Age, Gender and Communication Networks

Alina Stoica[a,b], Zbigniew Smoreda[a], Christophe Prieur[b], Jean-Loup Guillaume[c]
a. SENSe/Orange Labs (*alinamihaela.stoica, zbigniew.smoreda@orange-ftgroup.com*)
b. LIAFA/Université Paris 7 (*prieur@liafa.jussieu.fr*)
c. LIP6/Université Paris 6 (*jean-loup.guillaume@lip6.fr*)

***Abstract*—In this paper, we address some sociological and topological issues associated with mobile phone communication. Based on a dataset of a few million users, we use customers' age and gender information to study relation between these parameters and the average behavior of users in terms of number of calls, number of SMS and calls duration. We also study the dataset from a networking point of view: we define different profiles based on the topological properties of the personal network of each individual and study the relations between these profiles and the age of customers.**

Keywords: mobile phone, age, gender, network structure

I. INTRODUCTION

The ICTs landscape has been entirely changed by the cell phone diffusion. This individual and ubiquitous device, offering voice and text communication features, has transformed the frequency and the geography of communication as compared to older fixed phone practices. We are now virtually always accessible to others wherever we are. Moreover, the mobile phone gives us a direct access to a person: the phone line is no more filtered by household's or bureau's "switchboard" [4]. This offers a useful insight into individual behavior and personal and social network analysis.

The recent possibility to analyze large datasets of behavioral data coming from telecommunication operators gives us the opportunity to revisit some older research on telephone usages. It offers also a new prospect to work on nearly complete interpersonal communication networks. Among an increasing amount of behavioral traces collected by technical systems (internet, mail, IM, SNS…), the interpersonal communication data seem to be the best proxy of social interactions [1,2,3]. Indeed, we usually talk to people with whom we also have many other links, which is not always true in the case of communication with email contacts or SNS "friends". Such datasets therefore open the door to the analysis of close social relationship.

II. DATASET

The raw data analyzed in this contribution – the CDRs (Call Detail Records) – contain all mobile phone exchanges observed in 2006/2007 over a six-month period between Belgian customers of a local mobile operator. After data cleaning the dataset contains 3.3 million users that exchanged over 6 billion calls and short messages (SMS). In addition to communication details (date, hour, duration of call), this anonymous dataset also includes customers' age and gender.

We compared the age and gender distribution of the mobile phone customers in our dataset to the general national population and concluded that there is no systematic bias in operator's customer as regarding these two characteristics (except for people aged over 60 who are underrepresented amongst cellular users).

III. RESULTS

A. *Voice–text usages: a generational transformation?*

The mobile phone diffusion started in the mid-1990. Classically, it first touched the young and wealthy part of the developed countries population before the rapid, massive and nearly universal adoption [5]. From the usage point of view, it means that nowadays only the youngest groups of the population were entered in their "communication age" directly with a cell phone at hand. Hence, it seems interesting to look at some basic indicators of the mobile phone usages by age. Figure 1 shows the average number of calls and SMS, and the mean call duration by age. We observe that the differences in voice call frequency or duration between ages are relatively minor. The main distinction concerns SMS usage: younger users send more SMS than older ones. In the age group 18 to 25 this tendency is really impressive: the SMS is used 4 times more frequently than a conversational exchange!

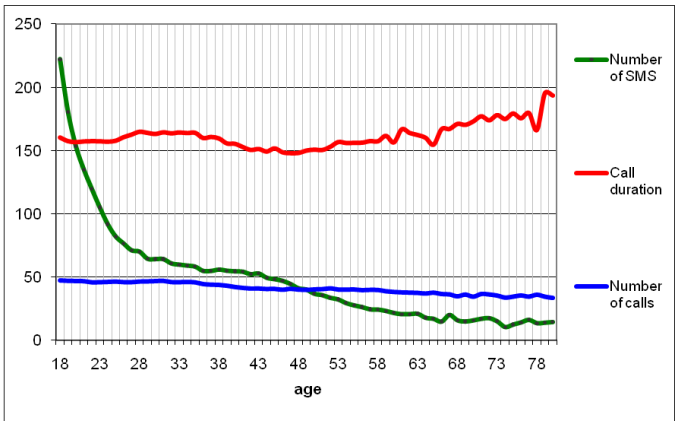

Figure 1. Average monthly number of calls, number of SMS and call duration as a function of phone user's age.

