Uncovering intimate and casual relationships from mobile phone communication

Mikaela Irene D. Fudolig[1], Daniel Monsivais[2], Kunal Bhattacharya[2], Hang-Hyun Jo[1,2,3], and Kimmo Kaski[2,4]

[1] Asia Pacific Center for Theoretical Physics, Pohang, Republic of Korea
mikaela.fudolig@apctp.org,
[2] Department of Computer Science, Aalto University, Espoo, Finland
[3] Department of Physics, Pohang University of Science and Technology, Pohang, Republic of Korea
[4] The Alan Turing Institute, London, United Kingdom

Abstract. We analyze a large-scale mobile phone call dataset with the metadata of the mobile phone users, including age, gender, and billing locality, to uncover the nature of relationships between peers or individuals of similar ages. We show that in addition to the age and gender of users, the information about the ranks of users to each other in their egocentric networks is crucial in characterizing intimate and casual relationships of peers. The opposite-gender pairs in intimate relationships are found to show the highest levels of call frequency and daily regularity, consistent with small-scale studies on romantic partners. This is followed by the same-gender pairs in intimate relationships, while the lowest call frequency and daily regularity are observed for the pairs in casual relationships. We also find that older pairs tend to call less frequently and less regularly than younger pairs, while the average call durations exhibit a more complex dependence on age. We expect that a more detailed analysis can help us better characterize the nature of peer relationships and distinguish various types of relations, such as siblings, friends, and romantic partners, more clearly.

Keywords: mobile phone call dataset · egocentric network · intimate relationship · casual relationship · romantic partner

1 Introduction

Traditionally, the studies of human relationships and human social networks have been conducted using questionnaire-based surveys [33,31]. As these surveys focus on detailed information about social ties between human individuals, they tend to be limited by the number of subjects, by the relative uniformity of subjects often recruited from the same social surrounding, and by the memory of the subjects filling the questionnaires. Recent digital technologies like mobile phones and Radio Frequency Identification (RFID) have enabled researchers to supplement the survey data with much more detailed relational data between subjects [12,18,7]. Although the data in these studies are still limited in size and

in the diversity of the subjects, they have paved the way for more accurate and quantitative description of social behavior of human individuals embedded in a social environment or network. An additional benefit of this type of studies is that they allow the cross-validation of the data gathered from different sources [29,30].

Large-scale mobile phone datasets have also become available due to rapid advance of digital mobile phone technology in the hands of people generating vast amount of data as traces of their behavior, which has facilitated a complementary approach to investigate human sociality even at the population level. The *call detail records* (CDRs) enable us to map the patterns of sociality at diverse scales, from studying the structure and dynamics of large-scale social networks [24,21], to the level of communities and groups [5], to immediate social neighborhood of individuals in terms of egocentric networks [25,19]. In these works, the strength of ties between individuals has often been quantified in terms of the frequency of contact between them. More recently, such information on the tie strength has been combined with the metadata, such as the age, gender, and billing post code of users, which in turn has enabled us to gain deeper insight into the nature of human sociality [25,19,4,11].

However, as the CDRs are anonymized, they do not carry the true nature of relationships between individuals. An approach to circumvent this issue is to utilize the demographic and/or locational information of the users and to make plausible assumptions about the nature of relationships between the users [9,10]. For example, for a given user (an ego), the contacts of the ego (alters) are ranked by the call frequency between the ego and each alter; a few of the top-ranked alters are selected for the study. Then the tie strengths of close relationships are correlated with the age, gender, and location information of the users. The findings in these studies turn out to be indicative and consistent with the well-understood life-course patterns of human sociality [16,20]. However, this approach can be refined to distinguish between pairs with the same demographic information but with different relationships, e.g., between opposite-gender friends and opposite-gender romantic partners.

In the present study we extend the above described approach to analyze the large-scale mobile phone dataset derived from CDRs focusing on getting insight into the nature of relationships between peers or individuals of similar ages. We combine the metadata of users, including age, gender, and billing post code, with information about the ranks of users to each other in their egocentric networks to characterize the nature of peer relationships as being either intimate or casual. We show that the rank information is crucial to distinguish intimate and casual relationships of peers. We find that such relationships are successfully distinguished by the calling patterns in terms of the average daily call frequency, daily regularity, and average call duration.

Our paper is organized as follows: In Section 2 we present the description of the data used in this study, followed by the methods of data preprocessing and statistical tests. Then in Section 3 we present the results focusing on average daily call frequency, daily regularity, and average call duration. This is followed by the discussion in Section 4, in which we focus our attention to intimate rela-

tionships of opposite-gender pairs, intimate relationships of same-gender pairs, and casual relationships. Finally we draw conclusions in Section 5.

2   Methods

2.1   Data description

We analyze the mobile phone call dataset of a European service provider for the first seven months of year 2007 (212 days). During this period, which is before the rise of smartphones and social network services, a significant part of the mobile communication was done through voice calls and Short Message Services (SMSs). The service provider had subscribers numbering around 20% of the population of the country [24].

The dataset contains the date and time for all the outgoing and incoming calls between subscribers or users. The duration is included for the calls between the users and for the outgoing calls from the users to those who subscribed to other providers, which we call non-company users. The duration is zero for incoming calls from non-company users to users, but the date and time of such calls are included. We discard the users whose contracts are known to begin or end within the period of interest, i.e., the first seven months in 2007.

