Proximity-Based Trust Inference for Mobile Social Networking⋆

Amir Seyedi, Rachid Saadi, and Valérie Issarny

ARLES Project-Team
INRIA CRI Paris-Rocquencourt, France
{name.surname}@inria.fr

Abstract. The growing trend to social networking and increased prevalence of new mobile devices lead to the emergence of mobile social networking applications where users are able to share experience in an impromptu way as they move. However, this is at risk for mobile users since they may not have any knowledge about the users they socially connect with. Trust management then appears as a promising decision support for mobile users in establishing social links. However, while the literature is rich of trust models, most approaches lack appropriate trust bootstrapping, i.e., the initialization of trust values. This paper addresses this challenge by introducing proximity-based trust initialization based on the users' behavioral data available from their mobile devices or other types of social interactions. The proposed approach is further assessed in the context of mobile social networking using users behavioral data collected by the MIT reality mining project. Results show that the inferred trust values correlate with the self-report survey of users relationships.

Keywords: Trust bootstrapping, mobile social network, small worlds.

1 Introduction

Portable devices have gained wide popularity and people are spending a considerable portion of their daily life using their mobile devices. This situation together with the success of social networking lead to the emergence of mobile social networking. However, anytime and anywhere interactions have a built-in risk factor. Development of trust-based collaborations is then the solution to reduce the vulnerability to risk and to fully exploit the potential of spontaneous social networking [5]. In our work, we aim at developing a trust management method for mobile social networking. Then, the challenge we are addressing here is how to initiate trust values and how to evaluate unknown mobile users using initiated trust values, to enable impromptu social networking.

Computational trust brings the human concept of trust into the digital world, which leads to a new kind of open social ecosystem [12]. In general, the notion of trust can be represented by a relation that links trustors to trustees. The

literature [17] includes two main categories of relations to set trust values for trustees, namely: (i) direct-based and (ii) recommendation-based relation.

Most existing trust models focus on assessing recommendation-based relationships [18] and lack the bootstrapping stage, which is how to initialize direct trust in order to efficiently start the trust model operation. This is very problematic and challenging, since recommendation-based relationships are built upon bootstrapped direct-based relationships. Indeed, most solutions that address trust assessment make one of the following assumptions:

– Trust initialization is not a problem of the model; it is the responsibility of the actors of the system   [8]. However, this task remains challenging, especially when it comes to evaluating trustees numerically (e.g., 0.1, 0.2, 0.15, etc.).
– The trust model initially evaluates trust relationships with a fixed value (e.g., 0.5 [9], a uniform Beta probabilistic distribution [10], etc.) or according to the trust disposition of the trustor [14] (i.e., pessimistic, optimistic, or undecided). In [16], trust is initialized by asking trustors to sort their trustees rather than assigning fixed trust values. There are other bootstrapping solutions [15, 2, 1] that assess trustees into different contexts (e.g., fixing a car, babysitting, etc.) and then automatically infer unknown trust values from known ones of similar or correlative contexts. However, if no prior related context exists, these solutions lack initialization of trust.

We have developed our trust model based on the hypothesis that it is possible to measure and bootstrap trust from human social behavior. Therefore, in this paper, we investigate a formal approach that quantifies human proximity from which possible trust relationships are transparently and automatically inferred and assessed on behalf of the trustor. We choose proximity between people as an effective measure for trust. Because, proximity between people is not only a matter of trust, but it increases trust affinity as well [4]. In other words, people spend more time with those whom they trust and, at the same time, if they start spending time with new people, it is likely that trust relationships will arise and evolve.

