This is a preprint version of a paper accepted for presentation at the IEEE Globecom 2013 Conference. The published version will be made available by IEEE in the IEEEXplore Digital Library.

How often social objects meet each other?

Analysis of the properties of a social network of IoT devices based on real data

Hamid Zargari Asl , Antonio Iera
University of Reggio Calabria
{name.surname}@unirc.it

Luigi Atzori
University of Cagliari
l.atzori@diee.unica.it

Giacomo Morabito
University of Catania
giacomo.morabito@dieei.unict.it

***Abstract*— Internet of Things (IoT) applications will be based on the interactions between smart objects. In many applications such interactions are possible (or meaningful) when objects are close to each others, i.e., there is a *co-presence*. Unfortunately, to date there are no data traces providing information about the co-presence of smart objects. Indeed, several mobility traces reporting humans' movements are available, but none of them contains information about the interactions between their objects. Objective of the work reported in this paper is to fill this gap. To this purpose, we start from user mobility patterns available from several datasets. We associate to each user a set of objects, based on a survey we have carried out over around 450 users. Accordingly, we analyze the statistics about the co-presence of objects. We carry out our analysis by exploiting the tools developed for the analysis of complex networks. Our objective is to identify the objects which are likely to play a key role in the interactions between smart objects in the IoT.**

Keywords—IoT, social graph, mobility data set

I. INTRODUCTION

In the web 2.0 era, social awareness has become a key enabler of significant changes in applications and protocols designed in the view of increasing the system performance and the Quality of Experience offered to users. In this context, relevant examples include Social oriented forwarding in Delay Tolerant Networks (DTN) [1], realistic mobility models based on social networks for VANET [2], social networking in peer-to-peer systems [3].

The Internet of Things domain is no exception to this trend. In the future Internet, the majority of connections will not be established among humans but among devices [4] (things, more or less "smart"). Therefore, typical notions, rules, modes of interaction, and dynamics of social networking must inevitably be extended to the networks of objects. Smart objects will need to operate in tremendously multifaceted contexts and it seems unlikely that a single (even very smart) object will ever have the capabilities to face the deriving complexity by themselves. A new generation of *social* objects may possibly interact with other objects to deal with this complexity.

In this vision, "Social IoT" (SIoT) [5] networks use concepts and technologies of social networks to foster resource visibility, service discovery, object reputation assessment, source crowding, and service composition. To users, this approach gives the possibility of enjoying new powerful applications, which rely on social interactions between mobile devices and environmental sensors to hide the complexity of the underlying IoT and to offer a more natural access to complex ICT environments (by also interacting with Social Networks of humans). To the underlying network layers, the approach gives the opportunity to improve the efficiency of policies for the management and selection of communication resources based on social concepts such as community and centrality.

A first approach to follow is to establish social relationships between devices (PC, mobile phones, etc.) that show a certain customary in exchanging data. It is easy to find examples of studies focusing on the analysis of social networks established between cellular phones. These rely on the availability of real logs of phone calls, which means that the social relationships of devices are derived from social interactions among their owners [6]. A complementary approach is used to build so called detected social networks [7] (i.e. devices that have a high probability to come into short-range coverage and establish social links), which have proven to be highly effective in supporting information exchange in DTN or opportunistic communications [8].

This paper starts from the following observations: (i) the cited studies are reductive because usually the only things considered are cellular phones that either remotely call one another or are mutually detected through short-range technologies (ii) the availability of real data relevant to the probability that different things come into contact is limited and restricted to specific events [9] (iii) in no study there is a real notion of social relationships among devices, on which IoT protocols and applications should rely.

The main focus of this paper is thus on interactions among things. Our objective is to understand if, and to which extent, these interactions can be exploited to set up a social graph of objects on which to base algorithms, protocols and applications of any future Social IoT. Co-presence of things is one of the means through which a social interaction among objects is established.