For each contract, some metadata of the users, such as age and gender, are included in the dataset; for most users, the billing post code is also included. We only consider users with known age and gender. Thus, we exclude non-company users whose age and gender are unknown. In addition, there are users with different identifiers associated with a single contract, which we also filter out to avoid inconsistencies. This is because in many cases, users of a single contract have exactly the same demographic information, and it is not possible to determine whether there is only one or more persons using many subscriptions.

2.2   Data preprocessing

For each user with known metadata, which we call an "ego", we enlist all the other users the ego communicated with, which we call "alters". The alters are ranked in descending order according to the total number of incoming and out-going calls made between the ego and each alter. By keeping the top five alters for each ego, we make the list of ego-alter pairs.

We are interested only in ego-alter pairs who have significant relationships, such as family, friends, and romantic partners. To filter out pairs which do not meet this criterion, we impose regularity by excluding purely transactional calls, which are characterized by lower call frequency and less regularity [15]. Specifically, we exclude pairs who have had calls in less than five out of the seven months. For example, if a pair has one thousand calls but only for the first month, we exclude that pair from the analysis. Further, we also exclude the ego-alter pairs in which the metadata of the alter does not include the age and gender. After these filtering steps, we are left with the users with known meta-data who make calls regularly to each other. Note that, since the filtering is done

Table 1: Numbers of pairs in nine demographic groups according to the genders and the younger user's age in each pair, with percentages in each group when decomposed by the ranks of users to each other, i.e., mutual top-rank (1-1), mutual non-top-rank (n-n), and non-mutual top-rank (1-n). Due to rounding errors, the percentages may not sum to exactly 100%.

		Opposite-gender (−)		Same-gender female (+f)		Same-gender male (+m)
Young adulthood		95,128		47,930		70,945
(Y)	1-1	38.0%	1-1	6.9%	1-1	11.0%
	n-n	45.6%	n-n	75.5%	n-n	69.9%
	1-n	16.5%	1-n	17.7%	1-n	19.2%
Middle adulthood		232,133		98,225		120,256
(M)	1-1	40.3%	1-1	9.9%	1-1	14.0%
	n-n	38.5%	n-n	68.8%	n-n	65.0%
	1-n	21.2%	1-n	21.3%	1-n	21.0%
Late adulthood		57,085		20,165		26,544
(L)	1-1	30.1%	1-1	14.7%	1-1	19.9%
	n-n	40.0%	n-n	59.3%	n-n	50.6%
	1-n	29.9%	1-n	26.0%	1-n	29.5%

after ranking, the ranking preserves the true importance of the alter to the ego, as far as call frequency is concerned.

It is possible that two users appear in each other's list of the top five alters. In such a case, the total number of incoming and outgoing calls and the total call duration are the same for both users, but this pair appears twice in the list of ego-alter pairs; we keep only one of these two.

After the above described ranking, filtering, and removing of duplicates, we are left with 322,823 users in 1,236,364 pairs. Of these, we consider the pairs more likely to be in a peer relationship rather than in a parent-child relationship. These two relationships can be distinguished using the age difference of the users; we set the cutoff to be 20 years, based on European census data [13]. Then we study pairs whose age difference is less than 20 years, which are then categorized into nine demographic groups. We first consider three combinations of genders of each pair: (i) opposite-gender, denoted by "−", (ii) same-gender female or "+f", and (iii) same-gender male or "+m". For each gender combination group, we consider three age groups according to the age of the younger user in each pair, being either 18–28 years old (young adulthood or "Y"), 29–45 years old (middle adulthood or "M"), or 46–55 years old (late adulthood or "L"), following the scheme of life stages used in Ref. [9]. Here, the pairs whose younger user is younger than 18 years old or older than 55 years old are not considered since the sample sizes for these demographic groups are not large enough. Consequently, we focus on 768,411 pairs in nine demographically separable groups, denoted by −Y, −M, −L, +fY, +fM, +fL, +mY, +mM, and +mL, respectively. Each of these groups has at least 20,000 ego-alter pairs, as summarized in Table 1. Also, although we consider the maximum age difference to be 20 years, we find that most of these pairs show an age difference of only 0–5 years, as shown in Table 2.

Table 2: Distributions of the age differences of the pairs. For all demographic groups, pairs with an age difference of 0–5 years comprise more than half of each group, followed by those with age differences in the range of 6–10 years. Although the peers are defined to have an age difference less than 20 years, most of the pairs in each group show age differences less than 10 years. Due to rounding errors, the percentages may not sum to exactly 100%.

	Age difference	**Opposite-gender ($-$)**			**Same-gender female ($+$f)**			**Same-gender male ($+$m)**		
	(years)	1-1	n-n	1-n	1-1	n-n	1-n	1-1	n-n	1-n
Y	0–5	73%	56%	62%	68%	64%	65%	71%	68%	71%
	6–10	20%	26%	23%	17%	23%	20%	20%	21%	19%
	11–19	7%	18%	15%	15%	13%	15%	9%	11%	10%
M	0–5	78%	58%	68%	69%	64%	65%	75%	67%	69%
	6–10	16%	24%	20%	16%	21%	17%	17%	21%	19%
	11–19	6%	18%	13%	15%	15%	17%	8%	13%	12%
L	0–5	76%	62%	72%	75%	61%	71%	83%	65%	76%
	6–10	17%	23%	19%	15%	23%	17%	12%	21%	15%
	11–19	6%	15%	9%	10%	16%	12%	6%	14%	8%

Each demographic group can be further divided into three subgroups according to the ranks of users in a pair to each other: (i) Both users in a pair are the top-rank alters of each other, which can be called *mutual top-rank* and denoted by "1-1", (ii) both users are not the top-rank alters of each other, i.e., *mutual non-top-rank* or "n-n", and (iii) one of the users is the top-rank alter of the other, but it is not mutual, i.e., *non-mutual top-rank* or "1-n". Table 1 shows that for all age groups, mutual top-rank pairs comprise a large portion in the opposite-gender groups, while they are a small minority in the same-gender groups.

