EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE

Open Access

Quantifying gender preferences in human social interactions using a large cellphone dataset

Asim Ghosh[1], Daniel Monsivais[1], Kunal Bhattacharya[1][*] , Robin I.M. Dunbar[2] and Kimmo Kaski[1]

[*]Correspondence:
kunal.bhattacharya@aalto.fi
[1]Department of Computer Science,
Aalto University School of Science,
Espoo, Finland
Full list of author information is
available at the end of the article

Abstract

In human relations individuals' gender and age play a key role in the structures and dynamics of their social arrangements. In order to analyze the gender preferences of individuals in interaction with others at different stages of their lives we study a large mobile phone dataset. To do this we consider four fundamental gender-related caller and callee combinations of human interactions, namely male to male, male to female, female to male, and female to female, which together with age, kinship, and different levels of friendship give rise to a wide scope of human sociality. Here we analyse the relative strength of these four types of interaction using call detail records. Our analysis suggests strong age dependence for an individual of one gender choosing to call an individual of either gender. We observe a strong bonding with the opposite gender across most of their reproductive age. However, older women show a strong tendency to connect to another female that is one generation younger in a way that is suggestive of the *grandmothering effect*. We also find that the relative strength among the four possible interactions depends on phone call duration. For calls of medium and long duration, opposite gender interactions are significantly more probable than same gender interactions during the reproductive years, suggesting potential emotional exchange between spouses. By measuring the fraction of calls to other generations we find that mothers tend to make calls more to their daughters than to their sons, whereas fathers make calls more to their sons than to their daughters. For younger callers, most of their calls go to the same generation contacts, while older people call the younger people more frequently, which supports the suggestion that *affection flows downward*. Our study primarily rests on resolving the nature of interactions by examining the durations of calls. In addition, we analyse the intensity of the observed effects using a score based on a null model.

Keywords: Social networks; Egocentric networks; Mobile phones; Life history; Gender differences; Sex differences

1 Introduction

In social interactions between humans, gender and age play a key role in the communities and social structures they form and the dynamics therein. For the caller–callee interactions in mobile communication there are four fundamental possibilities, namely male to male, male to female, female to male, and female to female, which together with age, kinship, and different levels of friendships affect the strengths of social interactions, giving

rise to a wide scope of human sociality. The studies of primate brain size and its relation to their average social group size suggest that humans are able to maintain of the order of 150 stable relationships (Dunbar number) [1–3]. In addition the Social Brain hypothesis suggests that on the basis of emotional closeness human social networks can be divided into four cumulative layers of 5, 15, 50 and 150 individuals, respectively [4]. The concept of emotional closeness is, in general, hard to quantify, but previous studies have shown how it can be associated with the frequency of communication between two individuals [5, 6]. This makes the concept quantifiable such that one can observe how much an individual shares social resources with his or her contacts of different gender and age.

Over the past decade or so, much research on human communication patterns has been done by using "digital footprints" data from modern communication technologies such as mobile phone calls and text messages as well as social media like Facebook, and Twitter [7–9]. Of these the mobile phone communication data of call detail records (CDRs) has turned out to help us in getting insight into the structure and dynamics of social networks, human mobility and behavioural patterns in much finer details than before [7]. It has also revealed how microscopic properties related to individuals translate to macroscopic features of their social organization such as networks. As a result of these studies we now have quite a good understanding of a number of structural properties of human social networks such as degree, strength, clustering coefficient, community structure, and motifs [10–12].

Apart from these basic structural properties of networks, more recent studies have given us insight into a number of other aspects of social networks, namely their dependence on temporal, geographic, demographic, and behavioral factors of individuals in the network [13–17]. One such observation pertains to the shifting patterns of human communication across the reproductive period of their lives, which appears to reflect parental care [18, 19]. Another is a study using the postal code information in the data to show that the tie strength is related to geographical distance [20]. In addition, it has been shown that there is a universal pattern of time allocation to differently ranked social contacts [21]. Finally, recent studies indicate variation in connections and the number of friends with the age and gender [22, 23]. The importance of the strength and significance of communication with top-ranked contacts have also been studied in detail [18, 22].

In the present study, we focus on measuring the relative strengths of the four possible pairwise caller–callee interactions over their lifespans as a function of the caller's age. From the point of view of call initiation, we find that females play a more active role during their reproductive years as well as during their grandmothering period [24, 25]. The grandmothering hypothesis is usually studied in the context of human longevity and evolutionary benefits. The notion deals with the focus of post-menopausal on their grandchildren. In general, the social focus of women are known to shift from the opposite gender in the same age cohort, when they are young, to the age cohort of their children, as they grow older. We observe that while females of grandmothering age are found to give more attention to their children, males up to the age of 50 years still keep stronger connection with their spouses of slightly younger age. Furthermore, the fraction of calls to individuals of different generations indicates that mothers tend to call their daughters more than their sons, whereas fathers call their sons more than their daughters. For younger individuals, most of their calls go to contacts of the same generation, whereas older people call younger people more frequently. The calling activity of older adults with the younger individuals who are below or around their reproductive age would signify parental and alloparental

care, that is, caring for the children of children. We group these kind of behaviour as *affection flows downward.*
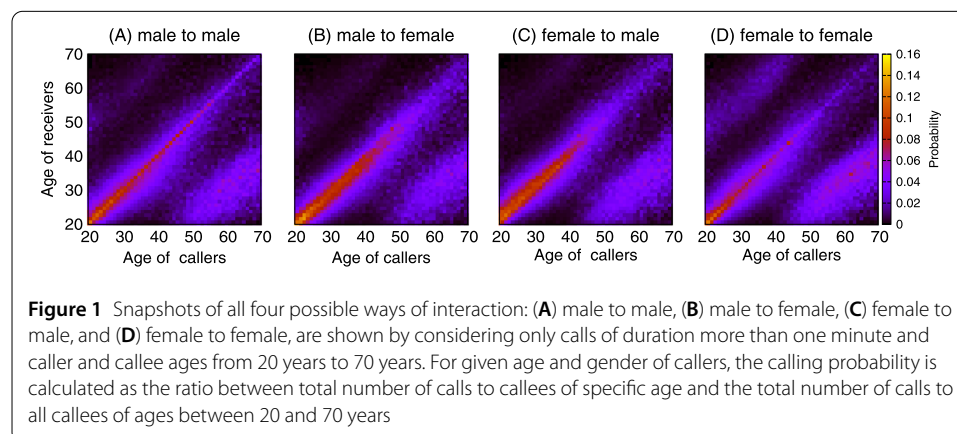
2 Methodology

