architecture integrates three main modules: capture, storage, and inference. The CM Service runs in background and is the component responsible for initiating all of the software modules as well as all configured networking interfaces. It

[1] The CM is being implemented in the context of WP5 and its final version shall be released in month 36. The development of the CM can be followed via the Senception's GitLab account: https://gitlab.com/UMOBILESenception/ContextualManager.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

ensures the modularity necessary for the easy plug-in of future sensors and classification modules. The CM service is also responsible for controlling data access in the local SQLite database, as well.

In terms of external interfaces, the CM considers an on-demand as well as a periodic delivery of computed weights to routing (1,2), and to naming (3).

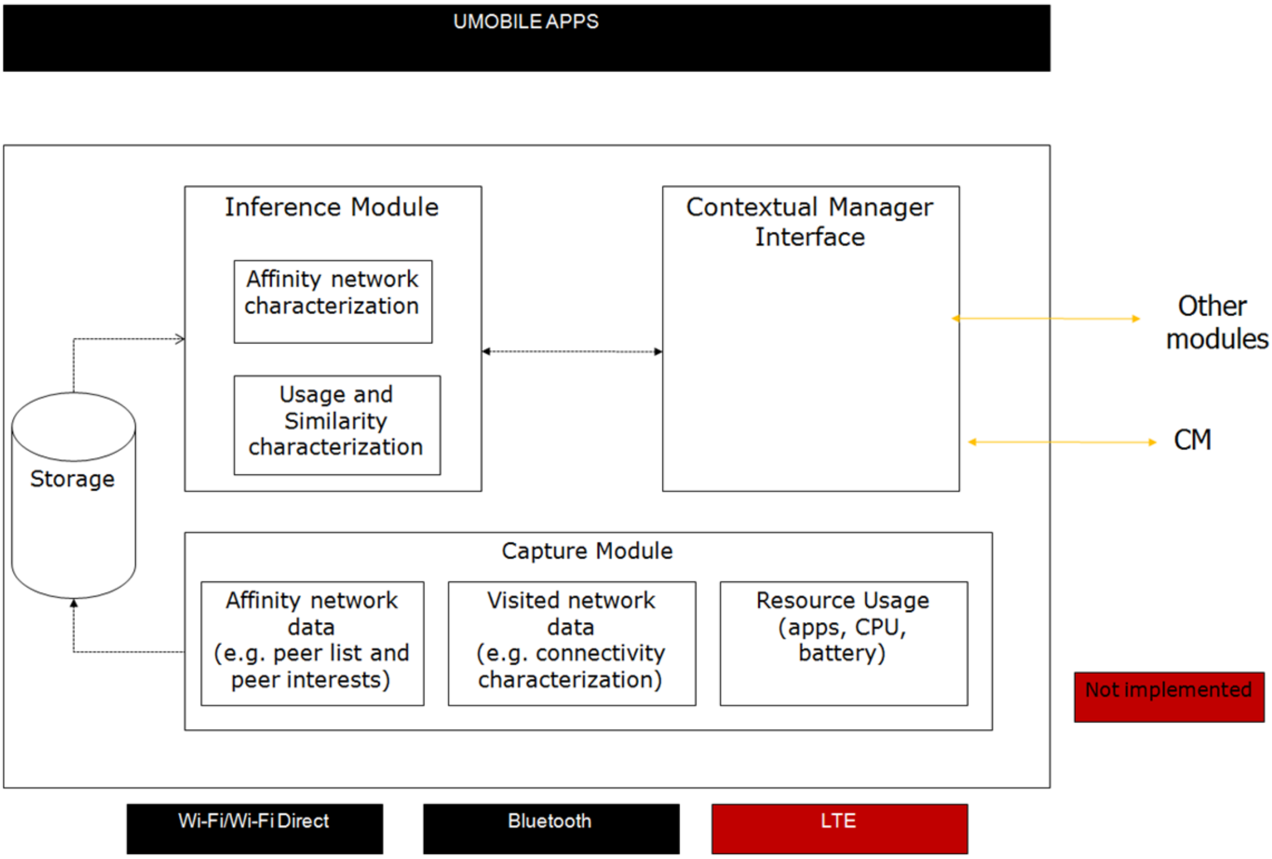


Figure 2: Contextual Manager architecture.

3.2.1. CM Data Capture

The current CM version integrates three capture pipelines as illustrated in Figure 3:

- **Visited networks' pipeline**. Data is captured via Wi-Fi and derived from regular Wi-Fi scanning thus integrating Access Points that are in the range of the device. The visited network characterization considers the following indicators:
    - **Id**. Corresponds to a sequential identifier of the network.
    - **Hashed SSID**. Hashed SSID for the access point.
    - **Hashed BSSID**. Hashed BSSID for the access point.
    - **Average visit duration**. Corresponds to the average visit duration based on an exponential moving average of all visits' duration in seconds. For APs that a device is within the range, but not authorized to access, such average visit corresponds to 0.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

   - **WeekArray**. Corresponds to an array of size 7 (days of the week), where the visits on each day of the week are registered.
   - **HourArray**. Corresponds to an array of size 24 (hours of the day) where the visits on each hour are registered. For instance, if an AP is frequently crossed on Saturdays, then HourArray[0] would hold the number of visits to that AP, independently of the day.
   - **Location**: based on geographical coordinates, obtained via fused location.

- **Affinity network**. Data is captured via Wi-Fi Direct as well as via Bluetooth, relying on regular scanning. The affinity network characterization considers the following indicators:
   - **Hashed MAC** for the peer[2].
   - **Peer device identifier** (UUID) hashed.
   - **Date and time** for the last encounter.
   - **Encounters**. Number of times a peer is seen.
   - **Location**: geographical coordinates, obtained via fused location.

- **Resource usage**. Data concerns device availability and hence, characterizes the use of internal resources over time and space (location) by capturing the following indicators:
   - **Energy consumption** over time and space.
   - **CPU usage status** over time and space.
   - **Memory usage status** over time and space.
   - **Storage usage status** over time and space.
   - **Application usage** over time and space.

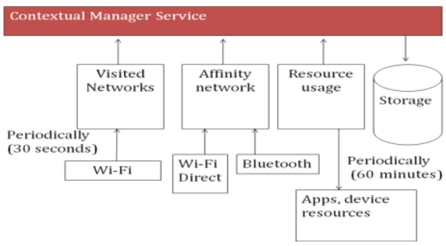


Figure 3: Data capture in the CM.

In addition to the aforementioned pipelines, the CM is expected to gather data from native applications, e.g., configured interests or other type of personal indicator preference. For instance, an application can

[2] We are considering a single hashed MAC independently of the data capture being served via Wi-Fi or Bluetooth.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

request a one-time configuration of categories of interests such as music, food, etc. Such meta-data is passed to the contextual manager, associated to the device UUID.

3.2.2. Storage Module

The data captured by the CM remains solely on the device. The database is based on SQLite, and resides on internal memory (thus just being accessible via the application). Tables 1 to 7 store Visited Networks' information per day, as illustrated in

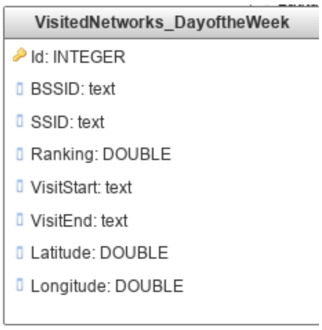


Figure 4, where ***Id*** corresponds to a sequential identifier of each visited network (for the purpose of storage only); ***BSSID*** and ***SSID*** correspond, respectively, to the hashed version of the visited network BSSID and SSID. ***Ranking*** corresponds to a measure of user preference of the visited network, based upon several parameters such as number of visits, average visit duration, as well as time gap between visits [21]. ***VisitStart*** and ***VisitEnd*** correspond, respectively, to the timestamps when the device is authorized to connect to an AP, and when the device disconnects/get disconnected from such AP. ***Latitude*** and ***longitude*** provide the geo-positioning coordinates of the device.

Tables 8 to 14 correspond to the affinity network data collection, for which the format is illustrated in Figure 5. Each entry corresponds to an encounter with a peer. ***Id*** is the entry sequential identifier; ***DeviceId*** corresponds to the hashed device identifier obtained via Bluetooth or Wi-Fi. ***MAC*** corresponds to the hashed MAC. ***Encounters*** is a variable that is incremented each time the node encounters the peer. The ***AverageEncounterDuration*** corresponds to an exponential moving average of the ***EncounterDuration***. ***Latitude*** and ***Longitude*** provide the location where the peer was last seen. **C** and **U** are measures, respectively, of the node's betweeness and availability (rf. to section 3.2.3 for inference details), given for slots of 1 hours, collected over 24 hours.