In addition to all of the above, we can extract the locational information of users with the help of the billing post code, which we assume to correspond to the user's home address. We will focus on whether the users of each pair have the same post code or different ones.

2.3   Statistical test

All the statistical tests are done on the log-transformed variable whenever necessary. To test for statistical significance, we use one-way ANOVA and Tukey's HSD post-hoc test when the variances are found equal by Levene's test [1]. If heteroscedasticity is obtained, Welch's ANOVA [32] and the Games-Howell post-hoc test [14] are used instead. The tests are implemented using Python's `scipy` and `statsmodels` as well as using R's `userfriendlyscience` packages.

Due to the large sample sizes in this study, the power of statistical tests is high [8], and true differences, no matter how small they are, are more likely to be found as significant. For brevity, we only mention the relevant results of the statistical tests where the null hypothesis cannot be rejected. Otherwise, the
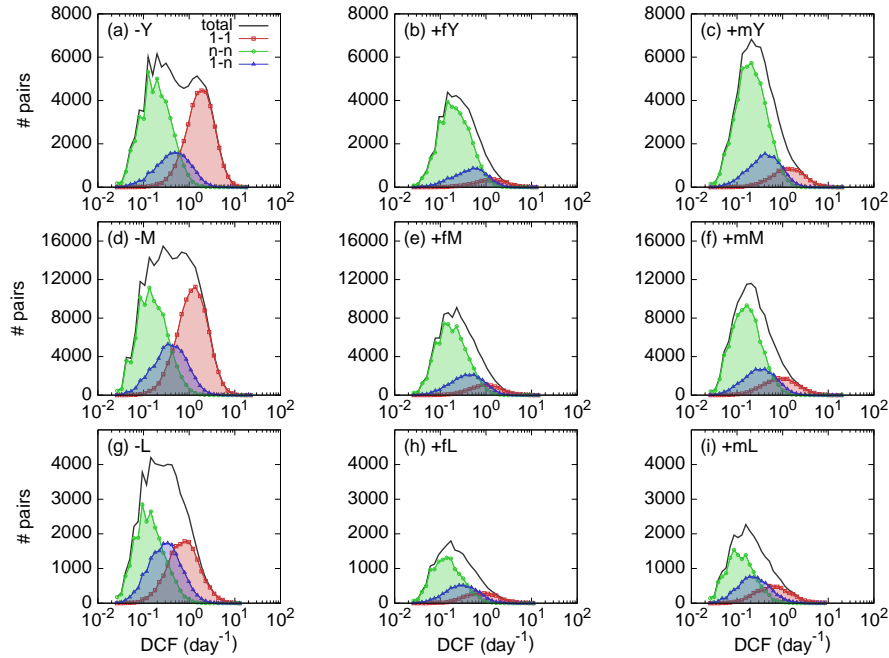
Fig. 1: **Distributions of the average daily call frequency (DCF).** For each demographic group, the total distribution (black solid line) is decomposed into three subgroups depending on the ranks of users to each other: mutual top-rank (1-1; red line with squares), mutual non-top-rank (n-n; green line with circles), and non-mutual top-rank (1-n; blue line with triangles). The unit of DCF is day^{-1}.

statistical tests either show a significant difference or are overruled by practical significance.

3 Results

In order to quantify the calling patterns of ego-alter pairs, we introduce three quantities, i.e., the average daily call frequency, fraction of days active, and average call duration. The distributions of these quantities are then systematically compared across different demographic groups.

3.1 Average daily call frequency

We first obtain the number of calls made by each pair, i.e., the call frequency. Dividing this call frequency by the number of days in the observation period, i.e., 212 days, we get the *average daily call frequency* (DCF) per pair to obtain its distributions. In Fig. 1 we find that the distribution of DCFs for each of
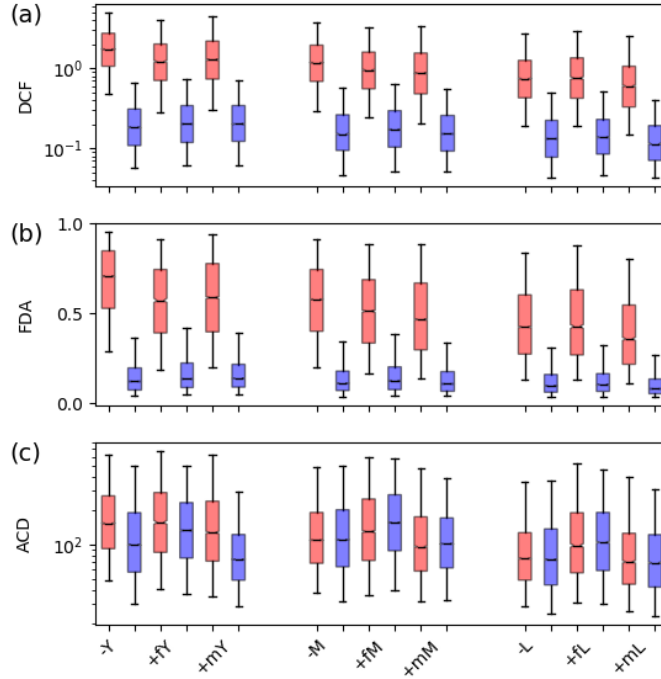
Fig. 2: **Summary of distributions of the average daily call frequency (DCF), the fraction of days active (FDA), and the average call duration (ACD).** The pink bars correspond to the mutual top-rank (1-1) pairs, while the blue bars are for the mutual non-top-rank (n-n) pairs. The black lines in the middle indicate the medians, the bars include data points from the 25th to the 75th percentile, and the whiskers show the 5th and 95th percentiles. The notches show confidence intervals generated by bootstrapping with 10000 resamples. The units of DCF and ACD are day^{-1} and seconds, respectively.