In order to better understand the contribution and evolution of proximity in the human society, consider the fact that a society is initiated by people who live in the same territory. Fukuyama [7], describing the role of trust in a society, mentions that people can build efficient economy and social organization, if they have wide and efficient trust networks. It shows clearly how trust and proximity of people are tied together to initiate a successful society. As a result, today we have different cultures and societies in the world simply because of their founders being at different location and proximity. Building on this social knowledge, this paper introduces a method for bootstrapping trust values in mobile environments, based on the *proximity* of people. However, in today's virtual world expanding the physical one, proximity is not just about the physical distance between people. Practically, while people who are physically close maybe detected using technologies such as Bluetooth [6], other types of proximity like phone calls, emails, social network interactions, etc. can be detected by

the implementation of virtual sensors. We classify the range of proximity-based trust values semantically for further judgments based on these values. Then, the initiated trust values can be used to calculate similarity between people from the standpoint of trust. Similar people can make good recommendations to each other. Hence, they can evaluate not-directly-known users on each others behalf. So, when mobile users are about to interact with unknown users, they may acquire the trust knowledge through known similar users. The process should be feasible in a limited number of hops because of the small world phenomenon [13].
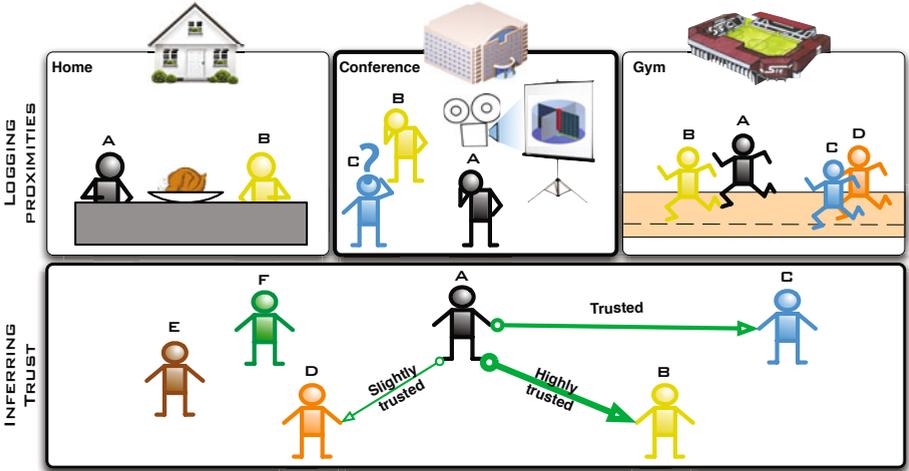

Fig. 1. User A evaluates others based on the observed proximity with others

The next section characterizes proximity towards trust assessment and is then followed by proposed proximity-based trust initialization in Section 3. Then, we evaluate the effectiveness of the proposed approach using the MIT reality mining dataset. Finally, we conclude in Section 5, summarizing our contribution and sketching our future work.

2   Trust and Proximity

We make the assumption that there is a strong correlation between proximity observations and real social relationship [6]. Proximity itself can be considered as a consequence of trust relationship, while at the same time the longer users are in proximity, the higher the probability of their friendship increases. Moreover, as noted by [4], proximity is a measure of trust as well as a cause to trust between users.

Thus, we argue that *proximity* is the *nearness* of any two persons in space or time. Let P be the proximity between two persons. Two persons are in *physical proximity* if the nearness happens in the same space and time. Two persons are

in *virtual proximity* if the nearness happens only in the time dimension. Physical proximity can be detected by various technologies (e.g., blue-tooth, Wi-Fi, etc.) and likewise virtual proximity through monitoring of social activities (e.g., chat, SMS, voice call, liking a content on facebook etc.). The collected proximity-related data provides information such as when, where, how frequently, and for how long people were in the proximity of each other.

In general, the definition of proximity takes several forms (from physical to virtual) and differs according to context (work, home, etc.) as well as it can be quantified by duration or frequency. From a social point of view, from the context in which proximity happens, we may identify a quality difference between the observed proximities. For instance, if the proximity happens at home, it is more intimate than a proximity in a professional meeting. Hence, in order to be able to aggregate of different types of proximity and consider the value difference between them, we characterize a proximity data type, namely η, as a tuple: $\eta =< p, l, t, d_s, d_d, s, m >$ where,