These data show that today interpersonal mobile communication is clearly distributed between voice and text exchanges.

B. Age groups and individual network profiles

To develop these observations, we decided to test whether there is a relationship between the age of a person and the way she is linked to others, or not. We thus analyzed the connections between individuals from the perspective of the communication network they belong to.

We represented the interpersonal communications by an undirected simple graph, where the vertices are the mobile phone customers. Two vertices are connected by an undirected edge if there had been at least one communication in each direction between the corresponding users during each month of the observation. This means that we only consider relatively strong interpersonal links. This gives us a social network with approximately 3 million vertices and 7 million edges that we use in order to characterize customers. Rather than a global approach, we decided to analyze the local structure of the graph around each vertex. More precisely we studied the personal (or ego-centered) network of each vertex (*ego*), i.e., the graph whose vertices are ego's neighbors and whose edges are the edges between the neighbors (note that ego is not included in its ego-centered network).

For every ego we computed several parameters of its ego-centered network: the number of vertices (ego's degree), the number of edges (the edges between ego's neighbors), the number of isolated vertices (the neighbors that are connected only to ego and not to any other neighbor of ego), the number of triangles (a group of 3 interconnected neighbors) and the number of "stars" (a group of 4 neighbors where one of them is connected to the other 3 that are unconnected between them). We use these simple network motives to identify specific individual profiles.

Note that the degree of vertices must be taken into account as the values of the different parameters are biased by it, we chose to compute profiles separately for each degree and distributed the vertices of each degree into 6 profiles defined as follows:

- profile 1: densely connected networks: the number of edges is high and the number of isolated vertices is low;
- profile 2: sparsely connected networks (the opposite situation): the number of edges is low and the number of isolated vertices is high;
- profile 3: mixed situation where there is a densely connected group of neighbors (many triangles) and a sparsely connected one (many isolated vertices);
- profile 4: medially dense networks but with many triangles: these networks do not belong to the first 3 profiles but have a high number of triangles;
- profile 5: medially dense networks but with many stars: these networks do not belong to the first 4 profiles but have a high number of stars;
- profile 6: medially dense networks with no special structure: unclassified vertices.

The question is to determine whether there is a connection between the different profiles and the age of a person or not. To answer this question, we computed, for each age from 18 to 60[1], the probability that an individual of that age belongs to a given profile (see: figure 2). The range of probabilities is different for different profiles which is due to the over-representation of profiles 2 and 6 caused by the heterogeneous distribution of parameters, with many small values. However, there are important differences between these probabilities for different ages. We observe that middle age people (30 to 45) are generally involved in sparser structures when younger and older groups are more densely connected. However, the oldest keep a densely connected group even if they have isolated contacts, while the youngest belong to some cliques (profile 4) or have one or more correspondents who are the "stars" of their ego-centered networks (profile 5).

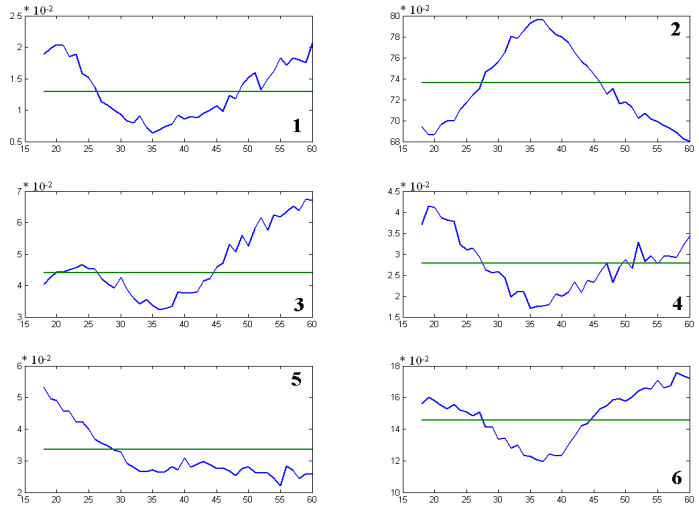

Figure 2. The probability of belonging to the 6 profiles by age (the green line is the average profile probability)