nine demographic groups can be overall described by unimodal distributions on a log-scale, except for the opposite-gender young adulthood $(-Y)$ case in Fig. 1(a), showing a clear bimodality. This bimodality is resolved by separating the pairs in the $-Y$ group according to the ranks of the users to each other. We observe that the mutual top-rank (1-1) pairs and mutual non-top-rank (n-n) pairs successfully account for the right and left peaks of the bimodal distribution, such that the median values for 1-1 and n-n pairs are around 1.75 and 0.18 calls per day, respectively. There is also a non-mutual top-rank (1-n) minority whose distribution shows a peak between those of mutual top-rank and mutual non-top-rank pairs, which will be discussed in Subsection 4.4. Although the bimodality found in the $-Y$ case is not evident in the rest of the groups, we find that the mutual top-rank pairs show, in general, largely different calling patterns from
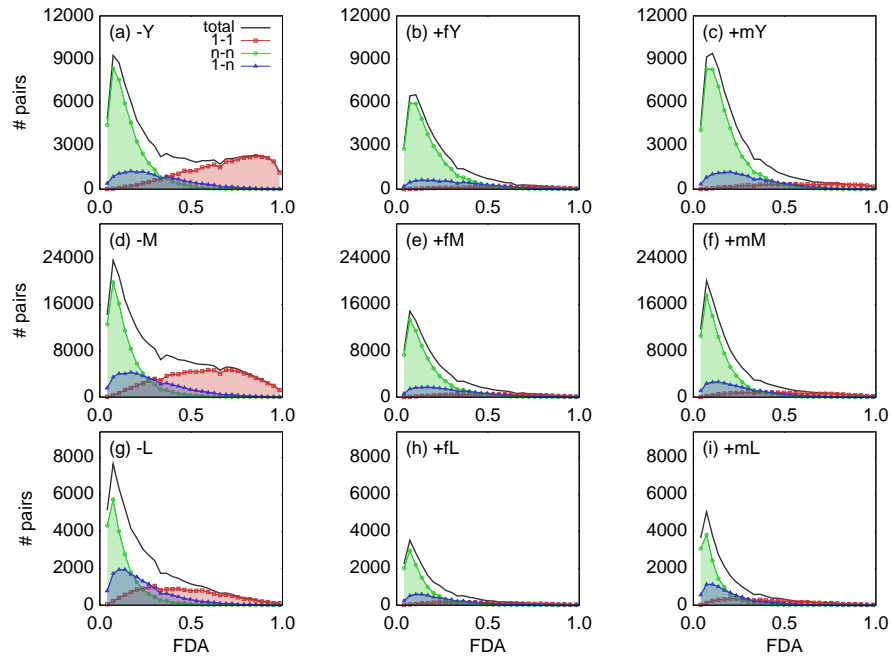
Fig. 3: **Distributions of the fraction of days active (FDA).** All notations are the same as those in Fig. 1.

the mutual non-top-rank pairs in all the other demographic groups, as depicted in Fig. 2(a). Note that although mutual top-rank pairs are expected to have more calls than mutual non-top-rank pairs, the successful decomposition of the bimodality by using the rank information is not straightforward.

We summarize other relevant findings from the results in Figs. 1 and 2(a). For all the gender combinations of pairs, younger pairs tend to call considerably (slightly) more often than older pairs in the mutual top-rank (mutual non-top-rank) case. For the mutual top-rank case, opposite-gender pairs call more frequently than their same-gender counterparts for both Y and M groups, while for the oldest (L) groups, there is no significant difference at $\alpha = 0.05$ between opposite-gender and same-gender female pairs ($p = 0.18$), but both opposite-gender and same-gender female pairs call more often than same-gender male pairs. On the other hand, for the mutual non-top-rank case, we find no clear gender dependence of the DCF for each age group.