Table 15 holds data collected concerning resource usage. Each entry corresponds to a type of resource. Currently, the CM considers 4 types of resources: ***Energy***, ***Storage***, ***CPU***, and ***Memory***. Data is collected by recurring to the Android library UsageStats[3]. Each entry holds the entry identifier (***Id***); ***Type***; an array of size 24 holding the average usage of the specific resource per hour; ***DayofTheWeek***, 1 (Sunday) to 7.

[3] https://developer.android.com/reference/android/app/usage/UsageStats.html.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

VisitedNetworks_DayoftheWeek
Id: INTEGER
BSSID: text
SSID: text
Ranking: DOUBLE
VisitStart: text
VisitEnd: text
Latitude: DOUBLE
Longitude: DOUBLE

Figure 4: Format for the Visited Networks' data collection (tables 1 to 7).

Affinity_DayoftheWeek
Id: INTEGER
DeviceId: text
MAC: text
Encounters: INTEGER
AverageEncounterDuration: text
Latitude: DOUBLE
Longitude: DOUBLE
C: VARCHAR(24)
U: VARCHAR(24)

Figure 5: Format for the Affinity Networks' data collection (tables 8 to 14).

ResourceUsage
Id: INTEGER
TypeofResource: text
AverageUsageHour: VARCHAR()
DayoftheWeek: INTEGER

Figure 6: Format for Resource Usage data collection (table 15).

Table 16 is used to store application usage for multiple categories of applications in a similar format to resource usage as illustrated in Figure 7. The fields in this table are: **Id** (sequential entry identifier); ***AppName***, standing for the application designation; ***AppCategory***, which corresponds to the Google commercial application category, which is provided via the application. ***AverageUsageHour*** is an array which holds the application usage for each hour period in a 24 hours period. ***DayoftheWeek*** holds a value of 1 (Sunday) to 7.

The final table, table 17, is used to store the measures derived from the data inference module, as described in the next section. Each entry holds the sequential identifier ***Id***; the time and date when the weight was stored (***dateTime***); weight **C** (node Eigencentrality); weight ***U*** (node availability)

AppUsage
Id: INTEGER
AppName: text
AppCategory: text
AverageUsageHour: VARCHAR()
DayoftheWeek: INTEGER

Figure 7: Applications Usage', database format (table 16).

Weights
Id: INTEGER
dateTime: dateTime
A: DOUBLE
U: DOUBLE
DayoftheWeek: INTEGER

Figure 8: Inference weights, database format (table 17).

3.2.3. Data Inference Module

The data inference module (rf. to Figure 9) takes care of using the different indicators stored, and combining them via different utility functions to characterize a node's affinity network (neighborhood and its variation over time and space) as well as to characterize a node's usage and to give a measure of similarity between adjacent nodes.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

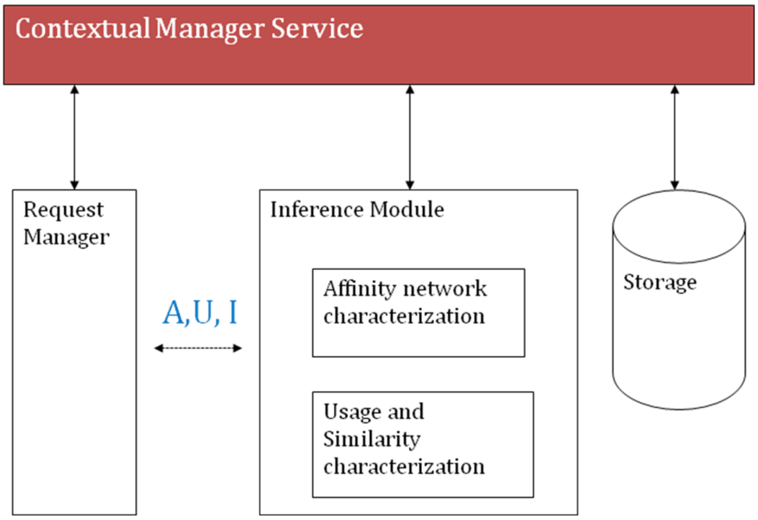


Figure 9 CM inference module.

To characterize a node's affinity network, the CM relies on the collected indicators, which can be provided raw to other modules, or combined via the available utility functions. Currently, the CM can provide directly to other modules the following list of indicators for a node i as well as for the peers of node i, for a specific instant in time, or during a specific period of time (e.g., between 16 h and 18 h, or for 1 day):

- **Peer list**. Corresponds to a list of hashed MAC addresses that identify the nodes around node I during the specified time period.
- **Resource status.** The CM provides the requested resource status (e.g. battery status, or storage status) for node I as well as for all the peers of node i during the specified time period.
- **Connectivity status.** Provides the minimum, average, and maximum values for connectivity period for each node during the specified time period.
- **Encounter duration.** Provides the minimum, average, and maximum values for encounter duration for node i towards its peers during the specified time period.
- **Node Degree.** Provides the average node degree for node I during the specified time period.
- **Visited network rank**. Provides the rank of a specific visited network for each of the peers of node i.

To characterize usage, the CM can provide directly to other modules the following indicators per day and per hour:

- **Status of resources** (e.g., 100% of battery at 10 a.m.; 10% at 11 a.m.; 80% of free storage at 10 a.m.).
- **Most used application categories**. Maximum of ten categories of applications for a node.

As mentioned before, these indicators can be passed on demand or periodically from the CM to other UMOBILE nodes. Furthermore, the CM inference module integrates different utility functions to provide other modules with weights for betweeness, usage, as well as level of similarity between different nodes, as explained next.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

3.2.4. Inference module, utility functions C, A, and I

Based upon the different collected indicators, the CM inference module periodically[4] computes three different weights and stores it in the inference table (cf. Figure 8):

- **C.** Corresponds to the affinity network level of node i and measures a node's centrality, i.e., a node's popularity.
- **A**. Corresponds to the internal usage weight of node I and measures the availability of the node.
- **I**: measures the (eigenvector) similarity between the selected resources of a node and one of its neighbors. For instance, I can provide a measure of battery similarity over time between nodes. Or, it can provide a measure of similarity between categories of applications.

3.2.4.1. C: Node Centrality

In what concerns a node centrality, there are several measures that can be considered. To select a specific measure of centrality, the following assumptions are relevant:

- The more visited networks a node has over a period of time, the more central a node is (increases the possibility for data transmission).
- The higher the number of connections a node has over a period of time, the more central a node is.
- The higher the node degree of node over a period of time, the higher is its centrality.
- The lower the distances traversed by the node are, the higher is its centrality.

Several measures of centrality could have been considered to implement the mentioned assumptions. For instance, *degree centrality* provides a measure of centrality based on the number of peers around (the more connected a node is, the higher the centrality is). *Closeness centrality* is another possible measure based on shortest-distances; however, it does not work well in disconnected networks (as may be the case in UMOBILE scenarios). *Betweeness centrality* takes into consideration the number of shortest-paths that a node holds, and for UMOBILE, this may not always be relevant, as the focus is on information-centricity.

We therefore consider as basis for centrality the *Eigenvector centrality* definition [38]. Eigenvector centrality of a node (also known as *Eigencentrality*) is a measure of that node's influence in the network. Eigenvector centrality provides a measure of the node's popularity based also on its neighbors' popularity and is derived from degree centrality. Our version of Eigencentrality follows this line of thought, but instead of considering the degree of neighboring nodes, it considers a centrality weight based on encounter duration and number of encounters for each neighbor, ***p(i)***. Equation 1 describes ***C(i)*** as well as ***p(i)***.

[4] Currently, the costs are computed every 10 minutes. The time window adjustment will be done upon validation, during the deployment of the UMOBILE proof-of-concept, WP5.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

$$C(i) = \frac{1}{\lambda} \sum_j A(j) * p(j)$$

Where

$$p(j) = \frac{encounter * average\_encounter\_duration}{d(i,j)+1}$$

$$\lambda \in [0,1]$$

$$A(j): adjancency$$

$$d(i,j) \in [0,100]$$

Equation 1: C, measure of a node;s centrality derived from Eigenvector centrality.

3.2.4.2. A: Node Availability

Node availability is derived from the node's usage rate over time and space. In usual IT terms, the metric used to measure availability concerns the percentage of time that a system is capable of serving its intended function. U is based on such notion, taking into consideration a composition of the multiple captured device resources per hour. Hence, the first step in measuring node availability concerns the computation of node resource usage ***r(i)***, given in Equation 2.

r(i) is computed based on the relative usage of each resource, namely, energy level; used memory status; used CPU status; used storage status. The lower each of these resources, the higher the resource usage r(i) is. The energy level status has more weight than the other components, as an energy drain causes more impact in the availability of the device. t corresponds to the sampling interval, currently set for 10 minutes. Then, on each hour, r(i) is stored based on an exponential moving average.

$$r_t(i) = \alpha * (b(i)^2 * cpu(i) * mem(i) * s(i)) + (\alpha - 1) * r_{t-1}(i)$$

where

$$b(i): energy\ available$$

$$cpu(i): available\ cpu$$

$$mem(i): available\ memory$$

$$s(i): available\ storage$$

$$\alpha: smoothing\ \ factor$$

Equation 2: r(i), resource status weight computation.