In this study we analyse mobile phone communication records of a particular European mobile service provider containing time series of call detail records or CDRs of caller–callee pairs. This dataset also includes demographic information such as age and gender of the callers. By using the gender information we measure the relative strengths for the four basic calling pattern such that we count the total fraction of calls for the caller–callee pairs of the same or of different genders by assuming a cut-off for the minimum call duration. We analyse all the CDRs for the year 2007 on a month-by-month basis for more than 2.4 million subscribers where both the caller's and the callee's demographics are known, totaling over 30 million calls. Since datasets of this kind are susceptible to error due to multiple subscriptions, we filtered out customers who have multiple subscriptions under the same contract number. Our study based on CDRs allows obtaining anonymized data from a very large population, but is in contrast to small-scale studies where volunteers are recruited and cross-validation of results is possible by collecting information from the participants through questionnaires [21, 26].
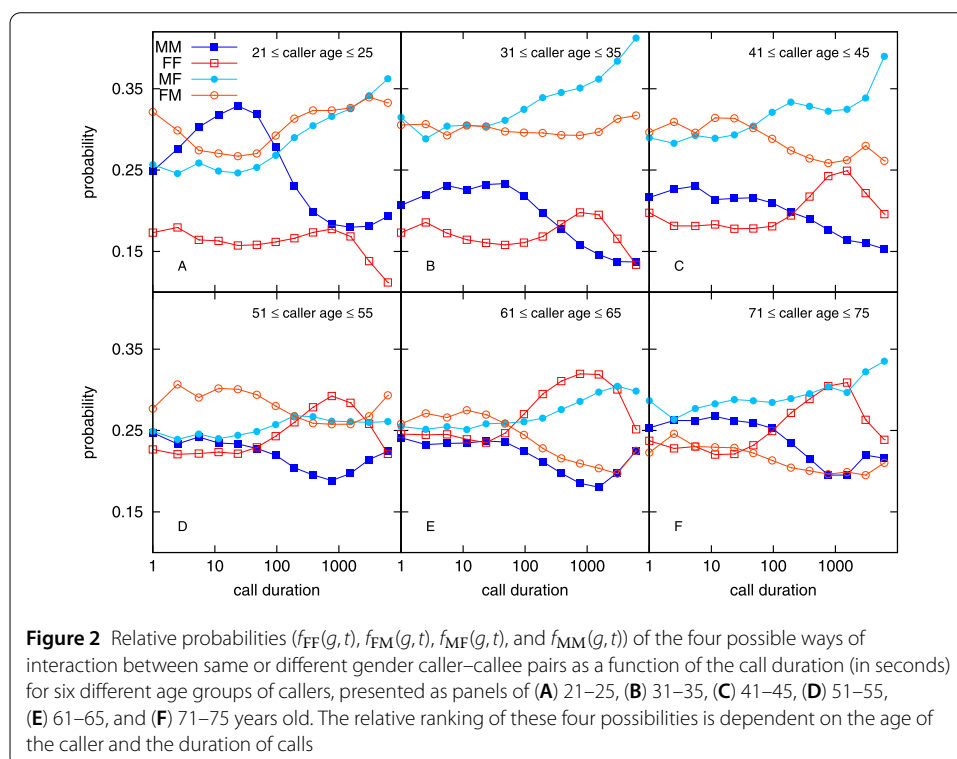
3 Results

In order to analyze the gender preferences of individuals in interaction with others at different stages of their lives we choose the age of the caller and count the total number of calls within a time window. We apply a threshold for minimum call duration, such that calls shorter than the threshold value are considered not to be indicative of emotional closeness while calls longer than that are taken to indicate a meaningful emotional or social exchange relationship between the caller and the callee. Then we calculate the relative probabilities for the four possible types of caller to callee interaction. As it is difficult to decide a priori where the borderline between meaningless and meaningful is, we will vary the threshold for minimum call duration in measuring how the probabilities of the four ways of interaction vary with age and gender of the callers.

In Fig. 1 we show snapshots of the four interaction types for a call duration threshold of one minute. By considering callers and callees between 20 and 70 years of age we have calculated the calling or interaction probabilities for the same and different gender pairs.

Figure 1 Snapshots of all four possible ways of interaction: (**A**) male to male, (**B**) male to female, (**C**) female to male, and (**D**) female to female, are shown by considering only calls of duration more than one minute and caller and callee ages from 20 years to 70 years. For given age and gender of callers, the calling probability is calculated as the ratio between total number of calls to callees of specific age and the total number of calls to all callees of ages between 20 and 70 years
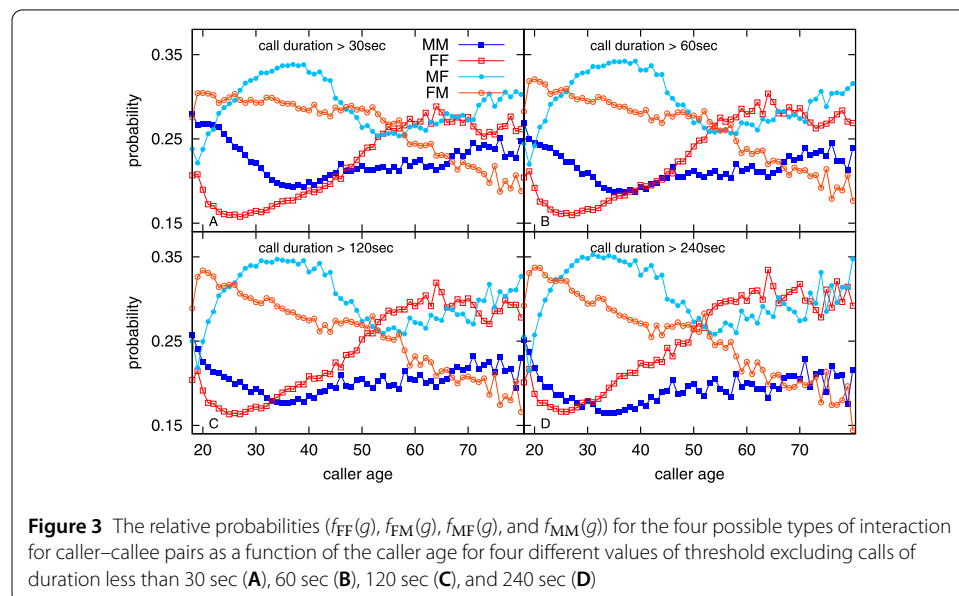
Here the probability is determined as the ratio between the total number of calls to the specific age callee and the total number of calls to all the callees within the age range of 20 to 70 years. From these communication patterns, the signature of a generation gap becomes evident. For example, the calling pattern between two females exhibits a rather clear signature of being triple lobed with the side lobes separated from the same age group center lobe by a generation gap up and down. This is indicative of frequent interactions between mothers and their daughters over the two generation gaps.
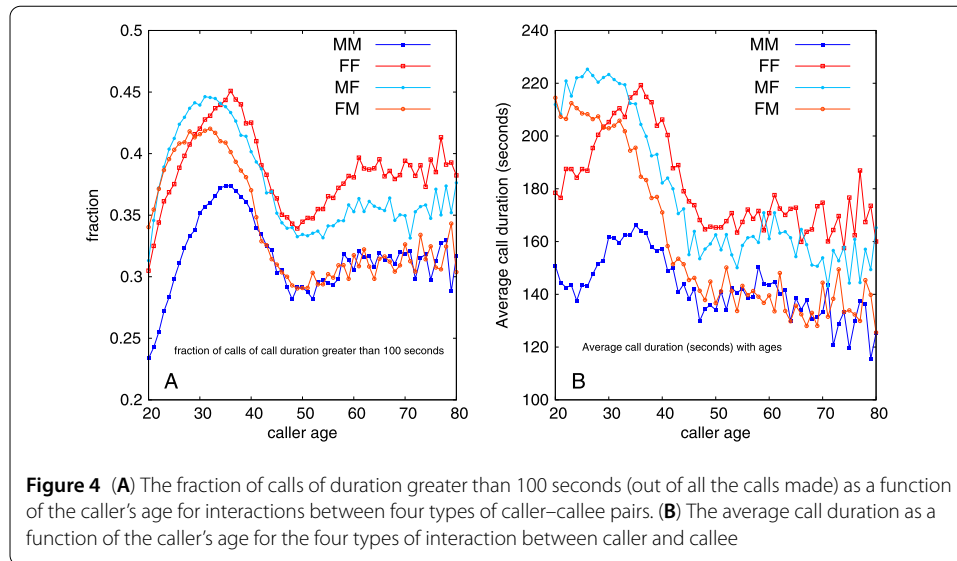
For callers with age g, we first divide the (outgoing) calls into four sets according to the genders of the caller–callee pairs, namely, FF, FM, MF and MM, where 'M' denotes males and 'F' denotes females. Then we further divide the sets by the duration of the calls (t, measured in seconds). For a given duration t, we calculate the relative probabilities among the four possible caller–callee pairs such that, $f_{FF}(g, t) + f_{FM}(g, t) + f_{MF}(g, t) + f_{MM}(g, t) = 1$. In case female-female pairs, $f_{FF}(g, t)$ = (total numbers of outgoing calls of duration t from female callers with age g to female callees)/(total number of outgoing calls of duration t from any caller (male or female) with age g), and likewise for other types of pairs. In Fig. 2 we show these probabilities as a function of the call duration t for different age groups of callers, i.e. 21–25, 31–35, 41–45, 51–55, 61–65, and 71–75 years. We find that the relative ranking between them is strongly dependent on call duration. At younger ages (21–25 years), the MM calls tend to be relatively short, with interactions peaking around 10 secs and being of the highest rank up to 100 secs then decaying, suggesting that these calls are concentrated on their same gender friends. However, as men age, they get married and change their interaction preference to their opposite gender partners (see the panels for the age groups of 31–35 and 41–45 years). At the same time the distribution of call duration becomes flatter making the average call duration longer, a trend also evident among the older age groups (51–55, 61–65, and 71–75 years). On the other hand, the ranking for