3.2 Daily regularity

In order to quantify the temporal regularity of the calling patterns on a daily basis, we define the *fraction of days active* (FDA) as the fraction of days in the observation period in which at least one call was made between the users of
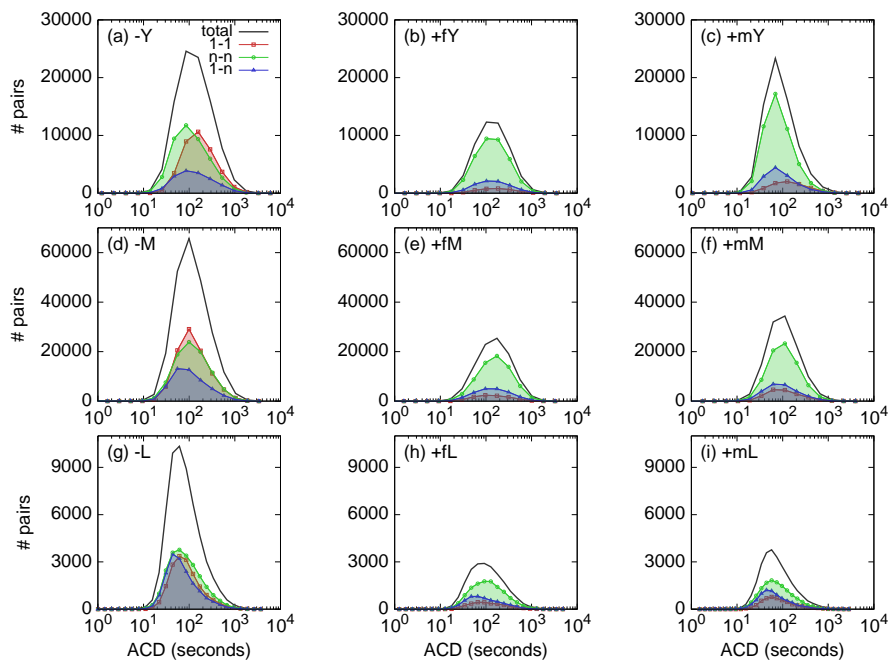
Fig. 4: **Distributions of the average call duration (ACD).** All notations are the same as those in Fig. 1, except that the unit of ACD is seconds.

each pair. By the FDA, one can distinguish, e.g., the case of 10 days with 10 calls per day from the case of 100 days with one call per day, which cannot be distinguished by the average daily call frequency (DCF).

We find that the FDA is highly correlated with the DCF ($r = 0.681$). It should be noted that the number of days active cannot be greater than the call frequency for each pair, which possibly enables the strongly positive correlation between FDA and DCF. However, how calls are distributed over the observation period is yet an interesting question, in particular, for pairs with high DCF: The pairs with a high DCF tend to have a high FDA, implying that the calls are made rather regularly instead of being lumped into a few days. The distributions of FDAs are presented in Fig. 3 and summarized in Fig. 2(b). Overall, we find similar behavior to that observed in the case of DCF, except that the shapes of the distributions are highly skewed either to the left or to the right, probably due to the intrinsic range of the quantity, i.e., FDA $\in [0, 1]$. The most pronounced difference between the mutual top-rank and the mutual non-top-rank pairs is observed again in the $-Y$ group as their median values are 0.71 and 0.13, respectively.

3.3   Average call duration

Finally, for studying the calling patterns in more detail, we calculate the average duration per call or the *average call duration* (ACD) for each pair by dividing the total call duration (in seconds) by the number of calls. The ACD turns out to be positively but only barely correlated with the DCF ($r = 0.087$) as well as with the FDA ($r = 0.075$). As shown in Figs. 4 and 2(c), unlike the DCF and FDA, there seems to be no clear demographic dependence of the ACD across the different age and gender groups. However, in the medians of the distributions we observe that for all the gender groups, younger pairs tend to have longer calls than older pairs only in the mutual top-rank case. Interestingly, the same-gender female pairs make longer calls than their opposite-gender and same-gender male counterparts for all the age groups, regardless of the ranks, except for one case; there is no significant difference between the mutual top-rank pairs in $-$Y and $+$fY groups ($p = 0.79$). We also find that in the young adulthood (Y) case, the mutual top-rank pairs have longer calls than mutual non-top-rank pairs for all gender combinations, while the opposite tendency is significantly observed for the $+$fM group.

4   Discussion

Based on the above empirical observations, we hypothesize that across all demographic groups, mutual top-rank pairs and mutual non-top-rank pairs have essentially different calling patterns, thus implying different types of relationships. In case of the mutual top-rank (1-1) pairs, the high number of calls and high daily regularity imply *intimate relationships*. On the other hand, mutual non-top-rank (n-n) pairs have fewer calls and very low regularity, enabling us to characterize them as *casual relationships*. As the calling patterns of mutual top-rank pairs are also differentiated by their genders, in the following we discuss three types of relationships: intimate relationships of opposite-gender pairs, intimate relationships of same-gender pairs, and casual relationships.

4.1   Intimate relationships of opposite-gender pairs

The opposite-gender, mutual top-rank pairs can be considered as being intimate or even romantic across all the age groups as they show the highest level of call frequency and regularity compared to all other gender and rank cases. This is consistent with the small-scale studies involving college students (corresponding to the age group of Y) in romantic relationships, where pairs with greater frequency or duration of phone calls have less relational uncertainty and higher intimacy [17,23]. Moreover, those in romantic relationships are, on average, found to call each other more regularly [23]. In addition, as the mutual top-rank pairs form a significant chunk in the opposite-gender groups, but only a small minority in the same-gender groups, romantic partnerships seem to be the most feasible characterization of these pairs.
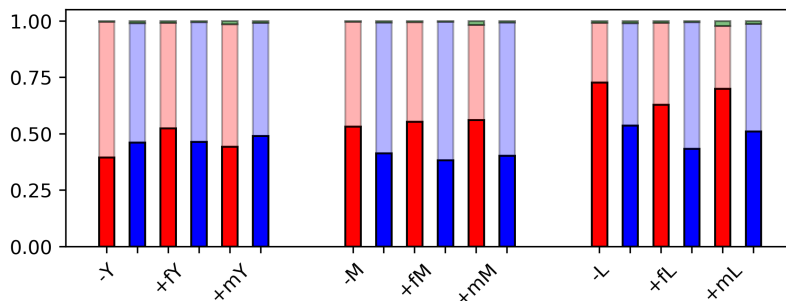
Fig. 5: **Decomposition of groups by the billing post code.** The red bars correspond to mutual top-rank (1-1) pairs, while the blue bars correspond to mutual non-top-rank (n-n) pairs for each group. The darker bars denote the fraction of pairs with the same post codes, while the lighter bars denote the fraction of those with different post codes. The fractions of pairs in which at least one post code is not available are denoted by the green bars.