The paper will thus address this aspect within the framework of the Social Internet of Things defined in [10] trying to answer questions like: which objects meet others in given places? How many times do such meetings occur? How long does any contact last? If reliable answers will be found then a model of objects' co-presence with relevant social ties will be available.

As we need a model that involves both fixed things and things in mobility, carried along by persons, we have to focus also on tracing human mobility and human presence in different every-day-life environments such as residences, shopping points, sport places and so on. By starting from this latter information, available from both real and simulated data sets, the presence of objects can be easily predicted if we have reliable information on the probability of carrying things along with their owner in any attended environment. We intend to follow this approach by relying on a data set available from real experiments and on the knowledge of human behaviors derived from a survey activity. These data, properly combined will allow us to come to an analysis of the social behavior of the members (sensors, devices, objects, etc.) of a future Social Internet of Things.

This study intends to provide a useful source of information on co-presence and interaction modalities of different fixed and mobile devices. Notably, our analysis has been conducted from real human mobility patterns and real information from a conducted survey. The social graph obtained by this study is the first of its kind as it refers to a Social Internet of Things which, if properly exploited, will open up enormous opportunities for research in the direction of inserting things in the human social loops.

II. THE FOLLOWED APPROACH

To build our model of objects' co-presence in the SIoT, we need to consider the type of places where the objects' contacts happen [5]. To this aim, we have grouped the places wherein a human being is supposed to spend her time for the different activities carried out during her life into the following sets: Working places (W), Car and open places (C), Sport (P) and Entertainment (E) Places, Shopping places (S) and Homes and residences (H).

In some studies, the physical coordinates of objects (or just their vicinity to Wi-Fi or GSM antennas) are revealed, but in none of them user coordinates are put into relationship with the corresponding *type* of place. This is why we use human mobility models for inferring the presence of a person in different kinds of places and, thereafter, retrieve the probability of presence of given objects. We refer to a real user mobility data set available from the literature in which some minimal information on the type of place wherein the users move is given. To assess the behavior of the objects in more complex scenarios, we also used the output of a simulator of user movements according to the "small world" model, SWIM (for details, please refer to [11]).

As already told, we had to know which objects are carried by individuals in different places. For solving this problem, we made an Internet survey to find for any object the probabilities to attend any place. In this survey, we collected the list of 12 different types of objects in the different places or situations listed.

A. *Probability of carrying objects*

For collecting the probabilities of presence of different objects in different places, a survey on the Web conducted by about 450 volunteers (mainly students and staff of the three Universities of the authors) was carried out. The result of this survey indicates the number and type of electronic gadgets carried to any place. In turn, the attendance probability of any object in any place has been easily estimated. Table I shows the considered places and objects and the computed *probability that one person brings a given device in a given type of place*. Detailed information on the survey results is available on http://www.social-iot.org/

TABLE I. RESULTS OF THE CONDUCTED SURVEY

	H	T	B	W	S	P	E	C
Desktop	0.62			0.61				
NTW Devices	0.74							
Laptop	0.82		0.52	0.55		0.01	0.01	
Camera		0.76	0.43					
MP3 reader	0.67	0.51	0.37	0.26	0.17	0.36	0.10	0.10
Navigator								0.06
Environment Sensors	0.15			0.36				
Vital Param. Sensors						0.06		
Printer				0.54				
Tablet	0.15	0.11	0.10	0.20	0.02	0.01	0.04	0.02
Cellular Phones	0.94	0.98	0.91	0.79	0.97	0.42	0.96	0.88
TV	0.91							

LEGENDA: H= Home ; T= Trip ; B= Business Trip ; W= Work ; S= Shops ; P=Sport ; E=Entertainment ; C=Car & open places

B. *Mobility data set*

Concerning mobility, after having analyzed a large number of data sets, the one collected by the University of Milan [12], hereafter called *Milan* data set, has been selected as the most appropriate for our purposes, because it offers the opportunity to recover some information about the kind of places where people meet.