Figure 2 Relative probabilities ($f_{FF}(g, t)$, $f_{FM}(g, t)$, $f_{MF}(g, t)$, and $f_{MM}(g, t)$) of the four possible ways of interaction between same or different gender caller–callee pairs as a function of the call duration (in seconds) for six different age groups of callers, presented as panels of (**A**) 21–25, (**B**) 31–35, (**C**) 41–45, (**D**) 51–55, (**E**) 61–65, and (**F**) 71–75 years old. The relative ranking of these four possibilities is dependent on the age of the caller and the duration of calls

the FF calls tend to be rather low for all the age groups up to a call duration of 100 secs. The distribution of call duration is initially quite flat and small in value, but it starts increasing at about the age of women bear their first child, peaking at around 1000 secs. This suggests frequent interactions between the daughter and her mother, and seems to indicate that the grandmothering effect has set in. As for opposite gender pairs, we find that below the age of 35 years, the FM and MF interactions show quite high values for medium to high call duration. This can be interpreted as an indication of strong bonding between spouses. But with age, the FM-interactions start decreasing while MF-interactions increase, thus showing inverse relationship from the age of 40 years onward for medium to high call duration. This observation suggests that as women age they shift their attention from their spouses to their children.

Next, aggregating calls over the different durations we calculate the relative probabilities as functions of age g, such that for a given g, we have, $f_{FF}(g) + f_{FM}(g) + f_{MF}(g) + f_{MM}(g) = 1$, where $f_{FF}(g)$ = (total numbers of outgoing calls from female callers with age g to female callees)/(total number of outgoing calls from any caller (male or female) with age g), and likewise. In Fig. 3, we depict these relative probabilities for the four caller–callee interaction categories as functions of the caller's age, for call durations greater than 30 sec, 60 sec, 120 sec, and 240 sec, respectively. The probabilities are rather stable when the calls of very low duration are filtered out. If we concentrate only on the behavior observed for threshold values of 120 sec and 240 sec (see the two bottom panels), our observations are as follows: For individuals older than 30 years, MM interactions become less frequent, which can probably be attributed to men getting married and thus giving priority to their opposite gender spouses over the same gender friends. This picture is also supported by observing the age-wise variation of the MF-interactions, where we see that up to the age of 45 years men call their spouses more than they call others. However, MF-interactions also show a minimum around the age of 50 years after which they start increasing again from the age of 55 years on. This may be attributed to men's more frequent interactions with women one generation younger which corresponds to the age cohort of the daughters. On the other hand, the FF-interaction curve starts from a low value at about 27 years
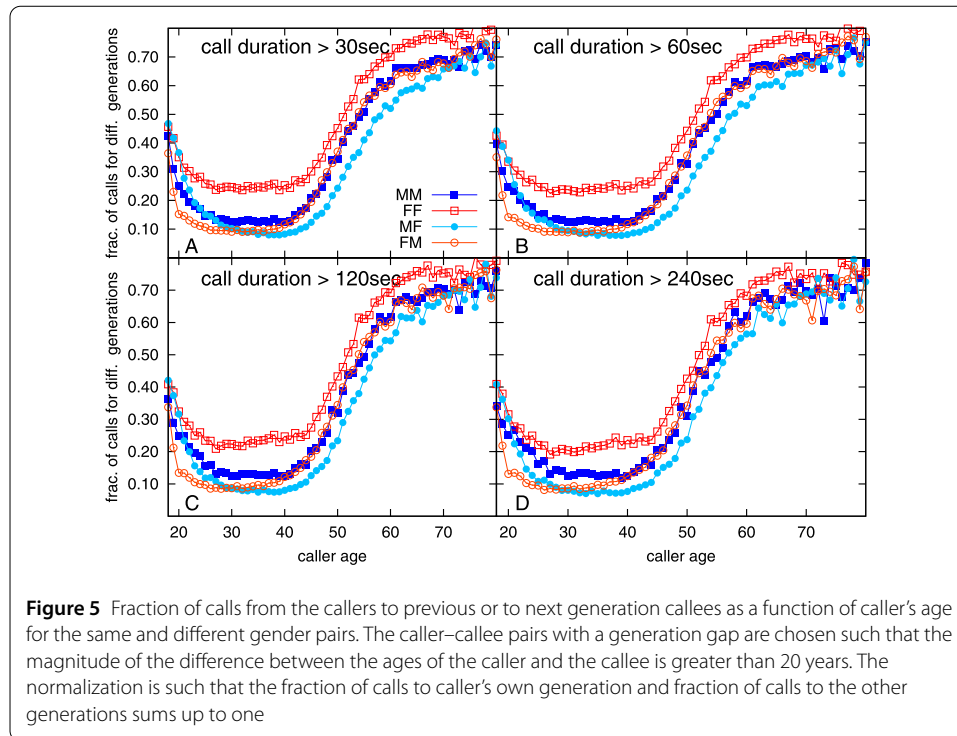

Figure 3 The relative probabilities ($f_{FF}(g)$, $f_{FM}(g)$, $f_{MF}(g)$, and $f_{MM}(g)$) for the four possible types of interaction for caller–callee pairs as a function of the caller age for four different values of threshold excluding calls of duration less than 30 sec (**A**), 60 sec (**B**), 120 sec (**C**), and 240 sec (**D**)

Figure 4 (**A**) The fraction of calls of duration greater than 100 seconds (out of all the calls made) as a function of the caller's age for interactions between four types of caller–callee pairs. (**B**) The average call duration as a function of the caller's age for the four types of interaction between caller and callee

of age, after which it shows a steadily increasing trend. This observation indicates again that before marriage, females call less frequently to other females. After the age of 27, the FF-interaction curve grows rapidly up to the age of about 65 years. This behavior lends support once more to the grandmothering effect. Finally, the curve for FM-interactions indicates that after the age of 35 years, the focus of women on their spouses starts progressively decreasing. A similar observation also presents itself when we consider only top-ranked calls (ranked by their call duration) as shown in the Appendix (see Fig. 7).