The availability of the node for a time period T is provided by U (cf. Equation 3) which is computed based on a specific time window T and takes into consideration all of the hourly values of r(i).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

$$U(i) = \frac{\sum_{h=1}^{t} r(i)}{T}$$

$where:$

$r(i): resource\ usage\ weight$

$h: hourly\ time\ slots$

$T:\ time\ period\ in\ hours$

Equation 3: U, measure of a node's availability.

3.2.4.3. I: Similarity

Similarity is the third measure provided by the CM to other UMOBILE modules, and it is a link measure, i.e., it provides a correlation cost between a node and its peers based on cosine similarity. Similarity is associated to a specific set of resources, e.g. visited networks; affinity network; app category, resource usage.

An example is as follows. Let us assume that node *i* has a set of application preferences corresponding to Music, Art and represented by the set A=[1,1,0,0,0,0,0,0,0,0,0,0] While node j has as main preferences Music, Literature, represented by the set B=[1,0,1,0,0,0,0,0,0,0,0]. The similarity weight **I** is computed as in Equation 4 and corresponds to 0.5.

$$I = \frac{\sum_{i=1}^{n} A(i) * B(i)}{\sqrt{\sum A(i)^2} * \sqrt{\sum B(i)^2}}, I \in [-1,1]$$

Equation 4: Similarity weight I based on cosine similarity.

This value can therefore be provided based on the different resources. To provide an example let us assume that we want to understand the similarity of the **U** weight between node *i* and its peer *j* over a specific period of time, e.g. 3 hours, where r(i)=[0.1, 0.3, 0.7] and r(j)=[0.7,0.5,0.7]. The similarity for resource battery between nodes *i* and *j* based on I would correspond to 0.378. Such analysis can assist in a better selection of peers for data transmission, for instance.

3.2.5. Interfaces

The CM interfaces concern other UMOBILE modules which may reside on the same node or not. For the sake of explanation we assume, however, that all modules are placed in the same end-user device.

The interfaces currently being deployed are:

- Routing interface (rf. to Figure 1, arrow 1 and 3).
  - The routing module can perform specific requests to the contextual manager, to get i) a set of indicators for a specific time window; ii) inferred values C, U, and I.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

- Periodically (currently set to 10 minutes) the CM sends values C,U, I, and the set of peers with the respective affinity network indicators.

- Interface to NREP. The interface between the Contextual Manager and NREP is bi-directional and has two different operational states (rf. to Deliverable D4.3):
  - NREP can perform a request to the Contextual manager to get a set of priorities (indicators) for a specific time window.
  - NREP can get (periodic) notifications for specific sets of indicators.

3.2.5.1. Example, Routing

In this section an example for a routing interface based on context similarity is provided. The definition of the routing computation as well as forwarding strategy is out-of-the-scope of this deliverable, but it can be followed in UMOBILE publications under development as well as in the deliverables of WP3 and WP5.

One important feature of NDN lies in the roles that are envisioned for routing and forwarding. Unlike the classical IP world where the forwarding is fixed, NDN offers the possibilities to implement strategies that can go beyond the simple Longest-Match prefix rule. The CM routing interface provides the routing module with period information concerning a node's centrality (C) and a node's availability as well as with the I similarity weight between peers. It is then up to the routing module to perform decisions on how to forward data.

Hence, the CM provides the routing module with interface ranking. For each name prefix in a RIB, a node therefore ranks the Faces based on C,U. The routing considers an additional measure D - time lapse between forwarding an Interest packet and getting data. D is therefore an indicator of distance (time) to the data (to the nearest copy). The routing then forwards Interest packets (control packets) through the Faces holding the highest ranks.

The Face ranking therefore increases with C and with U, and decreases with D. For the routing, C and U provide an indication of potential success in data transmission, while D reflects transmission quality.

4. Experimentation

Throughout this section, a visited or encountered wireless network corresponds to a wireless hotspot, and identified by a wireless Access Point (AP) SSID. While a connected wireless network or AP corresponds to a network that the user crosses and attaches to (uses the Internet).

The distance between visited wireless networks has been computed based on the latitude and longitude for two points identified as (lat1,long1) and (lat2, long2) by relying on the haversine formula given in Equation 5,

$$1000*(ArcCos(Cos(R(90-lat1))*Cos(R(90-lat2))+Sin(R(90-lat1)*SIN(R(90-lat2)*Cos(R(long2-long1)))*6371$$

Equation 5: distance computation based on geographical coordinates.

where:

- ArcCos(x) gives the arc cosine of x (argument in radians).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

- Cos(x) gives the cosine of x in (argument in radians).
- Sin(x) gives the size of x (argument in radians)
- 6371 corresponds to R, the earth's radius in meters
- R(x) corresponds to x in radians.

Roaming time corresponds to the time period during a day (over 24 hours) when the device first gets connected to a wireless network, until the device shuts down its wireless operation for the day. *Connectivity time* corresponds to the time period during a day (over 24 hours) when the device actually engages in Internet access via a wireless network. Therefore, if during a day a device connects to two different APs for periods of 1 hour, then the connectivity time for that day is of 2 hours.

4.1. Experiment I – Human Wireless Roaming Habits Study

This first experiment aimed at a better understanding of whether or not wireless tracking would be enough to characterize one's roaming behavior statistically; up to which point users that share social daily affinities (e.g. similar social routines) would exhibit similar roaming based on wireless; if the roaming routine could be characterized over time and space with some granularity (e.g. days, hours minutes). To look for answers to the mentioned questions, data was captured via seven different devices having PML installed and running for 24-hour periods, between the period of November 13th 2015 and December 18th 2015, in Lisbon, Portugal. Some of the carriers of these devices shared affiliation (4 in 7). The extracted traces [40] hold information such as encountered and accessed wireless hotspots (SSIDs and BSSIDs); duration of visits; geo-positioning; whether or not the device was connected to a specific hotspot, for how long and for how many times. Our study has considered time and spatial characterization. A longer version description is available via an initial technical report [5][5].

Roaming time corresponds to the time period during a day (over 24 hours) when the device first gets connected to a wireless network, until the device shuts down its wireless operation for the day. Connectivity time corresponds to the time period during a day (over 24 hours) when the device actually engages in Internet access via a wireless network. Therefore, if during a day a device connects to two different APs for periods of 1 hour, then the connectivity time for that day is of 2 hours.

4.1.1. Scenario A: Users with Strong Daily Routine Similarity

A first set of experiments in this study relied on two devices carried by users that share in their regular wireless routines some visited wireless hotspots and social routine interests (e.g. share affiliation; go to the same bistro), having carried the devices around for a period of 4 weeks in November and December 2015. Both devices were Android 5.1 smartphones with PML installed. This experiment was set to understand whether or not there is a roaming pattern for users based on wireless visited APs, and to characterize such roaming patterns in statistical terms, both in time and in space.

[5] Publication under submission, 2017.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

4.1.1.1. Time Characterization

Figure 10 depicts results obtained concerning the daily pattern of roaming time vs connectivity time, where the X-axis considers the 28 days observed, and the Y-axis corresponds to 1-hour periods over 24 hours. In terms of roaming times, both users exhibit a very close pattern where in average the roaming time is between 15 hours and 18 hours. This means that out of 24 hours, the devices have wireless coverage for around 15 hours. This period is significantly higher than the currently period of 8 hours often applied in networking simulations. Furthermore, there are at least three distinguishable usage peaks which should be considered and which seem to be tied to the daily habits of people. This is relevant not only in terms of modeling; it also shows that tracking and monitoring via Wi-Fi is today achievable with a reasonable level of accuracy.