We also observe that for this kind of relationship, younger pairs tend to have more frequent, more regular, and longer calls than older pairs, as depicted in Fig. 2. To study whether this tendency is due to the lower usage of mobile phones among the older generation [22] or due to the actual communication patterns of older users, more work is called for.

Next we analyze the location information of the pairs in intimate relationships by assuming that the billing post codes correspond to the home address of the users. It is known that the frequency of the face-to-face interaction, constrained by the location, is positively correlated with the frequency of contact by telephone and other media [27,26,28], enabling us to study how the locations of users in intimate relationships are related to each other. In Fig. 5 we find that the majority (60.1%) of mutual top-rank pairs in the −Y group have different post codes, possibly because they are not yet cohabiting. This trend is reversed for older age groups. The majority of mutual top-rank pairs in −M and −L groups have the same post codes because romantic pairs in these age groups are likely to be married and/or cohabiting. This tendency is consistent with the previous empirical findings using the same dataset [19]. There are, however, a significant chunk of pairs with different post codes (46.4% in −M and 26.5% in −L), which may correspond to dating pairs or possibly married pairs living in different locations.

4.2   Intimate relationships of same-gender pairs

The same-gender, mutual top-rank pairs can also be considered as being intimate across all age groups as their calling patterns are clearly more active and regular than their mutual non-top-rank counterparts. Yet, they are less active

and less regular than their opposite-gender counterparts, which implies that the same-gender, mutual top-rank pairs have a different type of relationships than the opposite-gender intimate relationships. However, such differences turn out to get smaller for older age groups. In addition, in terms of the median of the distribution of average call duration, the same-gender female pairs tend to have longer calls than the opposite-gender and same-gender male pairs, which is consistent with those in Refs. [4,9].

In order to characterize some of the same-gender intimate relationships as romantic, we need more supporting evidence for the communication patterns of homosexual romantic relationships. One can rather say that these pairs may be a mixture of romantic, familial, and other relationships.

4.3   Casual relationships

The mutual non-top-rank pairs in all demographic groups are here considered as casual relationships, as they are characterized by the lowest level of call frequency and daily regularity compared to their mutual top-rank counterparts. As for the average call duration, in terms of the median of its distribution, the mutual non-top-rank (casual) pairs tend to have shorter or similar call durations than the mutual top-rank (intimate) pairs in most cases, except for the +fM case, where the average call duration of casual pairs (around 157 seconds) is significantly larger than that of the intimate pairs (around 134 seconds). Moreover, among the mutual non-top-rank pairs, the +fM group shows the longest average call duration. This could be due to requirements of child rearing, job demands in the mid- to high-level careers, or other life events.

Similarly to the opposite-gender and same-gender intimate relationships, the average daily call frequency and the fraction of days active in casual relationships are decreasing with their age. However, the average call duration is the highest for the middle adulthood (M) group irrespective of gender.

4.4   Other relevant issues

So far we have focused on the ego-alter pairs as if they are separated from the rest of the social network. By incorporating the network structure surrounding those pairs, one can tackle some unresolved issues. For example, friends, family, and romantic relationships may be differentiated using the information about their common contacts, while the non-mutual top-rank (1-n) pairs may be studied in the context of directed relationships [3,2]. Since the peak of the DCF distribution in the 1-n case lies between the mutual top-rank (1-1) and mutual non-top-rank (n-n) peaks, we can hypothesize that they may exhibit different behaviors from both. They may also be composed of two subgroups, one resembling mutual top-rank pairs, the other resembling mutual non-top-rank pairs.

5   Conclusion

We have analyzed the large-scale call detail records (CDRs) with the metadata, such as the age, gender, and billing post code of mobile phone users, by focusing on around 770,000 peer relationships with an age difference of less than 20 years. We show that in addition to the metadata, the ranks of users to each other, determined by the call frequency between them, can be successfully used to uncover the nature of their relationships. In particular, mutual top-rank pairs have markedly different calling patterns from mutual non-top-rank pairs, not only in terms of call frequency but also in terms of daily regularity. These differences could enable us to characterize mutual top-rank pairs as *intimate relationships* and mutual non-top-rank pairs as *casual relationships*, respectively. By doing so, we could differentiate relationships of users with the same demographic information, such as friends and romantic partners.

We have found that mutual top-rank pairs are much more common among opposite-gender pairs. This, as well as the consistency of their calling patterns with those observed in romantic couples, makes it feasible that opposite-gender, mutual top-rank pairs reflect romantic relationships. On the other hand, although same-gender, mutual top-rank pairs also have relatively high call frequency and regularity, they have different calling patterns compared to their opposite-gender, mutual top-rank counterparts. This may be because this same-gender group is not solely composed of romantic partnerships, but may also include platonic or familial ties as well. In contrast to the mutual top-rank pairs, the mutual non-top-rank pairs exhibit the lowest levels of daily regularity and call frequency. We suppose that these pairs are very unlikely to be romantic partners, but instead they are more likely to be platonic or familial pairs.