In Fig. 4(A), the fraction of calls having duration greater than 100 seconds (out of all the calls made) is shown as a function of the caller's age. Here, the fraction of longer calls for the four different pairs of interactions all peak around callers aged 30 years, after which the interactions decrease till about 50 years of age, followed by an increase till about 60 years of age, at which point the interactions seem to plateau. It should be noted that, for the FF curve, the increase from 50 to 60 years of age can again be taken as clear evidence of grandmothering. In Fig. 4(B), we measure the average call duration as a function of caller's age for the four different types of interactions of the same or different gender pairs. From the MM curve, it is evident that the average call duration for male-to-male calls is low throughout their lifespan. The FM and MF curves show that at younger ages (i.e. before marriage) both male-to-female and female-to-male participate in long phone calls. But after typical marrying age for this population (27 years, as indicated in the national statistics), call duration drops significantly. The FF curve shows that initially the fraction increases with age (up to the age of 40 years), then rapidly falls. It is nevertheless clear that after the age of around 35 years, the call duration for female-to-female calls is the highest among the four possible types of interaction, which again can be interpreted as a signature of the grandmothering effect.

In Fig. 5, we show the fraction of outgoing calls from a caller to a callee who is either one generation older or one generation younger. The caller–callee pairs with a generation gap are chosen such that the magnitude of the difference between the age of the caller and the age of the callee is greater than 20 years. Here we observe that FF-interactions always have the highest value for any age, which can be taken as evidence of a large amount of communication between mothers and their daughters. Before the age of 27 years (the

Figure 5 Fraction of calls from the callers to previous or to next generation callees as a function of caller's age for the same and different gender pairs. The caller–callee pairs with a generation gap are chosen such that the magnitude of the difference between the ages of the caller and the callee is greater than 20 years. The normalization is such that the fraction of calls to caller's own generation and fraction of calls to the other generations sums up to one

average age of marrying in this population), measurement of MF interactions indicates that sons are also strongly attached to their mothers. After the age of 40 years, the MF and FM interactions are very close to each other, suggesting that sons get the same amount of attention from both parents. On the other hand, the tie strength between fathers and their sons are reflected in the curve for MM interactions, which show a similar trend as the other interaction types. Notice that female callers are, relatively speaking, closer to the other generation than the same age male callers. In addition, the calling patterns of older people suggest that sons and daughters get different amounts of attention from their parents. In other words, from the mothers' point of view, daughters get more attention than the sons, while sons get more attention from their fathers than daughters do. We find that, at younger ages, the fraction of calls going from one generation to another is around 10% to 30% of the total number of calls. On the other hand, when the age of the callers reaches 60 years, they are found to mostly communicate with their children (ranging from 50% to 70%), which supports the claim that affection flows downward. A similar pattern emerges from an analysis of just the top ranked calls, as elaborated in the Appendix, see Fig. 8.

4 Quantification using a null model

In order to benchmark our results we present a null model [27, 28] with the assumptions that the communication events are established at random with respect to gender and that the communication volume is proportional to the populations of caller and callee. We focus on the results shown in Fig. 2 and Fig. 3, where the deviations of the observed probabilities, like $f_{FF}(g, t)$ and $f_{FF}(g)$ with respect to the null model results provide a quantification of the amount of non-randomness. The model allows us to calculate a set of expectation values and standard deviations for the number of outgoing calls between different gender groups of the given age of callers. From the null model we obtain the set of probabilities

$\{p_{\mathrm{FF}}(g), p_{\mathrm{FM}}(g), p_{\mathrm{MF}}(g), p_{\mathrm{MM}}(g)\}$, where, $p_{\mathrm{FF}}(g) + p_{\mathrm{FM}}(g) + p_{\mathrm{MF}}(g) + p_{\mathrm{MM}}(g) = 1$. Here $p_{\mathrm{FF}}(g)$ denotes the probability of an outgoing call from a female caller of age g to a female callee, irrespective of the age of the latter. The probabilities for the other three gender pairs are similarly defined. We assume for a given age g, that the set of probabilities comprise of a multinomial distribution that signifies random reshuffling of the total number of outgoing calls across the four types of gender-based interactions. If $C_{\mathrm{tot}}(g, t)$ is the total number of outgoing calls of duration t from callers with age g found in the dataset, then the expectation value and the corresponding variance of the number of outgoing calls of duration t from female callers with age g to female callees are,

$$\langle C_{\mathrm{FF}}(g, t) \rangle = C_{\mathrm{tot}}(g, t) p_{\mathrm{FF}}(g); \qquad \sigma_{\mathrm{FF}}^2(g, t) = C_{\mathrm{tot}}(g, t) p_{\mathrm{FF}}(g) \big(1 - p_{\mathrm{FF}}(g)\big). \tag{1}$$

We obtain the probabilities for outgoing calls from one group to another following the calculation of edge probabilities in the configuration model of random graphs [29] such that the number of possible interactions between two sets of individuals is assumed to be the product of the populations of these sets. For instance, the probability $p_{\mathrm{FF}}(a) = k[n_{\mathrm{F}}(g)\{n_{\mathrm{F}}(g) - 1\} + n_{\mathrm{F}}(g)\{n_{\mathrm{F,tot}} - n_{\mathrm{F}}(g)\}]$, where $n_{\mathrm{F}}(g)$ is the number of female subscribers with age g, $n_{\mathrm{F,tot}}$ is the total number of female subscribers, and k is a proportionality constant that is independent of both the age and gender. The first term inside the brackets denotes the possible interactions between the females with age g and the second term accounts for the interactions with the females of other ages. By simplifying the expression and normalizing the probabilities we obtain

$$p_{\mathrm{FF}}(g) = \frac{n_{\mathrm{F}}(g) n_{\mathrm{F,tot}}}{\{n_{\mathrm{F}}(g) + n_{\mathrm{M}}(g)\}\{n_{\mathrm{F,tot}} + n_{\mathrm{M,tot}}\}}. \tag{2}$$

Note that it is also possible to obtain a further generalized form for the above probability, namely $p_{\mathrm{FF}}(g, t)$ that is dependent on the call duration t. This can be calculated by solely taking into account the subscribers who participated in calls of duration t. However, in the current scheme we do not include this dependence.