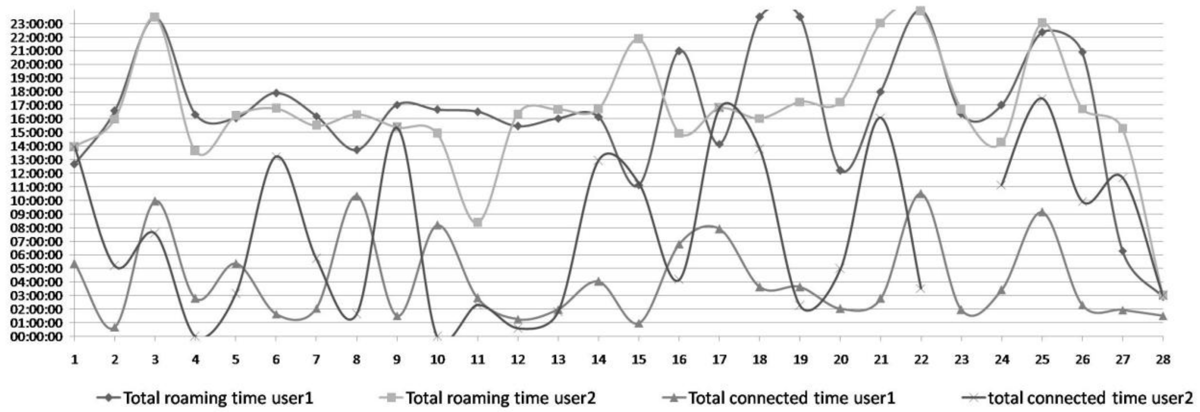


Figure 10: Roaming time characterization.

4.1.1.2. Spatial Characterization

The spatial characterization of human wireless roaming routine embodies multiple aspects, of which we have considered two: i) distance traversed between crossed wireless networks (average, minimum, and maximum); ii) encountered and connected hotspots. The aim is to understand whether or not tracking is achievable also in terms of the spatial routine of users, and up to which point users visit new networks, or are they regularly hopping between the same wireless networks.

Figure 11 characterizes maximum and average distance findings for user1 and user2, in terms of distance between two consecutive encountered wireless hotspots. The X-axis corresponds to the 28 days observed (day 1 being a Friday), and the Y-axis corresponds to distance in meters, shown in a logarithmic scale. In the figure, the minimum distance is not illustrated, as it has been found to always be 0.9 or 0 meters, possibly due to the way the Android fused location API provides information concerning overlapping APs (e.g., APs in the same building). Maximum (circa 10 kilometers) and minimum (0.9 meters) distances are similar for both users. Average path distances between two consecutive wireless hotspots are also similar. On days 9 and 10 (a weekend), user1 had the device on the same position, and hence distances between visited wireless networks were null. Days 23 and 24 (corresponding to a weekend) exhibit higher average and maximum distances, which is a consequence of larger paths being traversed (unusual patterns for both users). In terms of maximum and average distances, there is an observable similar pattern. Relevant to highlight is that on

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

weekends (days 2,3; 9,10; 16,17; 23,24) the distances are actually similar to the distances observed in working days, even though the devices connected time in weekends is lower than in weekdays.

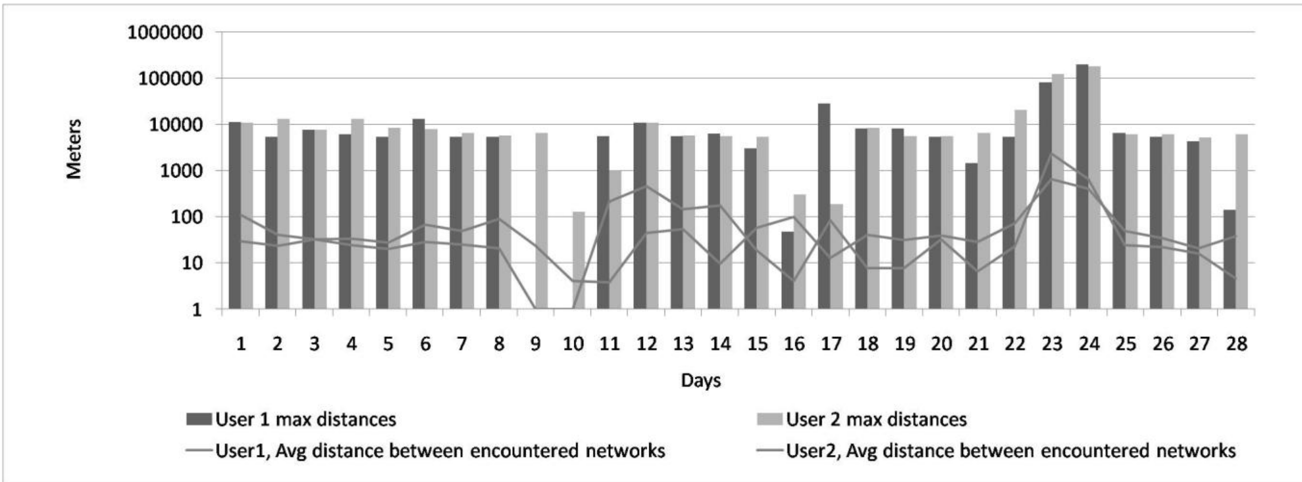
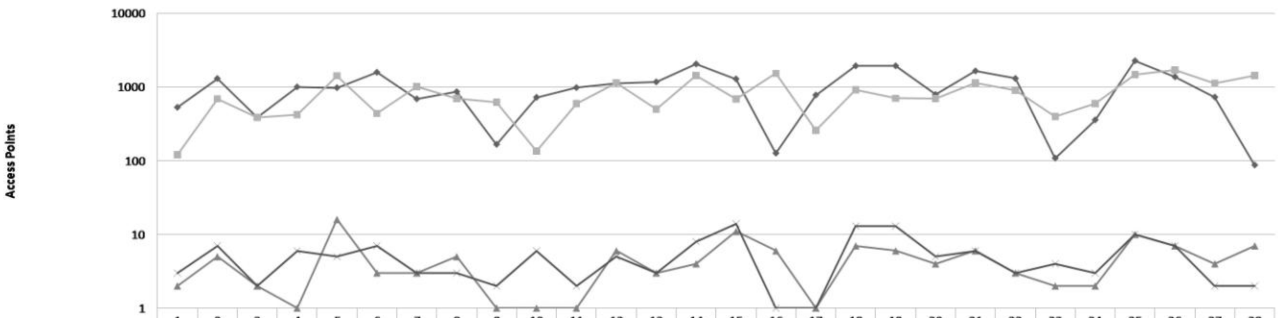


Figure 11: Distance characterization.

The total number of APs encountered during visits is similar for both users, and quite high: as shown in Figure 12, in average both users cross over 1000 wireless networks. What is interesting to reveal is that despite the density, both users exhibit patterns where they daily connect to a maximum of 6 APs, which is a huge discrepancy in comparison to the APs available.

Based on the traces obtained, we can state that this is not just a consequence of having most APs closed; in fact, this discrepancy actually relates to the daily activity of the users, as can be seen in the networking analysis in section 5.1.3. The average encountered wireless networks is quite high and exhibits a strong correlation. The sum provided for both encountered and connected wireless networks corresponds to the period of the 28 days, with an impressive 28268 networks for user1 and 23250 wireless networks for user2. As for connected networks, both users share similarity, connecting in average to 4-5 wireless networks per day. The maximum number of connected networks observed was between 14-16.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Aps encountered user1	529.0	1303.	388.0	1002.	977.0	1582.	689.0	862.0	166.0	722.0	980.0	1118.	1171.	2054.
Aps encountered user2	121.0	694.0	388.0	423.0	1424.	440.0	1019.	698.0	623.0	136.0	596.0	1148.	504.0	1436.
Aps connected user2	2.00	5.00	2.00	1.00	16.00	3.00	3.00	5.00	1.00	0.00	1.00	6.00	3.00	4.00
Aps connected user1	3.00	7.00	2.00	6.00	5.00	7.00	3.00	3.00	2.00	6.00	2.00	5.00	3.00	8.00

	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Aps encountered user1	1286	126	777	1935	1935	790	1648	1306	108	357	2269	1373	728	87
Aps encountered user2	688	1536	259	920	707	699	1141	908	400	596	1478	1711	1127	1430
Aps connected user2	11	6	1	7	6	4	6	3	2	2	10	7	4	7
Aps connected user1	14	1	1	13	13	5	6	3	4	3	10	7	2	2

Figure 12: Encountered vs. Connected APs.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

4.1.1.3. Scenario B: Weak Daily Routine Similarity

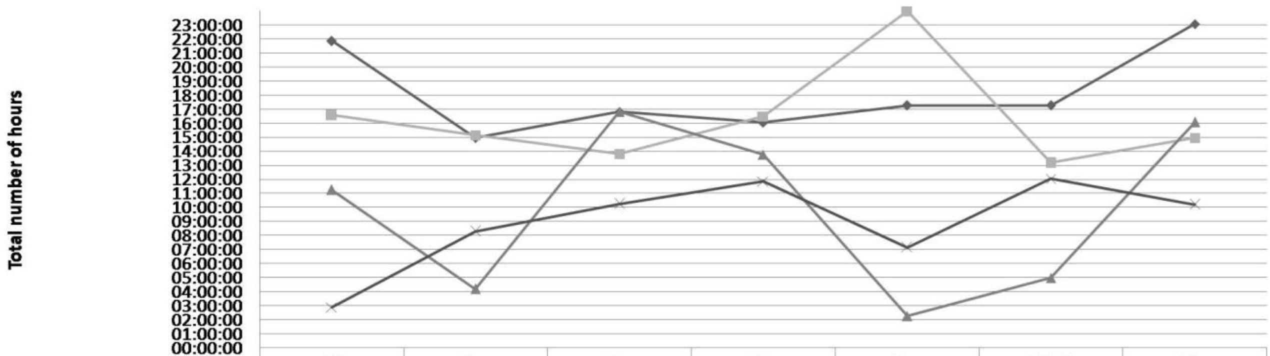
On a second set of experiments we have analyzed the behavior of two devices that have a low similarity in daily activity. The users carrying the devices are user2 (from the prior experiment) and user3. Users carrying these devices share affiliation partially, i.e., they are on the same affiliation place for circa 4 hours per day. The traces have been collected for one week, from 27.11.2015 to 03.12.2015. user2 carries an Android 5.1 smartphone device, while user3 carries an Android 5.0 device with PML installed.