- Proximity type (p): Proximity type has two modes, virtual and physical. For instance, this helps distinguishing between a face-to-face interaction and virtual proximity.
- Location (l): Location is the position in physical space, in which the proximity happens. Location meaning can be expanded semantically, by looking to social semantic aspects of different definitions for location. For instance, *home* is a location in which trust is included by definition. Location has an effect on intimacy, e.g., the difference between outdoor and indoor proximity.
- Time (t): The time context is the temporal measurement of an instance in which the proximity happens. However time definition can be expanded semantically. For instance, weekend or working time has different social values. We take into account the quality difference of proximity as time changes, e.g., during weekend, being in the proximity of friends is more likely and therefore is a more valuable proximity.
- Source device type (d_s): Device type helps to includes the nature of device in terms of mobility etc. e.g., mobile device like smart-phone versus laptop). The observed proximity from a mobile device is more reliable as people are more likely to have their mobile device always with them. Then, d_s is the source device that belongs to the observer user.
- Destination device type (d_d): d_d is the destination device of a user who has been observed.
- Sensing method (s) (e.g., physical sensors, virtual sensors): Sensing methods take into account the technology effect on measurement method. or example bluetooth detects people in a shorter range than wifi does. So the detected proximity by bluetooth is more reliable as it catches the closer users.
- Measurement type (m): Measurement type indicates the difference between duration and frequency of a proximity. Hence, we introduce a proximity coefficient, which is necessary for combining different types of proximity.

Proximity data types enable us to consider value difference between proximities. We in particular assign a coefficient for each proximity data. This can be

done using techniques such as fuzzy logic; logic can decide the weight of specific proximity data types in terms of trust. Thus, several sources of proximity can be combined by a weighted average using their coefficients. Let k_{η_i} be the coefficient for an observed proximity of type η_i. k_{η_i} is calculated for any given η_i by logical aggregation of proximity data type parameters. $K = \{k_{\eta_1}, k_{\eta_2}, k_{\eta_3}, ...\}$ is the set of coefficients for different types of proximity, coefficients are bounded to the range of $[0,1]$. Accordingly, K_B^A is the set of all proximity coefficients between users A and B. An example of three different proximity data types is shown in Table 1.

Table 1. Proximity Data Types

η	p	l	t	d_s	d_d	s	m
η_1	physical	anywhere	anytime	mobile	mobile	bluetooth	duration
η_2	physical	office	working time	mobile	laptop	WiFi	duration
η_3	virtual	anywhere	night	mobile	mobile	SMS	frequency

Given the above types of proximity we define proximity records as:

```
ProximityRecord=<UserID,η, Value>
```

where the `Proximity` tuple is composed by the `UserID`, which is the unique identifier of the observed user, η is the data type of the observed proximity; and `value` is the observed proximity, which is duration or frequency based on the data type. Hence, each user's device is assigned with a set of `Proximity` tuples called *observed set*, as exemplified in Table 2.

Table 2. An example of proximity duration data provided by user A device

UserID	η	Value
B	η_1	200h
C	η_1	20h
D	η_1	80h
E	η_1	30h
F	η_1	0,5h

3   Proximity-Based Trust Initialization

Proximity-based trust initiates trust values between nodes with a one-to-one trust relationship. Using an appropriate conversion method for various proximity data is the most challenging part of trust calculation. As a matter of fact, it is a difficult task to make a conversion from varied types of observed proximity to a range of trust values that are meaningful. For the conversion of proximity records to *Proximity-based trust* values, we use *standard score* formula. Standard score has a normalization effect on the amount of observed proximities according to the observer(user's) average activity.

3.1   Definitions

Normalization has several positive outcomes. First, social interaction quantity varies a lot according to user personality in a social network [11]; as a result the amount of proximity varies substantially for different nodes [6]. Second, there are multiple types of proximity; normalized scoring eases the process of combining trust values according to different proximity data types.

Definition 1 (Standard Score). *A* standard score *[3] indicates how many standard deviations (σ), an observation or datum(x) is above or below the mean(μ):*

$$Z = \frac{datum - mean}{standard\ deviation} = \frac{x - \mu}{\sigma} \tag{1}$$

As a result of using standard score, each peer normalizes trust values based on their average proximity duration with anyone. Therefore, each trust value is unbiased and bounded into a determined range, which hides the effect of variation of proximity duration due to peers having various behavior. Hence, by using standard score formula, we process the observed proximity as follows.