The calling patterns between peers have also been found to vary with the age of the users. We find that older pairs tend to call less frequently and less regularly than younger pairs. The mutual top-rank pairs tend to also have longer average call durations than the mutual non-top-rank pairs for younger pairs of 18–28 years old. For the older pairs, the difference between the mutual top-rank and the mutual non-top-rank pairs is smaller. Interestingly, we find that in the case of the female peers in the age range of 29–45, the mutual non-top-rank pairs make significantly longer calls than the mutual top-rank pairs. This age range corresponds to the period when most couples begin families; hence such calling patterns may be due to the demands of family and work on women of that age range. Our findings can be related to the shift in social focus: While both men and women in their young adulthood are likely to maintain stronger social focus on their partners [6], the attention of individuals in middle adulthood gets distributed to alters other than their partners due to time constraints and the increase in the number of familial ties [4].

Finally, we discuss possible future studies. While we have focused exclusively on peers, we can also investigate the child-parent relationships. In addition, as our analysis has focused on the ego-alter pairs, network analysis may help us to uncover more about the users' relationships, and even to distinguish between

other types of relationships, such as platonic and familial relationships. These are all interesting issues for future work.

References

1. NIST/SEMATECH e-Handbook of Statistical Methods (2013), `https://www.itl.nist.gov/div898/handbook/index.htm`
2. Almaatouq, A., Radaelli, L., Pentland, A., Shmueli, E.: Are You Your Friends Friend? Poor Perception of Friendship Ties Limits the Ability to Promote Behavioral Change. PLOS ONE **11**(3), e0151588 (3 2016). https://doi.org/10.1371/journal.pone.0151588, `http://dx.plos.org/10.1371/journal.pone.0151588`
3. Ball, B., Newman, M.: Friendship networks and social status. Network Science **1**(01), 16–30 (4 2013). https://doi.org/10.1017/nws.2012.4, `http://www.journals.cambridge.org/abstract_S2050124212000045`
4. Bhattacharya, K., Ghosh, A., Monsivais, D., Dunbar, R.I.M., Kaski, K.: Sex differences in social focus across the life cycle in humans. Royal Society Open Science **3**(4), 160097 (4 2016). https://doi.org/10.1098/rsos.160097, `http://rsos.royalsocietypublishing.org/lookup/doi/10.1098/rsos.160097`
5. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment **2008**(10), P10008 (10 2008). https://doi.org/10.1088/1742-5468/2008/10/P10008, `http://stacks.iop.org/1742-5468/2008/i=10/a=P10008?key=crossref.46968f6ec61eb8f907a760be1c5ace52`
6. Burton-Chellew, M.N., Dunbar, R.I.M.: Romance and reproduction are socially costly. Evolutionary Behavioral Sciences **9**(4), 229–241 (2015). https://doi.org/10.1037/ebs0000046, `http://doi.apa.org/getdoi.cfm?doi=10.1037/ebs0000046`
7. Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J.F., Vespignani, A.: Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. PLoS ONE **5**(7), e11596 (7 2010). https://doi.org/10.1371/journal.pone.0011596, `http://dx.plos.org/10.1371/journal.pone.0011596`
8. Columb, M., Atkinson, M.: Statistical analysis: sample size and power estimations. BJA Education **16**(5), 159–161 (5 2016). https://doi.org/10.1093/bjaed/mkv034, `http://linkinghub.elsevier.com/retrieve/pii/S2058534917300732`
9. David-Barrett, T., Kertesz, J., Rotkirch, A., Ghosh, A., Bhattacharya, K., Monsivais, D., Kaski, K.: Communication with Family and Friends across the Life Course. PLOS ONE **11**(11), e0165687 (11 2016). https://doi.org/10.1371/journal.pone.0165687, `http://dx.plos.org/10.1371/journal.pone.0165687`
10. David-Barrett, T., Rotkirch, A., Ghosh, A., Bhattacharya, K., Monsivais, D., Behncke, I., Kertesz, J., Kaski, K.: Peer relations with mobile phone data: Best friends and family formation (8 2017), `http://arxiv.org/abs/1708.07759`
11. Dong, Y., Yang, Y., Tang, J., Yang, Y., Chawla, N.V.: 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 - KDD '14. pp. 15–24. ACM Press, New York, New York, USA (2014). https://doi.org/10.1145/2623330.2623703, `http://dl.acm.org/citation.cfm?doid=2623330.2623703`