4.1.1.4. Time Characterization

Figure 13 provides the daily patterns of roaming time vs connectivity time for both users. The X-axis holds the 7 days observed (day 1 corresponding to Friday 27.11.2015), and the Y-axis holds time.

In terms of roaming times, both users exhibit periods above 15 hours, meaning that out of 24 hours, the devices have wireless coverage for over 15 hours. From a pattern perspective, there is again a reasonable correlation level between the usage of both devices.

When comparing these results to the results extracted for two users that share a strong correlation in terms of daily routine (cf. Figure 10), it can be observed that all roaming times exhibit usage for over 15 hours, while connected times exhibit usage for over 8 hours per day.



	Fri	Sat	Sun	Mon	Tue	Wed	Thu
Total roaming time user2	21:50:36	14:55:58	16:47:19	16:01:34	17:14:14	17:14:14	23:02:33
Total roaming time user3	16:32:40	15:07:08	13:46:57	16:28:13	23:57:07	13:09:34	14:56:36
total connected time user2	11:13:36	4:11:04	16:47:19	13:44:16	2:15:08	4:58:21	16:04:00
total connected time user3	2:51:39	8:16:12	10:13:35	11:50:15	7:07:52	12:00:59	10:11:24

Figure 13: Roaming time characterization.

4.1.1.5. Spatial Characterization

Figure 15 characterizes the maximum and average distance findings in terms of distance between two consecutive encountered wireless hotspots, where the X-axis corresponds to the 7 days observed (day 1 being a Friday), and the Y-axis corresponds to distance in meters, shown in a logarithmic scale. While maximum (circa 10 kilometers) and minimum (circa 0.9 meters) are similar, the average path distances between two consecutive wireless hotspots are close but not always similar (e.g. Saturday and Tuesday).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

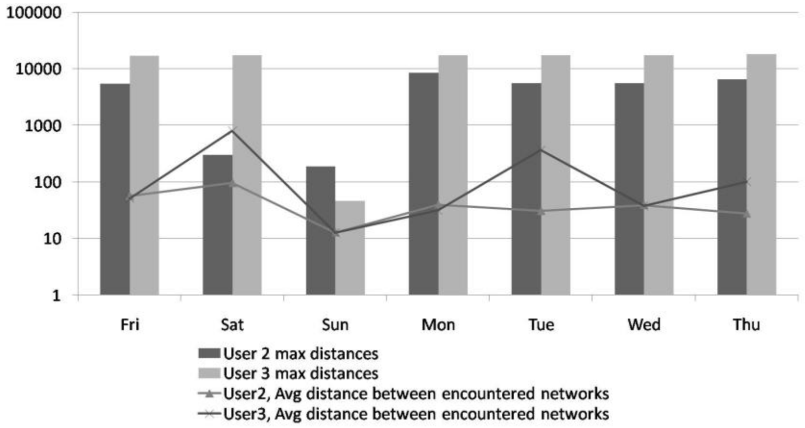


Figure 14: Distance characterization, experiment II.

Figure 15 depicts the total number of APs encountered vs. connected for both users. Such value is again significantly high, reaching over 1000 wireless networks in some days (e.g. Saturday for user2; Monday for user3). The number of connected wireless networks shows a huge discrepancy in comparison to the APs available. Similarly to what happened in the last experiment, we analyzed if this could be just a consequence of closed APs. In fact, the reason for this behavior relates with the roaming routine of the user, as we observed that users connect to the same wireless networks over time repeatedly.

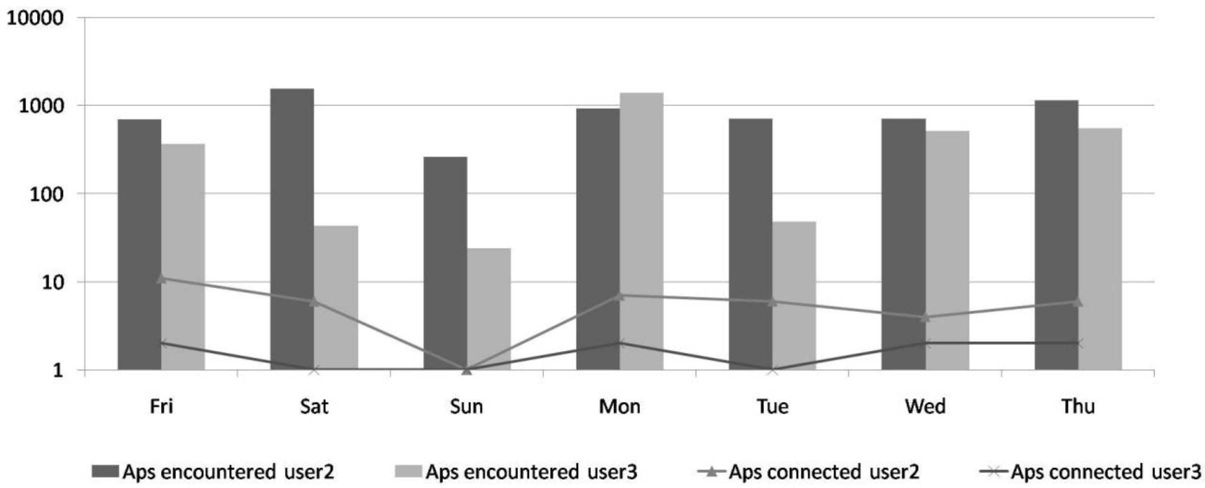


Figure 15: Encountered vs. Connected APs.

The sum provided for both encountered and connected wireless networks corresponds to the period of the 7 days for both users is still impressive. User2 crossed 5950 wireless networks, having used 41 (even though these are the same wireless networks, as the user relies on average in 5 to 6 wireless networks). While user3 crossed 2910 wireless networks, having used 11 (even though in average the user prefers 2 specific wireless networks).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

4.2. Scenario C: Variable Daily Similarity

On a third set of experiments we have considered 5 different devices having collected data for a period of 7 days, from 11.12.2015 to 18.12.2015, being users user1, user2, the users from the previous experiments. All users share affiliation partially, i.e., they are on the same premises together for around 5 hours per day. Users have been selected in terms also of their wireless roaming experience, namely: user1 and user2 are users heavily connected; user4 is an average connected user; users 6 and 7 exhibit low wireless usage.

4.2.1. Time Characterization

Table 1 shows the statistical analysis to assist in better correlating results obtained for the different users. The difference in terms of wireless usage shows that the routine of heavily connected users is in average above 15 hours, while the routine for the users (user6, user7) less engaged in wireless roaming can be as low as 2 hours per day. This is also a consequence of the fact that user6 and user7 had their devices off during weekends and is an aspect that requires more traces to allow us to better understand the time differences between low and high wireless usage.

Table 1: Roaming Time and Connectivity Time Correlation.