With $C_{\mathrm{FF}}(g, t)$ being the number of outgoing calls from female subscribers with age g to other females, we first calculate a scaled deviation or the Z-score given by, $Z(g, t) = \{C_{\mathrm{FF}}(g, t) - \langle C_{\mathrm{FF}}(g, t) \rangle\}/\sigma_{\mathrm{FF}}(g, t)$ [30]. This score scales the deviation of the actual number of calls from the expected number of calls in terms of the standard deviation as shown in Eq. (1). The expression for Z can also be written in the form: $Z(g, t) = \sqrt{C_{\mathrm{tot}}(g, t)}\{f_{\mathrm{FF}}(g, t) - p_{\mathrm{FF}}(g)\}/\sqrt{p_{\mathrm{FF}}(g)(1 - p_{\mathrm{FF}}(g))}$. The last expression shows that $Z(g, t)$ is the scaled deviation of the observed probability $f_{\mathrm{FF}}(g, t)$ shown in Fig. 2. However, $Z(g, t)$ gets amplified by the number of calls $C_{\mathrm{tot}}(g, t)$ which depends on the volume of calling having different durations as well as the volume of calling which differs across the ages. Therefore, to quantify the non-randomness that is independent of the volume of calling we use the following normalized score, which is similar to a measure for the effect sizes [31],

$$\text{Normalized score} = \frac{f_{\mathrm{FF}}(g, t) - p_{\mathrm{FF}}(g)}{\sqrt{p_{\mathrm{FF}}(g)(1 - p_{\mathrm{FF}}(g))}}. \tag{3}$$

In Fig. 6(A) we plot the normalized scores corresponding to Fig. 2. The probabilities in the null model are provided in Table 1. A comparison between Fig. 2 and Fig. 6(A) reveals
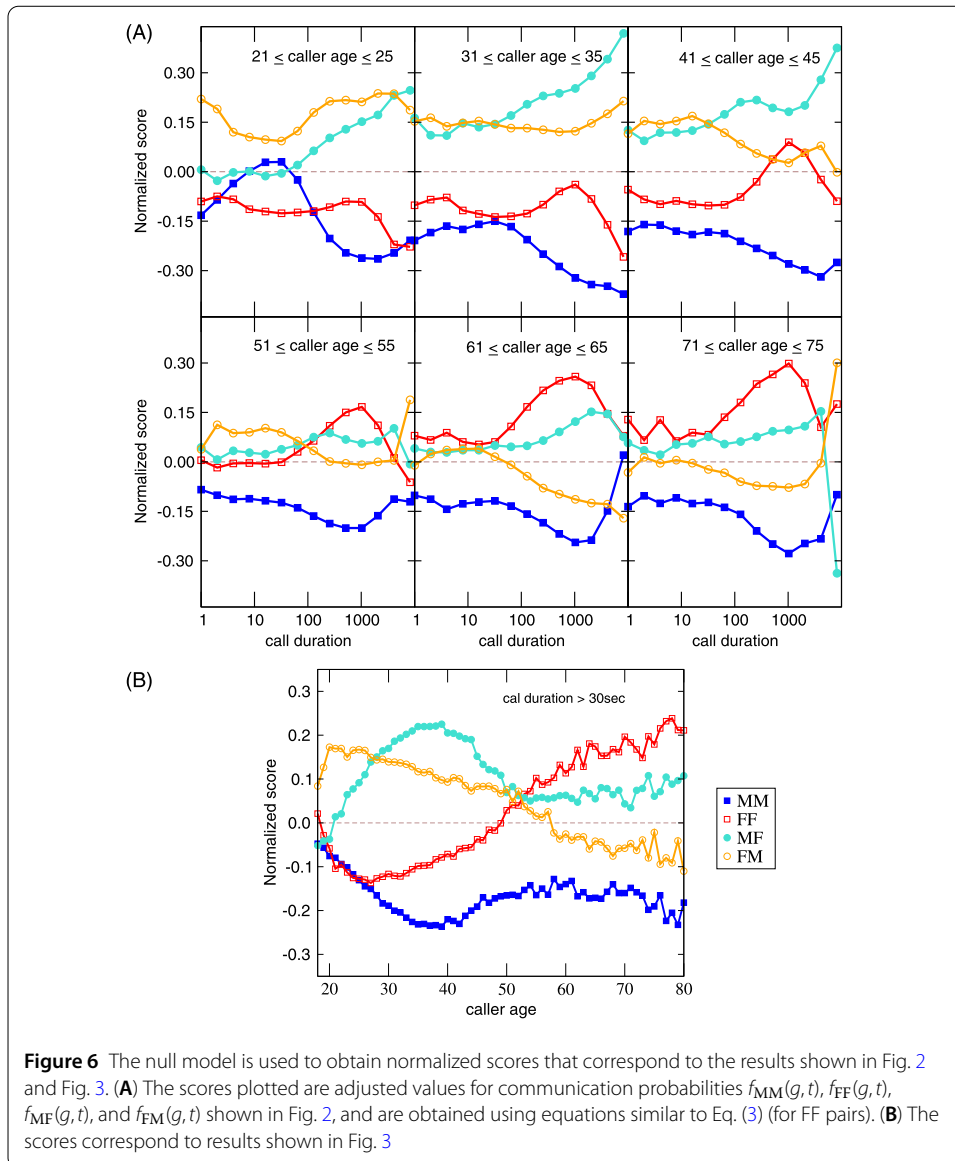
Figure 6 The null model is used to obtain normalized scores that correspond to the results shown in Fig. 2 and Fig. 3. (**A**) The scores plotted are adjusted values for communication probabilities $f_{MM}(g,t)$, $f_{FF}(g,t)$, $f_{MF}(g,t)$, and $f_{FM}(g,t)$ shown in Fig. 2, and are obtained using equations similar to Eq. (3) (for FF pairs). (**B**) The scores correspond to results shown in Fig. 3

Table 1 The probabilities used in the null model that are used to calculate the normalized scores corresponding to the results in Fig. 2. The values shown below correspond to the age brackets used in Fig. 2. The probabilities for the different pairs are calculated using equations similar to Eq. (2) (for FF pairs)

Age range	p_{FF}	p_{FM}	p_{MF}	p_{MM}
21–25	0.207	0.241	0.255	0.297
31–35	0.213	0.247	0.250	0.291
41–45	0.214	0.249	0.248	0.289
51–55	0.225	0.262	0.237	0.276
61–65	0.220	0.256	0.242	0.282
71–75	0.195	0.227	0.267	0.311

the following. First, a demotion of the MM calling probabilities in terms of the normalized score. The values of $f_{MM}(g,t)$ that are observed in Fig. 2 appear to be enhanced by the presence of larger number of male subscribers. In fact, the overall scores corresponding

to the MM pairs across different call durations and ages are negative, implying that the MM communication is lower than what is expected in the null model. Interestingly, our conclusions regarding the importance of short duration calls for the MM pairs for younger callers (aged 21−25) is still supported as the peak of the MM curve crosses over to a positive value in Fig. 6(A). For the MF pairs in Fig. 6(A), similar to Fig. 2, the patterns are largely unchanged across different age ranges of the caller and the score remains mostly positive. The curves for the FM and FF pairs appear to be contrasting cases. With the increase in the age of the caller, the FM scores show an overall decrease moving from positive to negative values, while the FF scores change from negative to positive. For older females, the effects is the strongest for calls at larger duration, where the FF calling appears to be higher than expected while the FM calling is lower than expected. This is consistent with our conclusions regarding the shift of focus for older women. In Fig. 6(B) we present the scores based on the results in Fig. 3 where the quantities of interest are, for example, $f_{FF}(g)$ instead of $f_{FF}(g, t)$. Here, the corresponding normalized score is $\{f_{FF}(g) - p_{FF}(g)\}/\sqrt{p_{FF}(g)(1 - p_{FF}(g))}$. We only show the scores corresponding to a threshold of 30 seconds on the duration of calls. The scores in Fig. 6(B) are in a way an encapsulation of the behaviour depicted in Fig. 6(A). The variation in the scores for different pairs relative to each other is mostly similar to the results in Fig. 3. However, and most notably the scores appear to provide additional clarity of the nature of variation of the MM, FF, MF and FM curves for caller ages above 50 years. Whereas, in Fig. 3, the curves overlap, the scores for the different pairs in Fig. 6(B) appear well differentiated. The communication for the FF pairs appear to take the lead over the other pairs. Also, in the case of the MM pairs, the adjustment with the null model shows that the communication is relatively much lower than what would be expected and that the original $f_{MM}(g)$ is amplified due to larger number of male subscribers.