Definition 2 (Observed Set). *The Observed Set (OS) of a user includes all the users that have been detected in proximity of a given user.*

We use a time finite subset of observed proximities for trust evaluation. Therefore, we define the proximity window function as:

Definition 3 (Proximity Window Function). *The proximity window function $P_{\eta_i}^w(A, B)$ accumulates the proximity of user B monitored by user A during the time window $\underline{w}$ and with proximity type η_i.*

Definition 4 (Proximity-based Trust Function). *Proximity-based trust is basically calculated using* standard score formula*. The proximity-based trust function is denoted by $T_{\eta_i}^t$ and is formally defined as:*

$$T_{\eta_i}^t : \mathbb{U} \times \mathbb{U} \to \mathbb{R} \qquad \mathbb{U} : set\ of\ Users$$
$$(A, B) \to T_{\eta_i}^t(A, B) \quad \mathbb{R} : set\ of\ real\ numbers$$

$$T_{\eta_i}^t(A, B) = \begin{cases} -\infty & if(B \notin OS_A) \\ \frac{P_{\eta_i}^w(A,B) - \mu_{\eta_i}^w}{\sigma_{\eta_i}^w} & if(B \in OS_A \wedge t \le w) \\ (1-\alpha) * T_{\eta_i}^p(A, B) + \alpha * \frac{P_{\eta_i}^w(A,B) - \mu_{\eta_i}^w}{\sigma_{\eta_i}^w} & if(B \in OS_A \wedge t > w) \end{cases} \tag{2}$$

where:

- *$T_{\eta_i}^t(A, B)$ is the proximity-based trust value given by user A for user B at the instant t with proximity data type of η_i.*
- *$T_{\eta_i}^p(A, B)$ is the past acquired proximity-based trust value by user A for user B at the instant $p = t - t\%w$ with proximity data type η_i.*
- *$P_{\eta_i}^w(A, B)$ is the cumulative proximity of B in the given period of time w with proximity data type η_i.*

- α is a coefficient which is in range of $]0,1[$ and defines how significant is the impact of new observed proximities on the last calculation of proximity based trust value.
- $\mu_{\eta_i}^w$ and $\sigma_{\eta_i}^w$ are, respectively, the observed period average and the standard deviation during the time window w with proximity data type η_i.

Definition 5 (Proximity-based Trust Aggregation Function). *The proximity-based trust aggregation function is for combining trust values, which is inferred from different proximity data types. It is formally defined as:*

$$T^t : \mathbb{U} \times \mathbb{U} \to \mathbb{R} \qquad \mathbb{U} : set\ of\ Users$$
$$(A,B) \to T^t(A,B) \ \mathbb{R} : set\ of\ real\ numbers$$

$$T^t(A,B) = \begin{cases} -\infty & if(B \notin OS_A) \\[2ex] \dfrac{\displaystyle\sum_{i=1}^{|K_B^A|} k_{\eta_i} * T_{\eta_i}^t(A,B)}{\displaystyle\sum_{i=1}^{|K_B^A|} k_{\eta_i}} & if(B \in OS_A) \end{cases} \tag{3}$$

where $T_{\eta_i}^t$ and k_{η_i} are the trust value and the coefficient for proximity type η_i, respectively K_B^A is the set of coefficients for all the observed proximity types between users A and B. Thus, by using the equation 2, we consider only the proximity that occurs during time window w. Then, for a new time window, the latest assessed trust value $(T_{\eta_i}^p)$ is used to serve as an input for new trust assessment.

3.2 Semantical Trust Inference

Given T^t, we are able to infer trust relationships between users. For the moment we do so by splitting the trust scale equally into four sections with respect to the normal distribution probability density in each **area** and according to the experiments of trust calculation we did using real proximity data from [6] (Illustrated in Figure 2), we define four trust levels and each level includes 25 percent of the observed peers, namely: *Unknown Trust, Slightly trusted, Moderately trusted* and *Highly trusted*. Unknown Trust represents all the persons whose trust (T^t) is assessed into interval $]-\infty, -0.67[$. For this category, we consider that the system has insufficient information to infer trust. All the other that are assessed over -0.67 are considered as trusted entities and they are classified into three other categories: (i) people who may be slightly trusted (i.e., people assessed in range of $[-0.67, 0[$), (ii) people who may be moderately trusted (i.e., ones assessed into the range $[0, 0.67[$) and (iii) people who may be highly trusted (i.e., ones assessed into the range $[0.67, +\infty[$).