Roaming						**Connectivity**				
	User1	*User2*	*User4*	*User6*	*User7*	*User1*	*User2*	*User4*	*User6*	*User7*
Mean	15:39:33	15:56:52	12:08:47	7:06:14	5:45:03	4:24:52	8:55:49	2:20:04	2:19:25	2:04:36
Standard Error	3:03:37	2:36:12	3:46:39	2:45:06	2:36:31	1:25:09	1:57:28	1:29:45	1:16:12	1:01:19
Median	17:00:22	15:29:21	13:56:33	6:51:58	2:33:40	2:18:36	9:55:02	0:00:01	0:00:00	0:00:00
Variance	2:43:54	1:58:37	4:09:43	2:12:30	1:59:05	0:35:15	1:07:04	0:39:09	0:28:14	0:18:17
Standard Deviation	8:05:49	6:53:17	9:59:39	7:16:48	6:54:06	3:45:18	5:10:46	3:57:28	3:21:37	2:42:14
Range	21:02:24	20:53:06	23:50:23	16:52:23	16:46:09	10:29:28	17:29:14	9:28:00	7:22:02	6:22:56
Minimum	2:55:15	3:01:04	0:00:00	0:00:00	0:00:00	1:31:21	3:01:04	0:00:00	0:00:00	0:00:00
Maximum	23:57:39	23:54:10	23:50:23	16:52:23	16:46:09	10:29:28	17:29:14	9:28:00	7:22:02	6:22:56

4.2.1.1. Spatial Characterization

The correlation between the results observed for the average, maximum, as well as minimum distances is provided in Table 2. The average distances between wireless networks are nonetheless again quite small (hundred meters) and show that the paths traversed have a strong density, relevant in terms of wireless tracking.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

Table 2: Distance correlation.

	Average					Minimum					Maximum				
	User1	*User2*	*User4*	*User6*	*User7*	*User1*	*User2*	*User4*	*User6*	*User7*	*User1*	*User2*	*User4*	*User6*	*User7*
Mean	109.83	94.11	10.68	18.55	2.64	0.09	0.09	0.11	7.07	0.04	31623.00	32488.00	2648.30	1586.00	32.08
Standard Error	90.57	51.31	5.63	9.25	1.80	0.00	0.00	0.05	7.04	0.02	27115.79	24185.36	2409.35	753.55	21.17
Median	23.24	48.22	5.44	0.00	0.00	0.09	0.09	0.09	0.00	0.00	5389.00	6095.00	223.00	0.00	0.00
Standard Deviation	239.62	135.74	14.90	24.48	4.77	0.00	0.00	0.14	18.62	0.05	71741.64	63988.44	6374.55	1993.69	56.01
Range	648.42	379.57	36.72	55.13	13.03	0.00	0.00	0.42	49.29	0.09	194251.00	177071.00	17089.00	4124.00	153.00
Minimum	4.58	20.43	0.00	0.00	0.00	0.09	0.09	0.00	0.00	0.00	141.00	5242.00	0.00	0.00	0.00
Maximum	653.00	400.00	36.72	55.13	13.03	0.09	0.09	0.42	49.29	0.09	194251.00	177071.00	17089.00	4124.00	153.00
Sum	768.84	658.77	74.74	129.82	18.47	0.63	0.63	0.78	49.47	0.27	221361.00	227416.00	18538.09	11102.00	224.57

The greater variability is observable in the maximum distances. This seems to be a consequence of the fact that in some days some of devices seemed to be off, and not necessarily a consequence of the path diversity that the users may cross. We intend to better analyze this aspect in already ongoing research.

The total number of APs encountered during visits shows the usage difference between users per day and is given in Table 3, which provides a statistical analysis for the correlation of the obtained results.

Table 3: Visited networks correlation.

Encountered APs						Connected APs				
	User1	*User2*	*User4*	*User6*	*User7*	*User1*	*User2*	*User4*	*User6*	*User7*
Mean	972.71	1181.29	244.14	97.57	20.14	4.29	5.14	1.29	0.43	0.57
Standard Error	278.42	144.50	117.24	59.51	6.49	1.15	1.10	0.42	0.20	0.20
Median	728.00	1127.00	32.00	33.00	21.00	3.00	4.00	1.00	0.00	1.00
Variance	542606.24	146170.57	96218.48	24787.62	295.14	9.24	8.48	1.24	0.29	0.29
Standard Deviation	736.62	382.32	310.19	157.44	17.18	3.04	2.91	1.11	0.53	0.53
Range	2182.00	1115.00	723.00	438.00	47.00	10.00	10.00	3.00	1.00	1.00
Minimum	87.00	596.00	0.00	0.00	0.00	2.00	2.00	0.00	0.00	0.00
Maximum	2269.00	1711.00	723.00	438.00	47.00	10.00	10.00	3.00	1.00	1.00
Sum	6809.00	8269.00	1709.00	683.00	141.00	30.00	36.00	9.00	3.00	4.00

The number of connected wireless networks shows again a strong discrepancy in comparison to the APs available. For this experiment this is both a consequence of the daily routine of users, as well as a consequence of the fact that some readings were not obtained.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

The sum provided for both encountered and connected wireless networks corresponds to the period of the 7 days for all users is still significant and shows that even for the cases of users that attain a low wireless usage footprint, tracking is achievable.

4.3. Experiment II – Social Interaction with Children

A second, larger and interdisciplinary experiment has been conducted in May 2017, with the purpose of giving insight into social interaction analysis in children. For this purpose, an interdisciplinary team involving Senception, COPELABS (a team of social psychologists) as well as a school in Lisbon (Escola EB 2,3 Pedro D'Orey da Cunha, Damai) has been run[6]. The experiment counted with 80 children aged 11 to 16 from 8 different classes of the same school, and involved 8-10 teachers and involved data collected with PML as well as social psychology surveys provided to children. The experiment had the purpose of assisting three main studies, currently under publication i) analysis of contact and prejudice in children; ii) analysis of well-being and spaces; iii) Study on clustering and time correlation of roaming habits/mobility patterns in children. Traces are expected to be publicly available until month 36.

The methodology for this experiment was as follows. Out of this set, 50 children were allowed (by their parents) to carry PML either on their smartphone, or on a smartphone provided to them. The data collected concerned roaming context during the school schedule (8a.m. to 5p.m.).

During a first week (Time I, 04.05.2017-12.05.2017) the children answered questionnaires with the school's educational psychologist. During the same period, installation of the software occurred. The data has been collected between 05.05.2017 until 06.06.2017. By the end of the month (Time II, 29.05.2017-05.06.2017), the children have again answered surveys, and data was extracted from PML.

In this report we provide a first analysis concerning clustering aspects of the collected data set concerning peer sightings. In Figure 16 a summary of the clustering analysis for affinity networks involving the 50 children is provided. The file has been treated to remove peers that could be other type of devices, e.g., TVs or printers. The resulting number of nodes correspond to the devices seen by the 50 children (source nodes). This implies that there are devices corresponding to the other children as well as "foreign" devices. As shown, the average path length is of 3 and the average number of neighbors corresponds to 5.

Clustering coefficient :	**0.0**	Number of nodes :	**75**
Connected components :	**2**	Network density :	**0.072**
Network diameter :	**7**	Network heterogeneity :	**1.067**
Network radius :	**1**	Isolated nodes :	**0**
Network centralization :	**0.384**	Number of self-loops :	**0**
Shortest paths :	**5258 (94%)**	Multi-edge node pairs :	**99**
Characteristic path length :	**3.008**	Analysis time (sec) :	**0.071**
Avg. number of neighbors :	**5.333**		

Figure 16: clustering analysis, school children experiment with PML light.

[6] The data analysis is currently being developed. In this report we provide an initial subset of results for clustering properties.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

A first clustering analysis was developed by recurring to the tool Cytoscape. A Cytoscape network has been formed having as source nodes the devices that did the PML readings; destination nodes corresponding to the peers; interconnection corresponding to time. Figure 17 provides the network layout where nodes correspond to the different devices; color reflects a continuous mapping of the node degree (the lighter the color the smaller the node degree, observed between 1 and 77); size reflects betweeness centrality. The nodes are placed in accordance with latitude (x position) and longitude (y position), and the links reflect the shared interaction over time.