5 Summary and conclusion

In this study, we have measured the relative interaction probabilities for the four possible caller–callee pairs of the same and opposite gender. We have observed that in general the interaction probabilities are strongly dependent on the age and gender of the caller in relation to the age and gender of the callee. Also, we observed the communication over the generation gap as depicted in Fig. 1 showing the lobed structure and in the Appendix in Fig. 9, where we depict the distribution of calls made by the callers of certain age to callees as a function of callees' age showing it to be bimodal [18, 19, 22].

Our findings from the study of the distributions of call duration for different age groups of the caller (Fig. 2) shows that the MF interactions tend to increase with call duration up to age 50 years, suggesting that men have a strong emotional connection with their opposite gender spouse of about the same age. In contrast, the FM interactions indicate that women are not as active after the age of 35 years, and have a decreasing trend for medium or long call duration with age. On the other hand, the MM interactions show initially greater probability for short call duration at younger ages, after which this becomes least probable for medium and higher call duration for any age of the caller. The FF interactions start with the lowest probability of all at younger ages, then shows a steadily increasing trend for medium and higher call duration with the age of the caller.

In the investigation of the relative probabilities for the four types of interaction as a function of the caller's age with calls above certain threshold value (30 sec, 60 sec, 120 sec
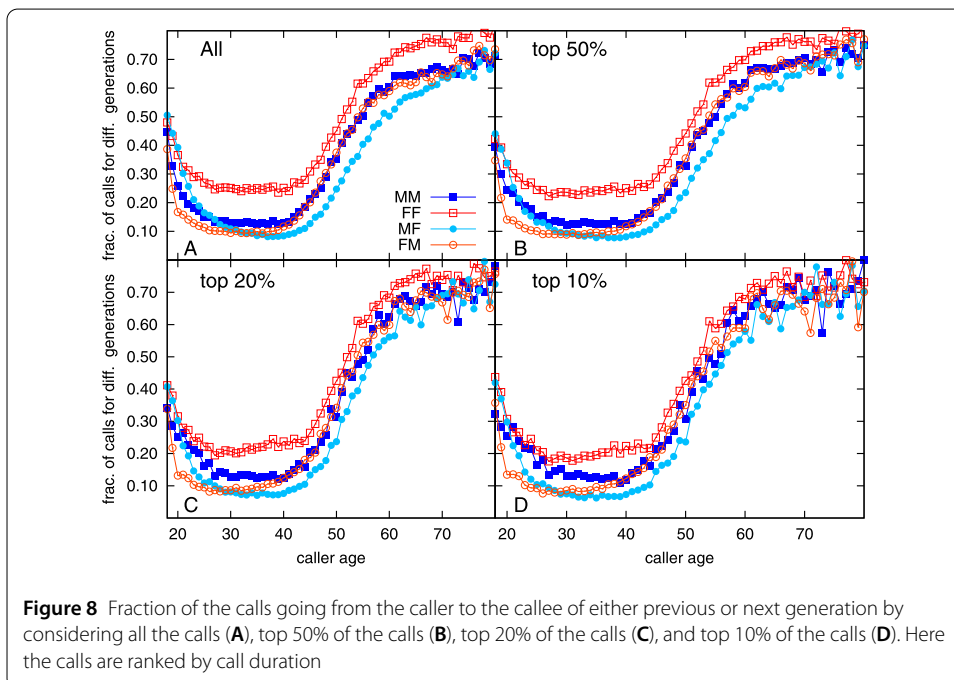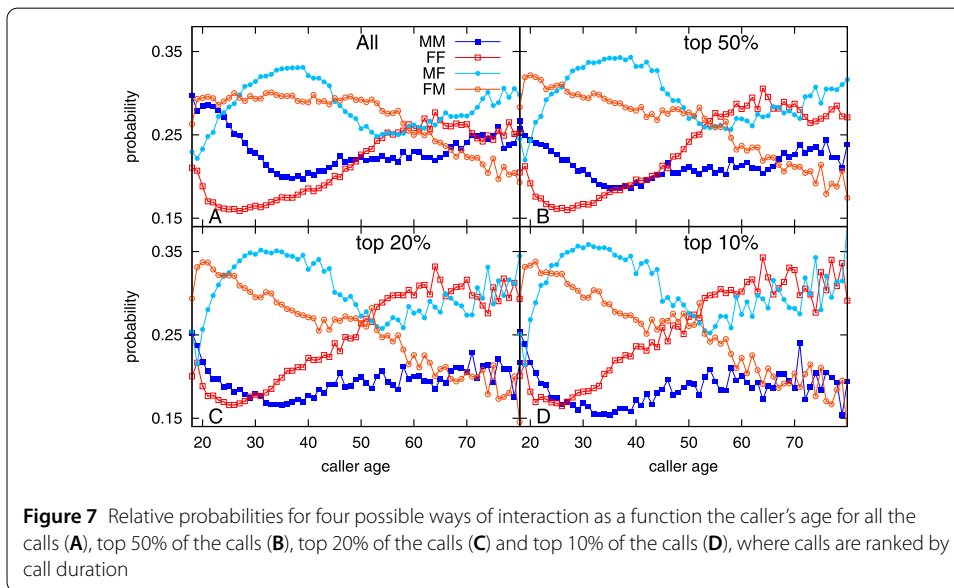
and 240 sec) (Fig. 3), we show that the FF interactions have an increasing trend with the caller's age. This is due to frequent interactions between the daughter and her mother, an indication that the grandmothering effect has set in. An opposite trend is observed for the FM interactions, i.e. the relative probability shows a decreasing trend with age. On the other hand, the MF interactions show a high probability for ages ranging from 20 to 50 years. After that, it shows a decreasing trend up to the age of 55 years, and then beyond that again shows an increasing trend. The MM interaction curve shows the weakest interaction after the age of 25 years.
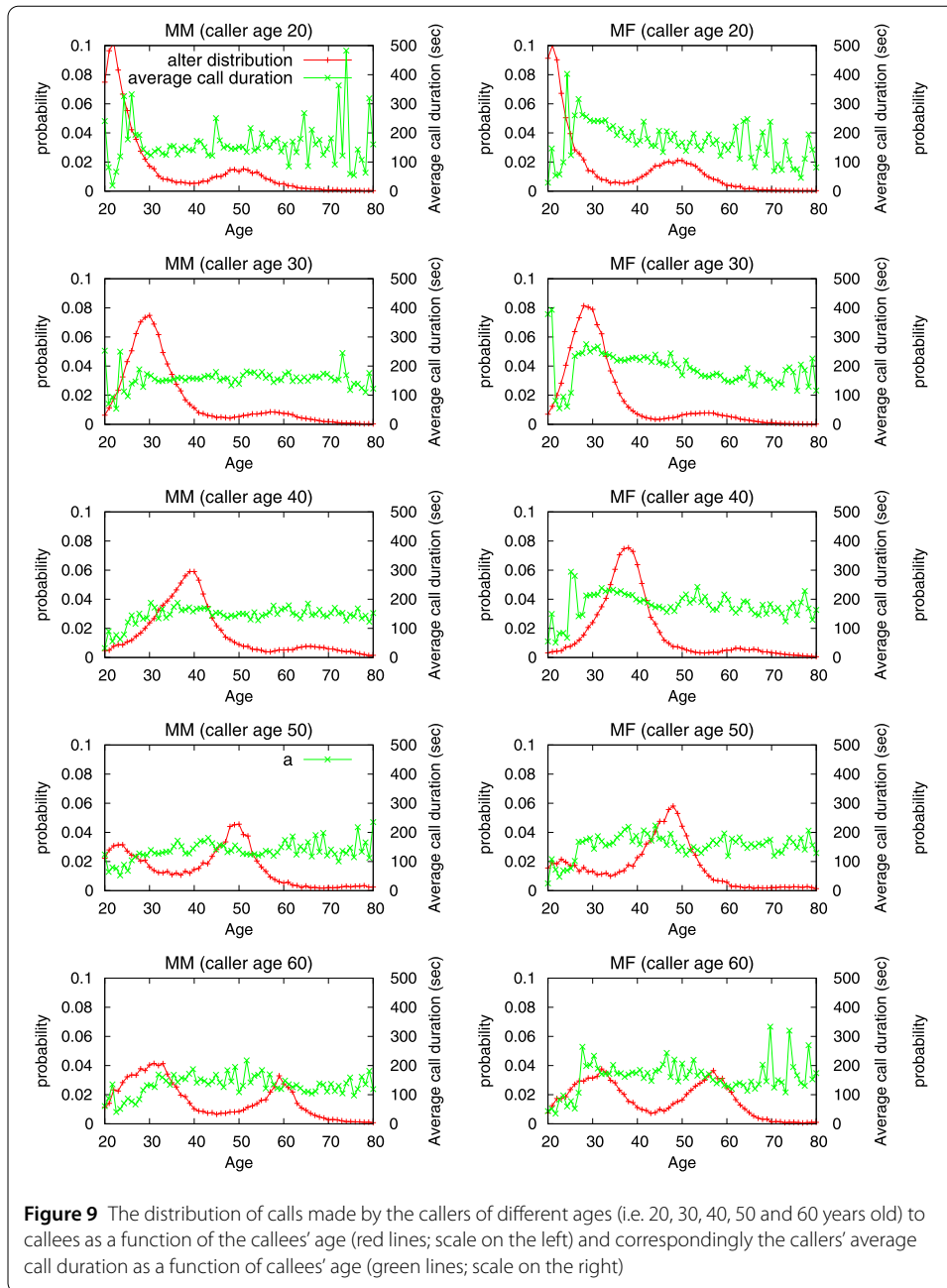
The effect of the difference in the number of male and female subscribers on the counting of pairs and the results shown in either of Fig. 2 and Fig. 3 is understood with the help of the scores calculated using the null model. These scores reveal how the populations of males and females influence the counting of pairs. Moreover, zero being a reference for the score, the latter is able to differentiate between contrasting cases, that is between negative (below the expectation with respect to the null model) and positive scores (above the expectation). For example, in Fig. 6(A) that corresponds to Fig. 2, scores for female-female calling are found to change from negative to positive with age of the caller. Overall, the scores lend support to our original results yet bringing clarity on the nature of the results.