Then, according to our classification, which is based on density of peers in each score interval and exprimental results, we consider that a proximity-based trust
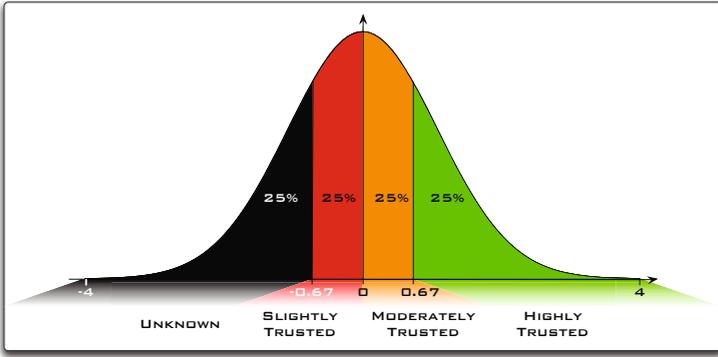
Fig. 2. Trust Scale for normalized Trust values

relationship is established if an inferred proximity-based trust value is higher than -0.67. We define this relation as follows:

Definition 6 (Proximity-based Trust relationship)
Let A and B be two users. If $T(A, B) \geq 0.67$, we say that the Trust relation is verified between A and B. Formally:

$$if\ T(A, B) \geq -0.67\ then\ A \xrightarrow{T} B$$

Example: To explain our approach, we use the following example. User A has a mobile device with a proximity logging application. The observed set of user A is composed by: $OS_A = \{B, C, D, E, F\}$. Table 2 introduced in Section 2 shows an example of cumulative proximity duration that can be provided by user A device with η_1 proximity data type. For instance: $P_{\eta_1}^w(A, B) = 200h$. Considering an unbounded time window (i.e., $w = +\infty$), the proximity duration average is: $\mu_{\eta_1}^w = 66.1$. For calculating the standard deviation, we first compute the difference of each data point from the mean, and then square the result: $(200 - 66.1)^2 = 17929.21$, $(80 - 66.1)^2 = 193.21$. We repeat the same process for the other values. Then, we calculate the standard deviation by dividing the sum of these values by the number of values and take the square root:

$$\sigma_{\eta_1}^w = \sqrt{\tfrac{17929.21 + 2125.21 + 193.21 + 1303.21 + 4303.36}{5}} = 71.90$$

The proximity-based trust value for the user B with 200 hours of proximity is then calculated using Formula: 3, $T_{\eta_i}^t(A, B) = \frac{200 - 66.1}{71.90} = 1.86$

Therefore, a one-to-one trust relation is established between users A and B (i.e., $A \xrightarrow{T} B$), which means that $B \in Tee(A)$ and $A \in Tor(B)$. Moreover, A may highly trust B because $T^t(A, B) \geq 0.67$.

In next section we evaluate our approach experimentally using the MIT real mobility data.

4   Experimental Evaluation

A full-fledged evaluation of our approach needs a large-scale proximity dataset with different data types, in order to have possibilities for combining trust values from different types of observed proximity. Also, for a multi-hop trust estimation, large number of users is needed. That aside, a survey of trust and/or other social facts such as friendship between the observed users is needed to make a comparison between inferred trust values and real social facts. While to the best of our knowledge there is no such kind of vast proximity dataset publicly available, we have used the reality mining dataset[1] [6] as the only existing public dataset of mobile phone proximity records with self-report survey data. The reality mining project was developed by the MIT Media Lab during years 2004-2005 by using 100 Nokia 6600s with Context logging software. They have gathered 330,000 hours of continuous behavioral data logged by the mobile phones of the subjects. They also did a survey in which they asked users about friendship and whom are they going to meet.

To illustrate the capability of our approach, we answer the following question: To what extent the bootstrapped trust values are in correlation with real social facts(e.g. friendship)?