Milan data set has been directly downloaded from the CRAWDAD website. In it, human mobility real data is collected by using sensors, although this refers to two types of places, 44 nodes, and 19 days only. The fixed nodes are considered as places, which are five and indicated by five different IDs. We received from the authors of the research relevant to the Milan data set a description of their sampling nodes and further information such as which are the fixed and moving nodes and which is the type of place where the fixed nodes were installed.

The collected data are very informative for our purpose as they allow to find place-aware data, though, it is not a comprehensive data set. Indeed, only two place types are considered. Classrooms and corridors ar supposed as working places (W) and cafeteria as entertainment place (S). In this case, there are five fixed sensors: three in cafeteria (nodes 45, 46, 47) and two in saloons (nodes 27, 43). We consider the five fixed nodes as places and look for co-present nodes. This means that if two (or more) nodes are seen by a fixed sensor, then we infer that these nodes are co-present; the overlapped times of being seen by the sensor are assumed to be their *co-presence time duration*. Clearly, the type of place, as indicated before, is recorded in any co-presence event.

C. *The simulated mobility patterns*

The output of the SWIM (Small World in Motion) simulation of the experiment Cambridge 06 is collected from the SWIM's official web site. In this data set, each line shows the node ID, its coordinates (such that $x,y \in [0,1]$), its meeting or leaving event, and timing of contacts. As in the SWIM there is no information about the types of place, we split a square with dimension of 1x1 into 100 different cells and then associated a "type of place" to each of them as shown in Figure 1.

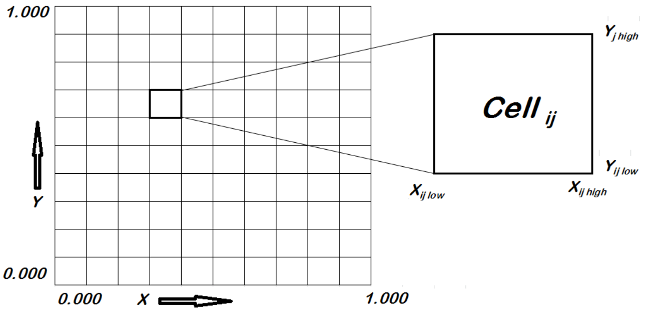

Fig. 1. Dividing SWIM area to one hundred cells

The distribution of the types of place was randomly generated, but the number of cells associated to each type of place was calculated according to real data. In fact, we used the information on the land usage of a portion of New York city available from the official web site of the Municipality of New York [13]. By processing the information extrapolated by such data, we obtained the values reported in Table II.

TABLE II. PERCENTAGES OF THE TYPES OF PLACES IN NEW YORK

Place type	H	W	S	P, E	C
%	43.85	7.12	3.24	7.45	38.34

We traced the nodes in every cell. If during the same time a node attends a cell another node arrives in, then this event is recorded. The presence of any node, l, in the (i,j)-th cell at time t, is recognized as follows:

$$(X_l(t) \leq X_{ij\,high}) \wedge (X_l(t) \geq X_{ij\,low}) \wedge (Y_l(t) \leq Y_{ij\,high}) \wedge (Y_l(t) \geq Y_{ij\,low})$$

where $X_l(t)$ and $Y_l(t)$ denote the coordinates of node l at time t. For locating any node, its coordinates were - all the times - compared with boundaries of any cell. This was done for any node so that fitting the coordinates of any node in the same cell and in the same time were recorded as an event of co-presence. We, thus, calculated the *number of contacts* between nodes, as well as their *co-presence time duration*.