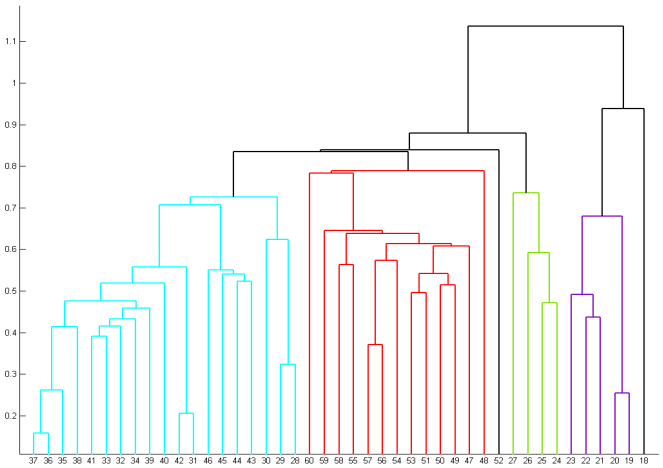

Figure 3. Hierarchical clustering of ages on probabilities in the 6 connectvity profiles

Let us now group together the ages that have similar probabilities for the 6 profiles. A hierarchical clustering was performed on profiles probabilities (see: figure 3). We observe

[1] The persons older than 60 are underrepresented amongst mobile phone users.

that there are 4 main homogeneous age groups very similar to life stages categories [6]: 18-23 (students), 24-27 (young people starting their active life), 28-46 (in couple, usually with children), and 47-60 (at the final stage/end of professional life, children are adult or living apart). Interestingly, here the classification is based exclusively on structural characteristics of their local communication network (where network size effect was neutralized).

C. Gender effect in mobile communication networks

The second personal characteristic of people in our network is their gender. Some years ago, in a French study on the residential use of the (fixed) telephone, the communication of several hundreds of adult men and women has been followed for 4 months using telephone billing records [5]. The study focused on the correlation between the observed duration of phone calls and the gender of callers and receivers. Data have shown that the duration of calls are correlated with the gender of the call receiver and is on average longer when a woman is called. Therefore the reasons why women speak more on the phone [6] seem more related to the gender homophily in telephone networks than to "their intrinsic tendency to talk." An in-depth Conversation Analysis work on a recorded telephone talk's dataset [7] has suggested that politeness rules governing the telephone call can explain in part why it is the gender of receiver that has the biggest effect on how the call is managed and on its overall duration. The conversations involving women tended to go through longer introductive sequences, to be more multi-thematic and digressive in nature with a corresponding lengthening and multiplication of closure sequences; and the conversations with men had a tendency to be linear and monothematic. In summary, the callers seem to adjust their interaction style to the gender of the receiver.

Ten years after, the mobile usages still fit this gender communication pattern. As we can see in the figure 4, mobile phone calls towards a woman are, in average, longer than calls to a man, whatever caller gender is.

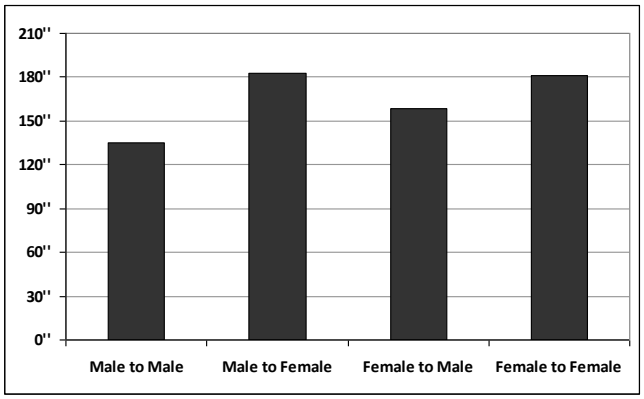

Figure 4. Mean call duration (in seconds) according to call initiator and receiver gender

To go further, we isolated all mixed-gender two-way communication pairs in our network and calculated average durations of call between them. We obtained 171 seconds when a male calls a female and 162 seconds when a female calls a male. As a consequence, it seems that we do not face a technological artifact but a more general social interaction pattern.