We run the evaluation with the following steps. First, we calculate the proximity-based trust values between users. Then, by comparing the calculated proximity-based trust values of each user to the answers he provided in the survey, we verify if the inferred trust values are coherent with friendship.

Table 3. Average $T^t(A, B)$ value for reported people in survey

Group	Average of minimum trust	Average of trust values
Friends	1.4070	2.0209
inLab	-0.3079	0.7696
outLab	0.0068	1.0460

From the reality mining dataset, we can calculate the proximity duration between two persons which has been detected by bluetooth. We apply the proximity-based trust function (Equation 2) to the proximity durations in order to obtain proximity-based trust values, $T^t(A, B)$, of each user. From the survey, each person predicts his possible future proximity with a friend, or if they are going to meet any other person inside or outside the lab. From this survey, we may tell that mentioned persons are either friends or they are important from the user point of view. We can make the judgment that it is probable that trust relationships exist between the reported users. For these groups (Friends, inLab, outLab), average trust value is shown in Table 3. To find out the relevance of the proximity-based trust values, as we can perceive, highest average of trust values are assigned to friends. Based on the given definitions (Figure 2), friends are

[1] http://reality.media.mit.edu/

Table 4. Average of similarity for reported people in survey

Group	Average of minimum similarities	Average of similarities
Friends	0.2913	0.4828
inLab	-0.0858	0.2520
outLab	0.0089	0.3372

assigned with highly trusted notion. For inLab group, which is the people that users meet inside the MIT Lab, the values are overall located in slightly trusted and moderately trusted classification. For outLab group, which usually consist of friends, family and friends of friends, that a person meets outside of working area, the values are around the barriers of highly trusted group. This experiment shows that trust values are related to the social strength of a relationship. For instance, highest values belong to friends. Additionally, we calculated similarity between users, which is used for the trust transitivity calculation. Table 4 shows that similarity values are behaving very similar to the proximity-based trust values, and they change with the characteristics of relationship. Knowing that similarity is a measure of trust, this arrangement evidences the social fact that friends are similar in their relationships and they are favorite recommender to each other.

The average of minimums is for showing the minimum value that is inferred by a user for each group.

Figure 3 shows the percentage of trust values in each semantical classification of trust, for the trust that is inferred for users in different groups of friends and known people inside and outside MIT Lab. Friends are removed from both
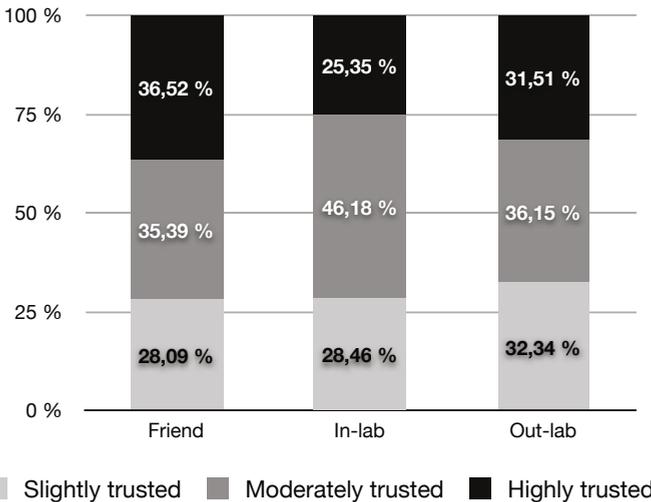

Fig. 3. Trust value distribution in different context of proximity

inLab and outLab groups. As we see, friends have the highest percentage of nodes assessed as highly trusted peers(36%), while inLab peers (e.g. colleagues) are often moderately trusted(46%). For the outLab group, the highly trusted nodes are more than inLab group, but the slightly trusted nodes are increasing. This can be commented by the fact that users meet more random people out of their working place and at the same time more intimate persons are around them than work.