12. Eagle, N., (Sandy) Pentland, A.: Reality mining: sensing complex social systems. Personal and Ubiquitous Computing **10**(4), 255–268 (5 2006). https://doi.org/10.1007/s00779-005-0046-3, `http://link.springer.com/10.1007/s00779-005-0046-3`
13. Eurostat: Eurostat Database, `http://ec.europa.eu/eurostat/data/database`
14. Games, P.A., Howell, J.F.: Pairwise Multiple Comparison Procedures with Unequal N's and/or Variances: A Monte Carlo Study. Journal of Educational Statistics **1**(2), 113 (22 1976). https://doi.org/10.2307/1164979, `https://www.jstor.org/stable/1164979?origin=crossref`
15. Granovetter, M.S.: The Strength of Weak Ties. American Journal of Sociology **78**(6), 1360–1380 (5 1973). https://doi.org/10.1086/225469, `https://www.journals.uchicago.edu/doi/10.1086/225469`
16. Hawkes, K.: The grandmother effect. Nature **428**(6979), 128–129 (3 2004). https://doi.org/10.1038/428128a, `http://www.nature.com/articles/428128a`
17. Jin, B., Peña, J.F.: Mobile Communication in Romantic Relationships: Mobile Phone Use, Relational Uncertainty, Love, Commitment, and Attachment Styles. Communication Reports **23**(1), 39–51 (4 2010). https://doi.org/10.1080/08934211003598742, `http://www.tandfonline.com/doi/abs/10.1080/08934211003598742`
18. Jo, H.H., Karsai, M., Karikoski, J., Kaski, K.: Spatiotemporal correlations of handset-based service usages. EPJ Data Science **1**(1), 10 (12 2012). https://doi.org/10.1140/epjds10, `http://www.epjdatascience.com/content/1/1/10`
19. Jo, H.H., Saramäki, J., Dunbar, R.I.M., Kaski, K.: Spatial patterns of close relationships across the lifespan. Scientific Reports **4**(1), 6988 (5 2014). https://doi.org/10.1038/srep06988, `http://www.nature.com/articles/srep06988`
20. Kalmijn, M.: Longitudinal analyses of the effects of age, marriage, and parenthood on social contacts and support. Advances in Life Course Research **17**(4), 177–190 (12 2012). https://doi.org/10.1016/J.ALCR.2012.08.002, `https://www.sciencedirect.com/science/article/pii/S1040260812000512`
21. Karsai, M., Kivelä, M., Pan, R.K., Kaski, K., Kertész, J., Barabási, A.L., Saramäki, J.: Small but slow world: How network topology and burstiness slow down spreading. Physical Review E **83**(2), 025102 (2 2011). https://doi.org/10.1103/PhysRevE.83.025102, `https://link.aps.org/doi/10.1103/PhysRevE.83.025102`
22. Kurniawan, S.: Older people and mobile phones: A multi-method investigation. International Journal of Human-Computer Studies **66**(12), 889–901 (12 2008). https://doi.org/10.1016/J.IJHCS.2008.03.002, `https://www.sciencedirect.com/science/article/pii/S1071581908000281#bib18`
23. Morey, J.N., Gentzler, A.L., Creasy, B., Oberhauser, A.M., Westerman, D.: Young adults use of communication technology within their romantic relationships and associations with attachment style. Computers in Human Behavior **29**(4), 1771–1778 (7 2013). https://doi.org/10.1016/J.CHB.2013.02.019, `https://www.sciencedirect.com/science/article/pii/S074756321300085X`
24. Onnela, J.P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., Barabási, A.L.: Structure and tie strengths in mobile communication networks. Proceedings of the National Academy of Sciences of the United States of America **104**(18), 7332–6 (5 2007). https://doi.org/10.1073/pnas.0610245104, `http://www.ncbi.nlm.nih.gov/pubmed/17456605http://www.ncbi.nlm.nih.gov/pubmed/17456605`

25. Palchykov, V., Kaski, K., Kertész, J., Barabási, A.L., Dunbar, R.I.M.: Sex differences in intimate relationships. Scientific Reports **2**(1), 370 (12 2012). https://doi.org/10.1038/srep00370, `http://www.nature.com/articles/srep00370`

26. Roberts, S.G.B., Dunbar, R.I.M.: Communication in social networks: Effects of kinship, network size, and emotional closeness. Personal Relationships **18**(3), 439–452 (9 2011). https://doi.org/10.1111/j.1475-6811.2010.01310.x, `http://doi.wiley.com/10.1111/j.1475-6811.2010.01310.x`

27. Roberts, S.G., Dunbar, R.I., Pollet, T.V., Kuppens, T.: Exploring variation in active network size: Constraints and ego characteristics. Social Networks **31**(2), 138–146 (5 2009). https://doi.org/10.1016/J.SOCNET.2008.12.002, `https://www.sciencedirect.com/science/article/pii/S0378873309000033?via%3Dihub`

28. Saramäki, J., Leicht, E.A., López, E., Roberts, S.G.B., Reed-Tsochas, F., Dunbar, R.I.M.: Persistence of social signatures in human communication. Proceedings of the National Academy of Sciences of the United States of America **111**(3), 942–7 (1 2014). https://doi.org/10.1073/pnas.1308540110, `http://www.ncbi.nlm.nih.gov/pubmed/24395777http://www.ncbi.nlm.nih.gov/pubmed/24395777`

29. Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M.M., Larsen, J.E., Lehmann, S.: Measuring Large-Scale Social Networks with High Resolution. PLoS ONE **9**(4), e95978 (4 2014). https://doi.org/10.1371/journal.pone.0095978, `http://dx.plos.org/10.1371/journal.pone.0095978`

30. Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., Campbell, A.T.: StudentLife. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct. pp. 3–14. ACM Press, New York, New York, USA (2014). https://doi.org/10.1145/2632048.2632054, `http://dl.acm.org/citation.cfm?doid=2632048.2632054`

31. Wasserman, S., Faust, K.: Social Network Analysis. Cambridge University Press, Cambridge (1994). https://doi.org/10.1017/CBO9780511815478, `http://ebooks.cambridge.org/ref/id/CBO9780511815478`

32. Welch, B.L.: The generalization of 'Student's' problem when several different population variances are involved. Biometrika **34**(1-2), 28–35 (1 1947). https://doi.org/10.1093/biomet/34.1-2.28, `https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/34.1-2.28`

33. Zachary, W.W.: An Information Flow Model for Conflict and Fission in Small Groups. Journal of Anthropological Research **33**(4), 452–473 (12 1977). https://doi.org/10.1086/jar.33.4.3629752, `https://www.journals.uchicago.edu/doi/10.1086/jar.33.4.3629752`