D. *Associating things to places*

The probabilities collected from the Internet survey are used to associate the right object/objects to any human in a given place. In so doing, we used the following procedure. After having found the co-presence of human nodes (persons) in any place, we exploited the probabilities from the survey to discover objects carried by each of them. Figure 2 shows a sample connection between two generic humans, $Node\_1$ and $Node\_2$, belonging to the set of human actors $\{Node\_l : l=1\ldots N\}$ and depicted like *graph vertices representing humans* HV_1, HV_2 $\in \{HV_l : l=1\ldots N\}$. Each human owns O_l objects (in Figure 2, both O_1 and O_2 are assumed to be 3), which can be represented as *graph vertices representing objects* OV^l_k with $k=0\ldots O_l$. In this sample arrangement, to map the human link onto objects' links, it is desirable to substitute the edge $HE_{1,2} \in \{HE_{u,v} : u=1\ldots N\ ;\ v=1\ldots N\}$ between humans with a set of potential edges $\{OE^{1,2}_{z,w} : z=1\ldots O_1\ ;\ w=1\ldots O_2\}$ between any couple of object vertices. In doing so, if one human owns O_1 objects and the other owns O_2 objects, then a maximum of $O_1 \times O_2$ potential inter-object links can be established. The *subset* of links *actually* established among objects is chosen according to the probability that a given object is actually brought along with the user. More specifically, one can consider P^i being the probability that the *i-th* object of a group of O_l possible objects of the *l_th* person is carried to a specific place (P^i is derived from the data in Table I).

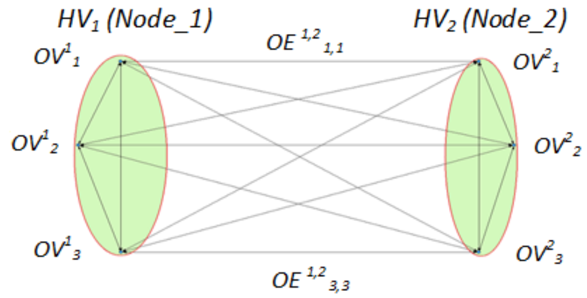

Fig. 2. Potential links among objects carried by two humans (social actors)

By using the values of P^i, human nodes may be replaced by objects in any kind of place where they meet, and, consequently, two sets of objects belonging to two connected humans will be connected. This operation can be seen as mapping human graph $G_h(V_h,E_h)$ onto the graph of objects $G_o(V_o,E_o)$. As connections between two fixed objects belonging to two persons is not acceptable (people do not bring with them their fixed objects, such as for example TV), we take this feature into due account in our study.

III. PERFORMANCE EVALUATION

By starting from the data obtained as shown above, we can have a clear idea on the things' social behavior both in terms of "total co-presence time" and "number of contacts".

A. *Co-presence time in Milan data set*

The focus is mainly on the Milan case, as it is representative of a *real mobility* trace and *real statistics* from the survey relevant to the things' displacement. Social graphs between objects are obtained by using the Netdraw tool visualization software and the analysis is made through the Ucinet Software.

For a better understanding of the characteristics of the graphs, original graphs have been progressively reduced. This is achieved by reducing the weights of the links and progressively deleting the edges with lower values than a threshold for eliminating the weaker links. This corresponds to progressively reducing the order of graph ($|E(G)|$). We also deleted pendants and isolate nodes at any step because these have not important roles and the graphs are not essentially changed by deleting them.

In a first analysis, the considered weight of the graph links is the *total co-presence time* expressed in seconds and between any pair of nodes (Figure 3).

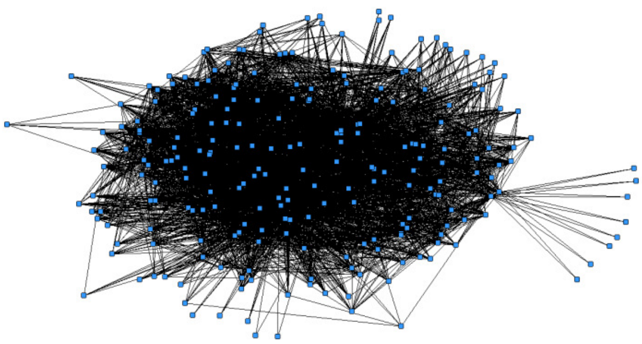

Fig. 3. Things' social graph for Milan data set based on co-presence time

It has been observed that when reducing the link weights by 2000, linear structures connecting portions of the graph begin to appear. This linear structure is always shown by going further in the reduction up to 4000 (graph in Figure 4). By considering this latter graph and by using Ucinet software, links with higher weights are highlighted by thicker edges.