Looking at the fractions of calls of duration more than 100 sec as a function of the caller's age (Fig. 4) revealed that for the FF interactions there is an increase from 50 to 60 years of age, which once again is taken as a clear evidence of grandmothering setting in. We have also found that on the basis of the average call duration that around 35 years of age the duration of female-to-female calls is highest among the four possible types of interaction. This can again be interpreted as a signature for the grandmothering effect. Furthermore, we showed (Fig. 5) that there is a fraction of total calls going from callers to callees who are either one generation older or one generation younger. Here we observed that for female callers the fraction of calls going to a different generation is, for all ages, always greater than for male callers. More precisely, the FF interactions show the highest probability, most likely reflecting strong ties between mothers and their daughters. On the other hand, at a younger age, a large fraction of calls go from males to females, suggesting that sons are strongly attached to their mothers before marriage. After the age of 40 years, the MF and FM interaction curves are very close to each other, suggesting that sons get the same amount of attention from both parents. More generally, we have found that younger callers, most of the calls (70–90%) to callees of the same generation. On the other hand, for older people, most of their calls go to their children (i.e. contacts who are younger by a generation), which supports the claim that affection flows downward. Broadly speaking, our conclusions regarding the preference of communicating individuals as a function of their age, and in particular, the preference of women in their post-reproductive period, is based on the observed variation in the quantities, namely the relative frequency of outgoing calls, the propensity to make calls of longer duration, and the proliferation of cross-generational calls. However, the consistency with alternative hypotheses could as well be investigated but those should be able to explain the patterns of communication taking into account the age dependent variation in kinship and different levels of friendship.

The advent of newer channels of digital communication in the last decade has rapidly supplemented the usage of mobile phones. The pattern of communication availing multiple modes, for example, voice calling in conjunction with text-messaging or through social networking services, is in turn able to characterize the nature of sociality [32–34]. How-

ever, the fact that variability in calls can also serve as an important factor, has been rather overlooked. In previous works by some of the authors we had focussed on undirected communication and tried distinguishing communication between peers, partners and kins for individuals of different gender and age [19]. Here, we investigate outgoing calls specific to gender and age, and consider duration of calls in the same spirit as multiple available channels. On one hand the current study stands consistent with the earlier works that revealed patterns in sociality like grandmothering, on the other the study might be indicative of gender and age differences in selecting different channels of communication. The understanding of the immediate social neighbourhood of individuals as well as the preferences

Figure 7 Relative probabilities for four possible ways of interaction as a function the caller's age for all the calls (**A**), top 50% of the calls (**B**), top 20% of the calls (**C**) and top 10% of the calls (**D**), where calls are ranked by call duration

Figure 8 Fraction of the calls going from the caller to the callee of either previous or next generation by considering all the calls (**A**), top 50% of the calls (**B**), top 20% of the calls (**C**), and top 10% of the calls (**D**). Here the calls are ranked by call duration

Figure 9 The distribution of calls made by the callers of different ages (i.e. 20, 30, 40, 50 and 60 years old) to callees as a function of the callees' age (red lines; scale on the left) and correspondingly the callers' average call duration as a function of callees' age (green lines; scale on the right)

in the style of communication would be valuable inputs to the developing fields of mobile health and medicine [35, 36].

Appendix

We repeat the analyses as in Fig. 3 and Fig. 5 by considering the top ranked calls where calls are ranked by call duration to show the robustness of long call duration. In Fig. 7, we show relative probabilities for the four possible types of interaction as a function of the caller's age by considering all the calls, the top 50% of calls, top 20% of the calls and top 10% of the calls (with calls ranked by duration). There are differences between all calls and top ranked calls. If we consider only potential calls, the main conclusions are (considering only top

20% of the calls and top 10% of the calls): (i) Beginning with a low value, the FF interaction curve has an increasing trend with caller age; (ii) the FM interaction curve starts with high value and then it shows a decreasing trend with age; (iii) the MF interaction curve shows a high value from age 30 to 40 years, after which it has a decreasing trend up to the age of 55 years, after which it shows an increasing trend; (iv) finally, the MM interaction curve differs from all the other curves by showing a low value for the probability for all the ages of the caller.