5   Conclusion and Future Work

In this paper, we proposed a novel method to bootstrap trust values from proximity between the people which can also be used for trust inference for unknown users. This approach is suitable for mobile social networking applications. We formalized different types of proximity and introduced proximity data types. The evaluation using real proximity data shows that inferred values are correlated with real social facts. For future work, we aim at creation and evaluation of large dataset of different types of proximity. At the same time, user opinions of trust should be surveyed and included in such kind of dataset. We aim at using fuzzy logic for aggregation of different proximity, according to the level of contribution they can provide to trust value. Also, a large body of work exists in the domain of estimation and recommendation, they may be adapted and evaluated for trust recommendation and transitivity within this approach. Hence, for further evaluation of our approach, and in order to better investigate other available possibilities in trust assessment, we are looking into the deployment of this approach as part of the yarta middle ware framework[2]. Yarta middle ware support the development of mobile social applications.

References

1. Ahamed, S., Hoque, E., Rahman, F., Zulkernine, M.: Towards Secure Trust Bootstrapping in Pervasive Computing Environment. In: 11th IEEE High Assurance Systems Engineering Symposium HASE 2008, pp. 89–96 (2008)
2. Ahamed, S., Monjur, M., Islam, M.: CCTB: Context correlation for trust bootstrapping in pervasive environment. In: IET 4th International Conference on Intelligent Environments, pp. 1–8 (2008)
3. Aron, A.: Statistics for the behavioral and social sciences. Prentice Hall, Englewood Cliffs (1997)
4. Bruneel, J., Spithoven, A., Maesen, A.: Building Trust: A Matter of Proximity? Babson College Entrepreneurship Research Conference, BCERC (2007)
5. Capra, L.: Engineering human trust in mobile system collaborations. SIGSOFT Softw. Eng. Notes 29(6), 107–116 (2004)
6. Eagle, N., Pentland, A.S., Lazer, D.: Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences 106(36), 15274–15278 (2009)

---

[2] https://gforge.inria.fr/projects/yarta/

7. Fukuyama, F.: Trust: the social virtues and the creation of prosperity. Free Press, New York (1995)
8. Golbeck, J., Hendler, J.: Filmtrust: movie recommendations using trust in web-based social networks. In: 3rd IEEE Consumer Communications and Networking Conference, CCNC 2006, vol. 1, pp. 282–286 (2006)
9. Haque, M., Ahamed, S.: An Omnipresent Formal Trust Model (FTM) for Pervasive Computing Environment. In: Proceedings of the 31st Annual International Computer Software and Applications Conference, vol. 01, pp. 49–56. IEEE Computer Society, Los Alamitos (2007)
10. Jøsang, A., Pope, S.: Semantic constraints for trust transitivity. In: APCCM 2005: Proceedings of the 2nd Asia-Pacific conference on Conceptual modelling, pp. 59–68. Australian Computer Society, Inc., Australia (2005)
11. Kumar, R., Novak, J., Tomkins, A.: Structure and evolution of online social networks. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD 2006, pp. 611–617. ACM, New York (2006)
12. Marsh, S.P.: Formalising trust as a computational concept (1994)
13. Newman, M.E.J., Watts, D.J., Strogatz, S.H.: Random graph models of social networks. Proceedings of the National Academy of Science 99, 2566–2572 (2002)
14. Perich, F., Undercoffer, J., Kagal, L., Joshi, A., Finin, T., Yesha, Y.: In reputation we believe: Query processing in mobile ad-hoc networks. In: Annual International Conference on Mobile and Ubiquitous Systems, pp. 326–334 (2004)
15. Quercia, D., Hailes, S., Capra, L.: TRULLO-local trust bootstrapping for ubiquitous devices. Proc. of IEEE Mobiquitous (2007)
16. Saadi, R., Pierson, J., Brunie, L.: T2D: A Peer to Peer trust management system based on Disposition to Trust. In: Proceedings of the 2010 ACM Symposium on Applied Computing, pp. 1472–1478. ACM, New York (2010)
17. Sabater, J., Sierra, C.: Review on Computational Trust and Reputation Models. Artificial Intelligence Review 24(1), 33–60 (2005)
18. Walter, F.E., Battiston, S., Schweitzer, F.: A model of a trust-based recommendation system on a social network. Autonomous Agents and Multi-Agent Systems 16, 57–74 (2008)