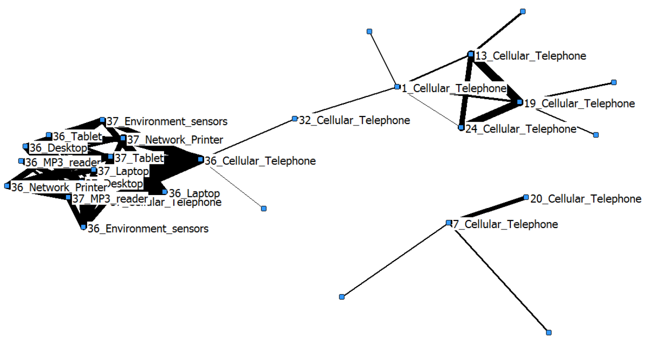

Fig. 4. Things' social graph for Milan data set with weights > 4000

In the Figure, the number in the label represents the object ID. It shows that the densest component at the left side of the graph is composed by different objects. The line , which is connecting two main components of graph, is totally made by cell phones. In the small separated graph there are cell phones in both sides of the thicker link.

To have a numeric estimation, the weights (co-presence times between any pair of objects) are sorted in descending order and the nodes that they connect are listed in Table III. As expected, cell phones are present in all lines of this Table and 13 things out of the 20 shown things are cell phones. The highest positions in the rank are also occupied by cell phones.

TABLE III. TOP 10 POWERFUL LINKS CONSIDERING THEIR CO-PRESENCE TIME IN SECONDS

Node 1	Node 2	Weight (s)
36_Cellular_Telephone	37_Cellular_Telephone	22392
36_Cellular_Telephone	37_Desktop	22392
36_Cellular_Telephone	37_Laptop	22392
36_Cellular_Telephone	37_Network_Printer	22392
13_Cellular_Telephone	19_Cellular_Telephone	21291
19_Cellular_Telephone	24_Cellular_Telephone	18999
37_Cellular_Telephone	36_Desktop	15967
37_Cellular_Telephone	36_Environment_sensors	15967
37_Cellular_Telephone	36_Laptop	15967
37_Cellular_Telephone	36_Network_Printer	15967

The probability distribution function of co-presence times between any pair of objects is also shown in Figure 5.

Last, it is worth observing that our studies have demonstrated that, as expected, the highest values in all kinds of Centralities (*Degree, Bonacich, 2step closeness, ARD, Eigenvector*, and *Betweenness* computed through Ucinet) are associated to cell phones; the second position in the rank is usually occupied by desktops and the third one by cell phones again. Laptops occupy the majority of fourth positions, while other objects have lower scores. In terms of Betweenness, Degree, and Bonacich power centralities, the cell phones have strictly higher values than the second objects in the ranks.

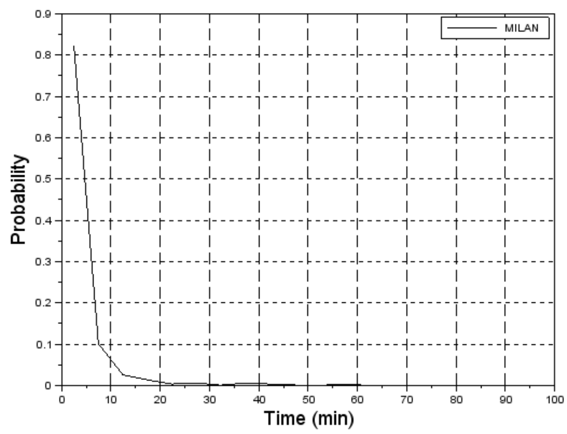

Fig. 5. Probability Distribution Function of co-presence for Milan data set.