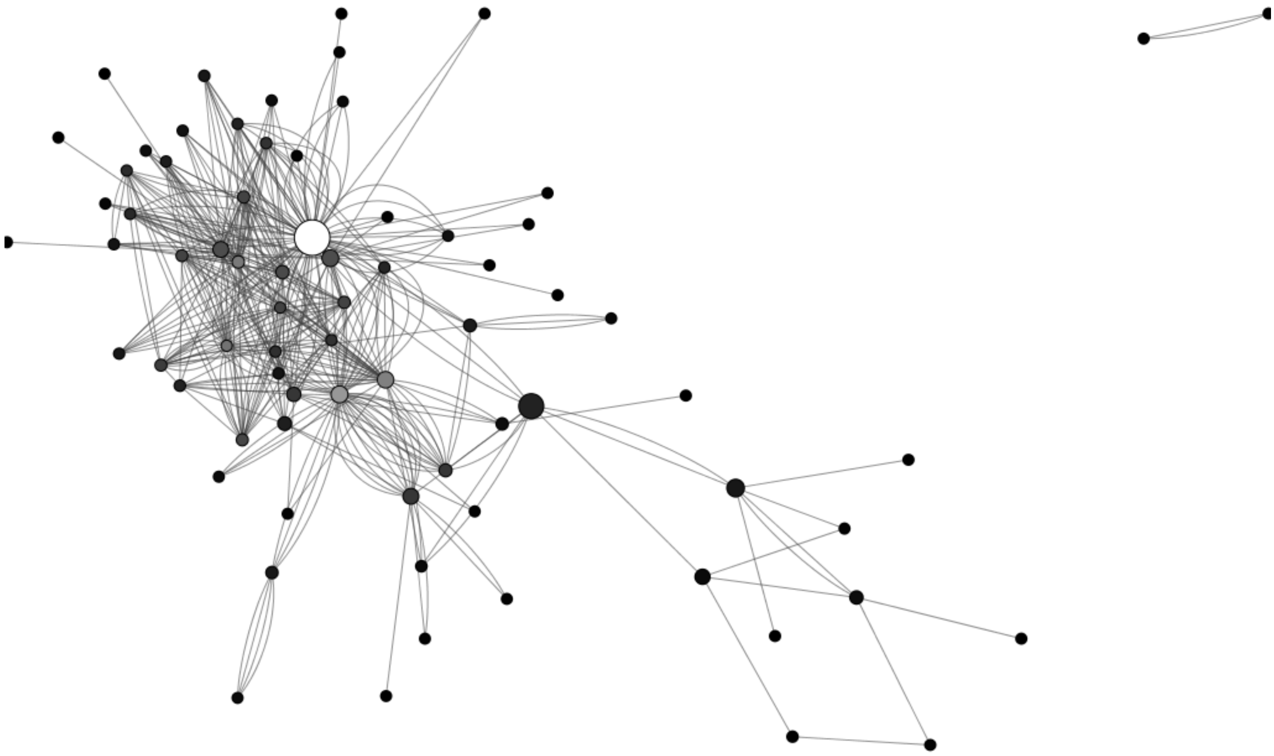


Figure 17: School children network, organic layout where the shared interaction corresponds to time (peers and time correlation).

Figure 18 provides a geographical distribution of the clusters where each color represents a different student. The concentration over time and space relates with the fact that the study was performed during school hours.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

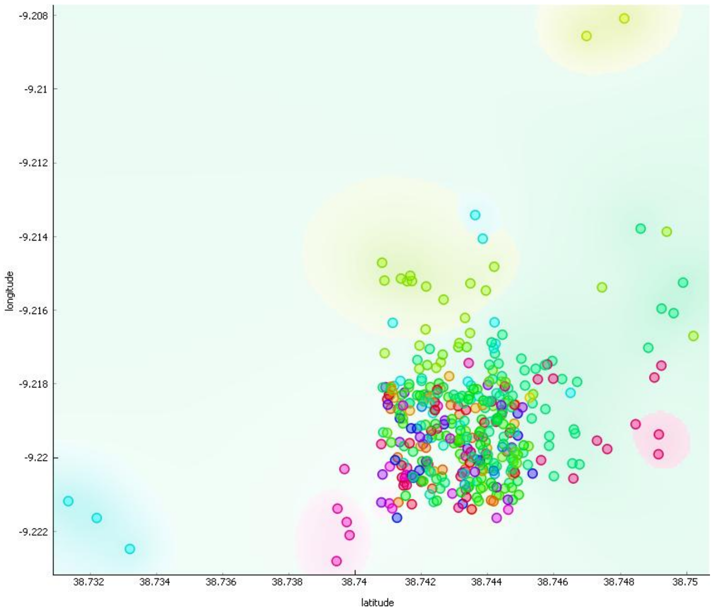


Figure 18: Distribution per latitude and longitude.

Figure 20 provides the neighborhood connectivity, where the red line corresponds to a Power Law line, R-squared value of 0.745, and holding a correlation of 0.846. Figure 20 corresponds to the shared neighborhood distribution, where the most frequent groups are formed by 1 or 2 peers. Hence, in this study children have a tendency to cluster in small groups. In terms of betweeness centrality (cf. Figure 21, where the red line corresponds to a fitted Power Law with correlation = 0.716 and R-squared equal to 0.492), there are not a significant number of nodes that concentrate a higher level of between centrality, which seems to imply that the different clusters exhibit a similar pattern of interaction over time and space.

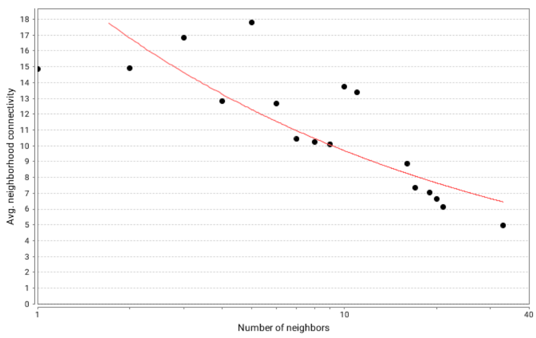


Figure 19: Neighborhood connectivity distribution.

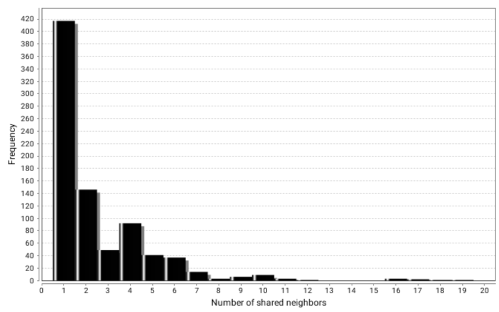


Figure 20: Shared neighborhood distribution.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

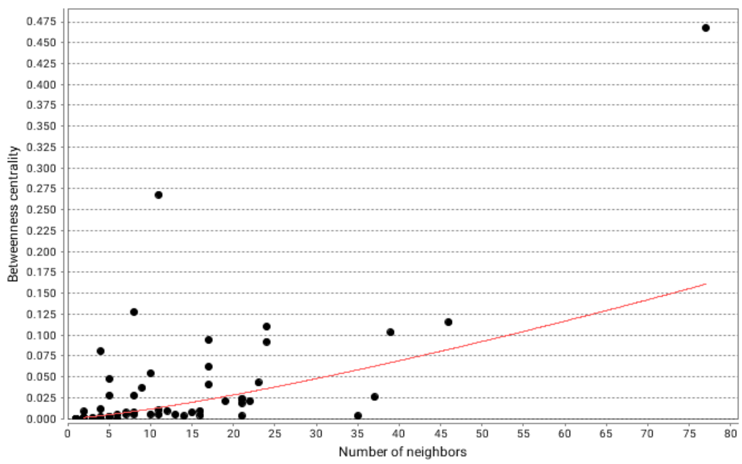


Figure 21: Betweeness centrality distribution.

5. Conclusions

In this deliverable we present the UMOBILE contextual plane, explaining the relevancy of such a plane in the context of UMOBILE. The deliverable corresponds to the outcome of task 4.2, and the main goal of this task was to give insight into i) how to develop mechanisms for processing of sensor data through context understanding; ii) to show how such mechanisms apply in UMOBILE, and why they are relevant.

The deliverable starts by introducing context-awareness, its role in networking as well as in UMOBILE. It then describes the contextual plane of UMOBILE, detailing its modules, and how it interacts towards other modules. The contextual plane, supported in the UMOBILE proof-of-concept by the Contextual Manager service, is currently being developed in the context of WP5.

The deliverable gives insight into the tools that were developed for the purpose of assisting contextualization in UMOBILE, and describes experiments that have been pursued, providing as well open-access traces collected during such experiments.

In addition to the implementation of the CM (WP5), further experiments are being developed to assist the validation of the CM. Such additional experiments shall be detailed in the context of WP6, dissemination.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

6. References

[1] Abowd, G., Dey, A., Brown, P., Davies, N., Smith, M. and Steggles, P., 1999. Towards a better understanding of context and context-awareness. In Handheld and ubiquitous computing (pp. 304-307). Springer Berlin/Heidelberg.Perera, Charith, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. "Context-aware computing for the internet of things: A survey." IEEE Communications Surveys & Tutorials 16, no. 1 (2014): 414-454.

[2] Bonomi, Flavio, Rodolfo Milito, Jiang Zhu, and Sateesh Addepalli. "Fog computing and its role in the internet of things." In Proceedings of the first edition of the MCC workshop on Mobile cloud computing, pp. 13-16. ACM, 2012.

[3] Guo, Bin, Daqing Zhang, Zhu Wang, Zhiwen Yu, and Xingshe Zhou. "Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things." Journal of Network and Computer Applications 36, no. 6 (2013): 1531-1539.