The personal networks composition was also scrutinized looking at the fraction of men in ego communication network (see figure 5). We observed that the overall gender homophily in communication networks evolves with age for men and, less, in women's networks. The life cycle transitions modify sociability patterns—from external to the household contacts for young people to more and more family-oriented links for older individuals [10]—and influences the shape of gender relations. In fact, as domestic and familial spheres are still associated with the role of woman, with age the male's network starts to be populated by females. At the end of lifecycle, there are more women than men in the man's mobile phone directory.

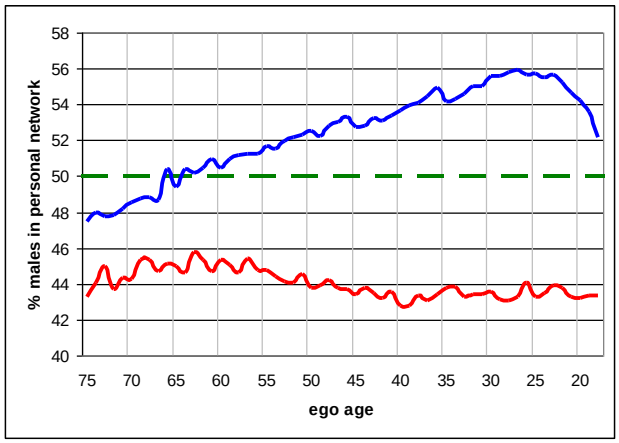

Figure 5. Fraction of males in the communication network of females (red) and males (blue) by age

Going back to communication practices, we can speculate about their hypothetical transformation. Of course, we do not have long-term time series, thus our interpretation remains tentative. However, some tendencies (see: figure 6) indicate differentiation of gendered mobile usages by age. The SMS usages seem to be more "feminine" in general (fig. 6-a): for younger part of population (aged 18-25), we also notice that the between-gender "texting" is particularly popular. And, in fact, it develops at the expense of voice calls (fig. 6-c); the number of voice communication is going down in young adults. The duration of calls varies less, but for the young adults it diminishes sharply for same gender calls. The mixed-gender conversation length remains at the level of other age groups.

We can consider that in younger generations the mobile phone appropriation was deeper as it includes both communication functions offered by cell phones: text and voice. Some authors indicate that heavy SMS use in youngster's relation with other gender is related to seduction tactics where a direct voice contact can be more "risky" for interlocutors [11]. Anthropologists also emphasized the propensity of girls to write personal diaries, letters, etc., as well as women responsibility in familial correspondence [12]. The changing balance between voice and text in general and in between-gender communications can be in part a reflection of

new form of accommodation to a new communication channel offered by a popular technology.