B. *Number of contacts in Milan data set*

The probability distribution function of the number of contacts is shown in Figure 6. This Gamma like distribution begins from 0.65 for one contact and decreases exponentially until zero in near 7 contacts. The corresponding graph is also progressively reduced and it is observed that up to a reduction of the graph weights by 7 times, there are no separated components.

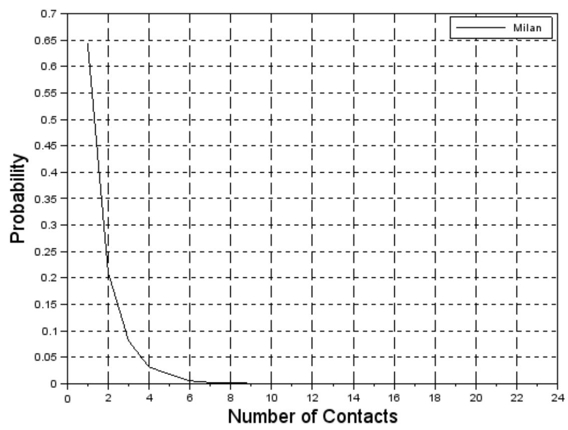

Fig. 6. Probability Distribution Function of nr. of contacts in Milan data set

The final graph without deleting the pendants shows a linear structure which is connected to a group of nodes. Following the weight reduction by 2, two big ego networks ("hub" like) are observed. We focused on the relevant graph (shown in Fig. 7). In the figure, the nodes with the highest centrality degree are highlighted. It shows that objects in center of left ego are limited to cell phones (majority), laptop, desktop, and network printer. In the right ego, there are only cell phones as central nodes.

The prepared graph after a reduction of the edge weights by 7 times is depicted, with relevant labels, in Figure 8 (the link thicknesses represent the weights). It shows that the linear structure is composed of cell phones only and this linear structure at one end is connected to a group of nodes. There is also a separate component that is made solely by cell phones. Further studies conducted with Ucinet on this reduced graph demonstrate, as expected, that the twelve highest weights are absolutely connecting cell phones to each other and in the first 20 strongest links only four desktops, four laptops and four network devices, one laptop, and one network printer out of 40 objects are observed.

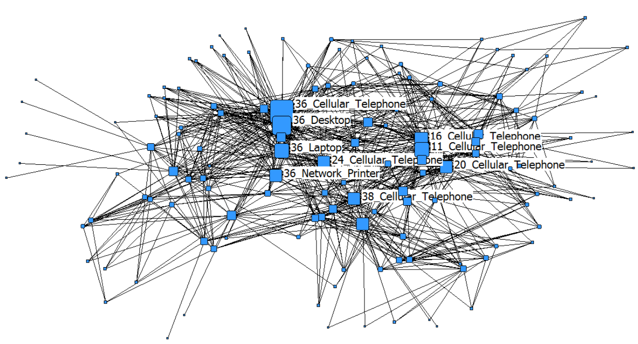

Fig. 7. Central objects of big egos in the reduced graph for Milan data set.

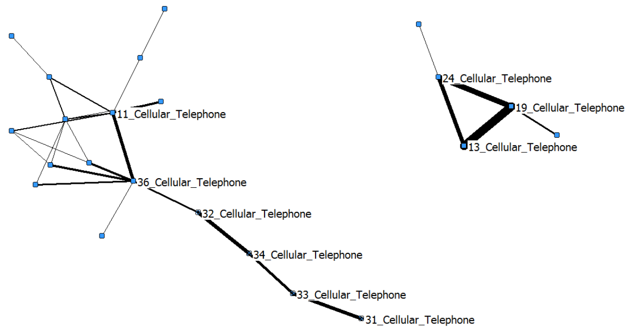

Fig. 8. Graph structure in Milan data set after subtracting any weight by 7

C. Co-presence times in SWIM data set

The results commented so far refer to a real mobility data set which, anyhow, considers a limited number of places. To confirm the results in more complex scenarios (more place types and more players), we also analyzed the system behavior with simulated data by SWIM, as described in Section II.b.