[4] Christos-Alexandros Sarros, Sotiris Diamantopoulos, Sergi Rene, Ioannis Psaras, Adisorn Lertsinsrubtavee, Carlos Molina-Jimenez, Paulo Mendes, Rute Sofia, Arjuna Sathiaseelan, George Pavlou, Jon Crowcroft, Vassilis Tsaoussidis. Connecting the edges: A universal, mobile-centric and opportunistic communications architecture. IEEE Communications Magazine, Special Issue on Human-Driven Edge Computing and Communication, November 2017.

[5] R. Sofia, P. Mendes. A characterization study of Human Wireless Footprints with PerSense Mobile Light. Technical Report, Senception and COPELABS. DOI: 10.13140/RG.2.1.2789.4647. March 2016.

[6] Schilit, Bill N., and Marvin M. Theimer. "Disseminating active map information to mobile hosts." IEEE network 8.5 (1994): 22-32.

[7] Chen G , Kotz D. A survey of context-aware mobile computing research. Dartmouth: Dartmouth Computer Science Technical Report; 2004

[8] Perera, Charith, et al. "Context-aware computing for the internet of things: A survey." IEEE Communications Surveys & Tutorials 16.1 (2014): 414-454.

[9] . D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles, Towards a better understanding of context and context-awareness," inProceedings of the 1st nternationalsymposium on Handheld and Ubiquitous Computing, ser. HUC '99.London, UK: Springer-Verlag, 1999, pp. 304–307. [Online]. Available:http://dl.acm.org/citation.cfm?id=647985.743843

[10]Makris, Prodromos, Dimitrios N. Skoutas, and Charalabos Skianis. "A survey on context-aware mobile and wireless networking: On networking and computing environments' integration." IEEE communications surveys & tutorials 15.1 (2013): 362-386.

[11]Perera, Charith, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. "Ca4iot: Context-awareness for internet of things." In Green Computing and Communications (GreenCom), 2012 IEEE International Conference on, pp. 775-782. IEEE, 2012.

[12]Bonomi, Flavio, Rodolfo Milito, Jiang Zhu, and Sateesh Addepalli. "Fog computing and its role in the internet of things." In Proceedings of the first edition of the MCC workshop on Mobile cloud computing, pp. 13-16. ACM, 2012.

[13]Perera, Charith, et al. "Context-aware sensor search, selection and ranking model for internet of things middleware." Mobile Data Management (MDM), 2013 IEEE 14th International Conference on. Vol. 1. IEEE, 2013.

[14]Wei, Qiang, and Zhi Jin. "Service discovery for internet of things: a context-awareness perspective." In Proceedings of the Fourth Asia-Pacific Symposium on Internetware, p. 25. ACM, 2012

[15]Scott, James, Jon Crowcroft, Pan Hui, and Christophe Diot. "Haggle: A networking architecture designed around mobile users." In WONS 2006: Third Annual Conference on Wireless On-demand Network Systems and Services, pp. 78-86. 2006.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

[16]Musolesi, Mirco, et al. "Writing on the clean slate: Implementing a socially-aware protocol in Haggle." World of Wireless, Mobile and Multimedia Networks, 2008. WoWMoM 2008. 2008 International Symposium on A. IEEE, 2008.
[17]Förster, Anna, Kamini Garg, Hoang Anh Nguyen, and Silvia Giordano. "On context-awareness and social distance in human mobility traces." In Proceedings of the third ACM international workshop on Mobile Opportunistic Networks, pp. 5-12. ACM, 2012.
[18]EU IST FP7 User-centric Wireless Local Loop. 2010-2013. http://copelabs.ulusofona.pt/~uloop
[19]Nascimento, A., Sofia, R.C., Condeixa, T. and Sargento, S., 2011, July. A Characterization of Mobility Management in User-Centric Networks. In NEW2AN (pp. 314-325)
[20]Barcelo-Arroyo, F., Gorawski, M., Grochla, K., Martín-Escalona, I., Połys, K., Ribeiro, A.G., Sofia, R. and Zola, E., 2014. New Trends in Mobility Modelling and Handover Prediction. In Wireless Networking for Moving Objects (pp. 88-114). Springer International Publishing.
[21]Sofia, Rute. "A Tool to Estimate Roaming Behavior in Wireless Architectures." In International Conference on Wired/Wireless Internet Communication, pp. 247-258. Springer, Cham, 2015.
[22]Bettstetter, Christian. "Mobility modeling in wireless networks: categorization, smooth movement, and border effects." ACM SIGMOBILE Mobile Computing and Communications Review 5, no. 3 (2001): 55-66.
[23]Ribeiro, A.; Sofia, R. A survey on Mobility Models for Wireless Networks. Technical Report, SITI-TR-11-01. February 2011.
[24]Hossmann, Theus, Thrasyvoulos Spyropoulos, and Franck Legendre. "Putting contacts into context: Mobility modeling beyond inter-contact times." Proceedings of the Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing. ACM, 2011.
[25] Hsu, W.J., Spyropoulos, T., Psounis, K. and Helmy, A., 2007, May. Modeling time-variant user mobility in wireless mobile networks. In INFOCOM 2007. 26th IEEE international conference on computer communications. IEEE (pp. 758-766). IEEE.
[26]Andréa Ribeiro, Rute C. Sofia and André Zúquete, Modeling Pause Time in Social Mobility Models. inProc. of the 2012 International Symposium on Wireless Communication Systems (ISWCS), pp 656-660, August 2012. DOI: 10.1109/ISWCS.2012.6328449
[27]Musolesi, Mirco, and Cecilia Mascolo. "A community based mobility model for ad hoc network research." In Proceedings of the 2nd international workshop on Multi-hop ad hoc networks: from theory to reality, pp. 31-38. ACM, 2006.
[28]Borrel, Vincent, Franck Legendre, Marcelo Dias De Amorim, and Serge Fdida. "Simps: Using sociology for personal mobility." IEEE/ACM Transactions on Networking 17, no. 3 (2009): 831-842.
[29]Boldrini, Chiara, Marco Conti, and Andrea Passarella. "ContentPlace: social-aware data dissemination in opportunistic networks." Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems. ACM, 2008.
[30]Distl, Bernhard, et al. "Extending the reach of online social networks to opportunistic networks with PodNet." Proceedings of the Second International Workshop on Mobile Opportunistic Networking. ACM, 2010.
[31]Álvarez-García, Juan Antonio and García, Álvaro Arcos and Chessa, Stefano and Fortunati, Luigi and Girolami, …, "Detecting Social Interactions in Working Environments Through Sensing Technologies", Springer International Publishing (2016), 21--29.
[32]Greenberg, Saul and Marquardt, Nicolai and Ballendat, Till and Diaz-Marino, Rob and Wang, Miaosen, "Proxemic interactions", interactions (2011), 42.
[33]Osmani, Venet and Carreras, Iacopo and Matic, Aleksandar and Saar, Piret, "An analysis of distance estimation to detect proximity in social interactions", Journal of Ambient Intelligence and Humanized Computing (2014), 297--306.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

[34]Palaghias, Niklas and Hoseinitabatabaei, Seyed Amir and Nati, Michele and Gluhak, Alexander and Moessner, Klaus, "Accurate detection of real-world social interactions with smartphones", IEEE (2015), 579--585.
[35]Choudhury, Tanzeem and Pentland, Alex, "Sensing and Modeling Human Networks Using the Sociometer", IEEE Computer Society (2003), 216----.
[36]Rachuri, Kiran K and Mascolo, Cecilia and Musolesi, Mirco and Rentfrow, Peter J, "SociableSense: Exploring the Trade-offs of Adaptive Sampling and Computation Offloading for Social Sensing", ACM (2011), 73--84.
[37]Sofia, R. C. and Firdose, S. and Lopes, L. A. and Moreira, W. and Mendes, P., "NSense: A People-centric, non-intrusive Opportunistic Sensing Tool for Contextualizing Nearness. DOI 10.1109/HealthCom.2016.77…", IEEE (2016).
[38]Bonacich, Phillip. 1987. “Power and Centrality: A Family of Measures.” American Journal of Sociology 92: 1170-1182.
[39]Linton C. Freeman: Centrality in networks: I. Conceptual clarification. Social Networks 1:215-239, 1979. http://leonidzhukov.ru/hse/2013/socialnetworks/papers/freeman79-centrality.pdf
[40]PerSense Mobile Light traces set 1. 2015. Available via https://www.researchgate.net/profile/Rute_Sofia/publication/320058498_PML_traces1/data/59cb88c6a6fdcc451d5cabb5/PML-traces1.zip. DOI: 10.13140/RG.2.2.33952.69126.
[41]PerSense Mobile Light. Available at https://play.google.com/store/apps/details?id=com.senception.persenselight&hl=pt_PT.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124