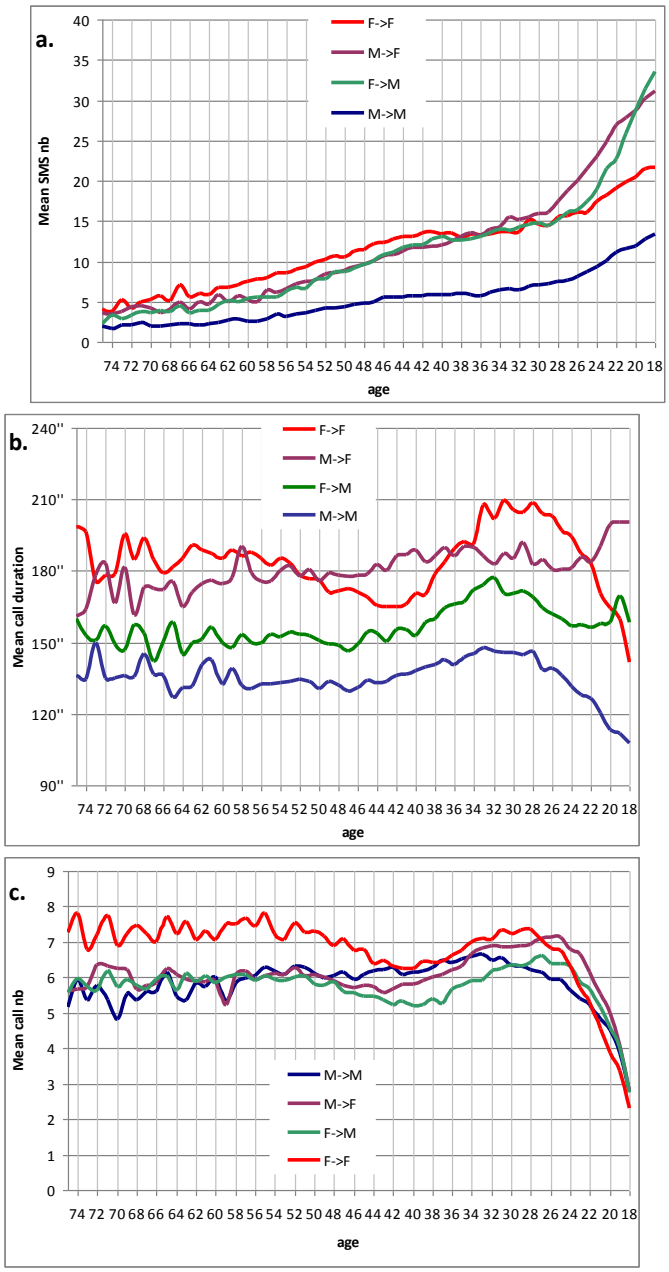

Figure 6. Mobile communication practices by age and gender: (a) mean monthly SMS number, (b) average call duration, (c) mean monthly number of calls.

IV. CONCLUSION

We presented here a stage report of an ongoing research. Based on CDR of a mobile phone operator, we studied the relations between phone usages and networks properties of customer's and their age and gender. We have shown that some already known sociological results are still valid in the context of mobile phone. In particular, there are very different behavior depending on the age of the users and their gender: the gender of the receiver of phone calls is strongly correlated with the duration of the call, while the gender of the caller is less important.

We also built ego-centered networks for each customer and tried to find some correlations between the topological structure of their neighborhood and their behavior. Automatic profiling based only on simple topological properties yielded groups of users which correspond to life stages categories.

Much work remains to be done in this context, in particular more complex network properties and their correlations with age or gender could be studied. We could also propose models taking such behaviors into account.

V. REFERENCES

[1] J-P.Onnela et al., "Analysis of a large-scale weighted network of one-to-one human communication", New Journal of Physics, vol. 9, 179, 2007.

[2] C.A.Hidalgo, C.Rodriguez-Sickert, "The dynamics of a mobile phone network", Physica A, vol. 387, pp. 3017–3024, 2008.

[3] R.Lambiotte et al., "Geographical dispersal of mobile communication networks", Physica A, vol. 387, pp. 5317-5325, 2008.

[4] R.Ling, The Mobile Connection: The Cell Phone's Impact on Society, San Francisco, Morgan Kaufman Publ., 2004.

[5] See: http://www.itu.int/ITU-D/ict/statistics/index.html

[6] N.Shulman, "Life-cycle variations in patterns of close relationship", Journal of Marriage and Family, vol. 37, pp. 813–821, 1975.

[7] Z.Smoreda, C.Licoppe, "Gender-Specific Use of the Domestic Telephone", Social Psychology Quarterly, vol. 63, pp. 238-252, 2000.

[8] L.R.Rakow, Gender on the Line. Women, the Telephone, and Community Life, University of Illinois Press, Chicago, 1992.

[9] R.Akers-Porrini, "Efficacité féminine, courtoisie masculine: la durée inégale des appels téléphoniques", Réseaux, vol. 18(103), pp. 143-182, 2000.

[10] A.Degenne, M.Forsé, Introducing Social Networks, London Sage, 1999.

[11] R.Ling, B.Yttri, 2003. "Kontroll, frigjøring og status: Mobiltelefon og maktforhold i familier og ungdomsgrupper." in F.Engelstad, G.Ødegård (eds.) *På terskelen: makt, mening og motstand blant unge*, Oslo: Gyldendal Akademisk (english version: http://www.richardling.com/papers/2004_Control_Emancipation_and_status.pdf)

[12] D.Fabre (ed.), Ecritures ordinaires, Paris, P.O.L, 1993.