As expected, the social graph among objects, built on the co-presence time, is richer in links than before. Compared to previous results, in the SWIM case, the nodes which have more powerful links does not refer to one kind of thing only (although still cellular phones have a prominent role, as expected). Different things are contributing to the graph. The measure in which this happens can be derived by the values in Table IV, where the information about the links with higher weights are shown.

TABLE IV. TEN MORE POWERFUL LINKS IN SWIM DATA SET

Node 1	Node 2	Weight
7_Cellular_Teleph1one	14_Cellular_Telephone	159749.0
12_Cellular_Telephone	8_Cellular_Telephone	151447.3
12_Cellular_Telephone	8_Laptop	151447.3
12_Cellular_Telephone	8_TV	151447.3
12_Cellular_Telephone	8_Desktop	151447.3
12_Cellular_Telephone	8_MP3_reader	151447.3
12_Cellular_Telephone	8_Network_Devices	151447.3
12_Laptop	8_Cellular_Telephone	151447.3
12_Laptop	8_Laptop	151447.3
12_Laptop	8_TV	151447.3

By focusing on the twenty objects with the highest centrality values (*Degree*, *Bonacich*, *2step closeness*, *ARD*, *Eigenvector*, and *Betweenness*, computed through Ucinet) it has been confirmed that the highest centrality values belong to cellular phones and the second objects with higher centralities are laptops. The third ones are only MP3 readers and there are no other kinds of objects in the rank. In 2-step closeness and Betweenness centralities the role of cellular phones is more relevant, so that 18 objects out of 20 high ranked objects are cell phones. Last, the probability distribution function of the *co-presence time* and of the *number of contacts* between nodes (things) show very similar trends to the Milan case. This testify to the high reliability of the results found by relying on the real mobility data set and the object displacement probabilities obtained through our survey.

D. Discussions and final remarks

Our findings suggest regarding the cell phones as the most important object in different kinds of centralities. This results show that for service discovery purposes they have higher capabilities as well as higher routing and transmitting abilities. Laptops and desktops are in the next steps and can also be suitable alternatives in lack of cellular phones. Higher scores of different centralities for cellular phones show that they are the most suitable devices for service discovery, service providing, and for routing data across the network. A further interesting aspect is that all the charts on the number of contacts and co-presence have similarities to Gamma distributions. This also holds for persons in any place.

Our study is intended as a basis for future researches on the social behavior of things in the IoT. Actually, the research needs to be repeated for more complete place aware data sets as soon as these will be available. Further work needs to be done to better characterize the social relationships that things can establish in given places. What we intend is that one possible research step is to map rough information, like those shown that only refer to number of contacts and co-presence durations, onto well-defined *relational categories* and *types of social relationship* like those defined in [5].

A further future step will be to study the properties of the graphs resulting from the social ties established according to the different defined relationship types. Thank to this analysis, it will be possible to define the properties that the graph shall possess and modify the mechanism that trigger any social link establishment accordingly (for example properly define the threshold on the frequency of contacts and the co-presence duration that allow to trigger a given kind of social relationship among objects). Besides, the behavior of different algorithm for the information diffusion in Social Networks of humans can be assessed in the Social Internet of Things graphs and, if required, new algorithms can be suitable designed and their properties investigated. Equally important is also to evaluate, by starting from the observed object behavior, the importance of any object within the social network by taking into account also the intermittent availability typical of both mobile devices (due to, for example, battery charge expiration) and fixed devices (switching on and off of PC and Printers, etc.). What happens to the social network properties if the objects are not always available but they have a typical high dynamic profile?

In our opinion, it is manifest that the proposed research, may contribute to fill the void represented by the lack of actual statistics on the movement of objects carried by their owners across several locations in their everyday life. Besides, it has the potentials to open many new research paths towards future IoT paradigms that mimic the social networks of human.