In Fig. 8, we depict the fraction of the calls going from the caller to the callee of either the previous or next generation. The main conclusions are as follows: (i) the FF-interaction curve shows the highest value for any age of the caller as clear indication of the mothers and daughters frequent connections; (ii) the MF-interaction curve indicates that sons are more attached to their mothers before they marry; (iii) after the age of 40 years, the MF and FM interaction curves are very close each other, indicating that sons get the same amount of attention from their parents. Also the behavior of all the four possible combinations of social interaction after the age of 40 tells us that mothers more frequently call their daughters than their sons, and fathers more frequently call their sons than their daughters. In Fig. 9, we show the distribution of calls made by the callers of certain age to callees as a function of callees age (red lines) and the corresponding average call durations (green lines). The distributions of calls turn out to be bimodal, with a maximum at around caller's own age and another maximum at an age difference of one generation [18, 19, 22].

Abbreviations
FF, female caller–female callee; FM, female caller–male callee; MF, male caller–female callee; MM, male caller–male callee.

Availability of data and materials
The datasets analysed during the current study are not publicly available due to a signed non-disclosure agreement. The dataset contains sensitive information of the subscribers, particularly age, gender, postal address and the location of the most accessed cell tower.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
Conceptualization: AG, DM, KB; methodology: All authors; formal analysis: AG; resources: KK; data curation: DM; writing (original draft preparation): AG, KB, RD; writing (review and editing): All authors; visualization: AG; supervision: RD, KK; funding acquisition: RD, KK. All authors gave final approval for publication.

Author details
[1]Department of Computer Science, Aalto University School of Science, Espoo, Finland. [2]Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Dunbar RIM (1992) Neocortex size as a constraint on group size in primates. J Hum Evol 22(6):469–493

2. De Ruiter J, Weston G, Lyon SM (2011) Dunbar's number: group size and brain physiology in humans reexamined. Am Anthropol 113(4):557–568
3. Gonçalves B, Perra N, Vespignani A (2011) Modeling users' activity on Twitter networks: validation of Dunbar's number. PLoS ONE 6(8):e22656
4. Hill RA, Dunbar RIM (2003) Social network size in humans. Hum Nat 14(1):53–72
5. Dunbar R, Arnaboldi V, Conti M, Passarella A (2015) The structure of online social networks mirrors those in the offline world. Soc Netw 43:39–47
6. MacCarron P, Kaski K, Dunbar R (2016) Calling Dunbar's numbers. Soc Netw 47:151–155
7. Blondel VD, Decuyper A, Krings G (2015) A survey of results on mobile phone datasets analysis. EPJ Data Sci 4(1):1
8. Ellison NB, Steinfield C, Lampe C (2007) The benefits of Facebook "friends:" social capital and college students' use of online social network sites. J Comput-Mediat Commun 12(4):1143–1168
9. Kwak H, Lee C, Park H, Moon S (2010) What is Twitter, a social network or a news media? In: Proceedings of the 19th international conference on World wide web. ACM, New York, pp 591–600
10. Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A-L (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci 104(18):7332–7336
11. Onnela J-P, Saramäki J, Hyvönen J, Szabó G, De Menezes MA, Kaski K, Barabási AL, Kertész J (2007) Analysis of a large-scale weighted network of one-to-one human communication. New J Phys 9(6):179
12. Kovanen L, Karsai M, Kaski K, Kertész J, Saramäki J (2011) Temporal motifs in time-dependent networks. J Stat Mech Theory Exp 2011(11):P11005
13. Aledavood T, López E, Roberts SGB, Reed-Tsochas F, Moro E, Dunbar RIM, Saramäki J (2015) Daily rhythms in mobile telephone communication. PLoS ONE 10(9):e0138098
14. Kujala R, Aledavood T, Saramäki J (2016) Estimation and monitoring of city-to-city travel times using call detail records. EPJ Data Sci 5(1):6
15. Mones E, Stopczynski A, Lehmann S (2017) Contact activity and dynamics of the social core. EPJ Data Sci 6(1):6
16. Monsivais D, Bhattacharya K, Ghosh A, Dunbar RIM, Kaski K (2017) Seasonal and geographical impact on human resting periods. Sci Rep 7(1):10717
17. Monsivais D, Ghosh A, Bhattacharya K, Dunbar RIM, Kaski K (2017) Tracking urban human activity from mobile phone calling patterns. PLoS Comput Biol 13(11):e1005824
18. Palchykov V, Kaski K, Kertész J, Barabási A-L, Dunbar RIM (2012) Sex differences in intimate relationships. Sci Rep 2
19. David-Barrett T, Kertesz J, Rotkirch A, Ghosh A, Bhattacharya K, Monsivais D, Kaski K (2016) Communication with family and friends across the life course. PLoS ONE 11(11):e0165687
20. Onnela J-P, Arbesman S, González MC, Barabási A-L, Christakis NA (2011) Geographic constraints on social network groups. PLoS ONE 6(4):e16939
21. Saramäki J, Leicht EA, López E, Roberts SGB, Reed-Tsochas F, Dunbar RIM (2014) Persistence of social signatures in human communication. Proc Natl Acad Sci 111(3):942–947
22. Bhattacharya K, Ghosh A, Monsivais D, Dunbar RIM, Kaski K (2016) Sex differences in social focus across the lifecycle in humans. R Soc Open Sci 3(4):160097
23. Dong Y, Yang Y, Tang J, Yang Y, Chawla NV (2014) Inferring user demographics and social strategies in mobile social networks. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 15–24
24. Hawkes K (2004) Human longevity: the grandmother effect. Nature 428(6979):128–129
25. Coall DA, Hertwig R (2010) Grandparental investment: past, present, and future. Behav Brain Sci 33(1):1–19
26. Karikoski J, Nelimarkka M (2011) Measuring social relations with multiple datasets. Int J Soc Comput Cyber-Phys Syst 1(1):98–113
27. Wagner C, Graells-Garrido E, Garcia D, Menczer F (2016) Women through the glass ceiling: gender asymmetries in Wikipedia. EPJ Data Sci 5(1):5
28. Mueller W, Silva TH, Almeida JM, Loureiro AAF (2017) Gender matters! Analyzing global cultural gender preferences for venues using social sensing. EPJ Data Sci 6(1):5
29. Newman M (2010) Networks: an introduction. Oxford University Press, New York
30. Zar JH (2013) Biostatistical analysis: Pearson new international edition. Pearson Education, Upper Saddle River
31. Rosenthal R, DiMatteo MR (2001) Meta-analysis: recent developments in quantitative methods for literature reviews. Annu Rev Psychol 52(1):59–82
32. Wang Y, Faloutsos M, Zang H (2013) On the usage patterns of multi-modal communication: countries and evolution. In: Computer communications workshops (INFOCOM WKSHPS), 2013 IEEE conference on. IEEE, pp 97–102
33. Zignani M, Quadri C, Bernardinello S, Gaito S, Rossi GP (2014) Calling and texting: social interactions in a multidimensional telecom graph. In: Signal-image technology and Internet-based systems (SITIS), 2014 tenth international conference on. IEEE, pp 408–415
34. Quadri C, Zignani M, Capra L, Gaito S, Rossi GP (2014) Multidimensional human dynamics in mobile phone communications. PLoS ONE 9(7):e103183
35. Onnela J-P, Rauch SL (2016) Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 41(7):1691
36. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, Rozario S (2016) A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. PLoS ONE 11(5):e0154248