IV. CONCLUSIONS

This study aimed at finding social rules and structures between objects of a Social Internet of Things to be used in other researches to foster things visibility, service discovery, thing reputation assessment, source crowding, and service composition. The conducted study is based on real data relevant to user mobility and real statistics about the displacement of objects obtained through a surveying activity. The property of the graphs of the inter-object ties, based on co-presence durations and number of contacts between objects, have been analyzed and their main properties highlighted. Also the output of the SWIM simulator for the Cambridge 06 experiment has been used to generalize the results obtained with real data and assess their reliability. The research output is a good starting point for future researches on the exciting issue of thing-to-thing information exchange based on social notions, typical of Social networks of humans, with an eye to the deployment of a future Social Internet of Things.

ACKNOWLEDGMENT

Special thanks go to Prof. Elena Pagani from the University of Milan for the valuable additional information on the Milan data set she has kindly shared with us.

REFERENCES

[1] Ying Zhu, Bin Xu, , Xinghua Shi, and Yu Wang, “A Survey of Social-Based Routing in Delay Tolerant Networks: Positive and Negative Social Effects”, IEEE Comm. Surveys & Tutorials, V. 15, N. 1, 2013

[2] Ana Gainaru, Ciprian Dobre, Valentin Cristea, “A Realistic Mobility Model Based on Social Networks for the Simulation of VANETs”, Proceedings of the 69th IEEE VTC Sping 2009 Conference

[3] Yamini Upadrashta, Julita Vassileva, Winfried Grassmann. "Social networks in peer-to-peer systems", IEEE ICSS 2005 Conference.

[4] Miguel Castro, Antonio J. Jara, and Antonio F. Skarmeta. "An analysis of M2M platforms: challenges and opportunities for the Internet of Things.", IEEE IMIS 2012

[5] L. Atzori, A. Iera, G. Morabito, M. Nitti, “The Social Internet of Things (SIoT) – When social networks meet the Internet of Things: Concept, architecture and network characterization”, Computer Networks, Volume 56, Issue 16, 14 Nov. 2012, Pages 3594–3608, Elsevier

[6] Eagle, N., Pentland, A. S., and Lazer, D., “Inferring friendship network structure by using mobile phone data”, Proceedings of the National Academy of Sciences, 106(36), 15274-15278.

[7] Vassilis Kostakos and Jayant Venkatanathan, “Making friends in life and online: Equivalence, micro-correlation and value in spatial and transpatial social networks”, in Proceedings of the 2010 IEEE, SOCIALCOM Conference, pages 587–594, Washington, DC, USA

[8] G. Bigwood, D. Rehunathan, M. Bateman, T. Henderson, and S. Bhatti, “Exploiting self-reported social networks for routing in ubiquitous computing environments”, Proceedings of the IEEE Int. Conference on Wireless and Mobile Computing, WIMOB ’08, Oct. 2008

[9] Anna-Kaisa Pietilainen and Christophe Diot, CRAWDAD data set thlab/sigcomm2009 (v. 2012-07-15), downloaded from http://crawdad.cs.dartmouth.edu/thlab/sigcomm2009.

[10] L. Atzori, A. Iera, G. Morabito, “SIoT: Giving a Social Structure to the Internet of Things”, IEEE Communications Letters, vol. 15, 2011

[11] Alessandro Mei and Julinda Stefa, "SWIM: A simple model to generate small mobile worlds." INFOCOM 2009, IEEE. IEEE, 2009

[12] Paolo Meroni, Sabrina Gaito , Elena Pagani, and Gian Paolo Rossi, CRAWDAD data set unimi/pmtr (v. 2008-12-01), Downloaded from http://crawdad.cs.dartmouth.edu/unimi/pmtr

[13] http://www.nyc.gov/html/dcp/pdf/landusefacts/landuse_tables.pdf