EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE Open Access

Temporal patterns behind the strength of persistent ties

Henry Navarro[1], Giovanna Miritello[1,2], Arturo Canales[3] and Esteban Moro[1]*

*Correspondence:
emoro@math.uc3m.es
[1] Departamento de Matemáticas & GISC, Universidad Carlos III de Madrid, Avenida de la Universidad 30, Leganés, 28911, Spain
Full list of author information is available at the end of the article

Abstract

Social networks are made out of strong and weak ties having very different structural and dynamical properties. But what features of human interaction build a strong tie? Here we approach this question from a practical way by finding what are the properties of social interactions that make ties more persistent and thus stronger to maintain social interactions in the future. Using a large longitudinal mobile phone database we build a predictive model of tie persistence based on intensity, intimacy, structural and temporal patterns of social interaction. While our results confirm that structural (embeddedness) and intensity (number of calls) features are correlated with tie persistence, temporal features of communication events are better and more efficient predictors for tie persistence. Specifically, although communication within ties is always bursty we find that ties that are more bursty than the average are more likely to decay, signaling that tie strength is not only reflected in the intensity or topology of the network, but also on how individuals distribute time or attention across their relationships. We also found that stable relationships have and require a constant rhythm and if communication is halted for more than 8 times the previous communication frequency, most likely the tie will decay. Our results not only are important to understand the strength of social relationships but also to unveil the entanglement between the different temporal scales in networks, from microscopic tie burstiness and rhythm to macroscopic network evolution.

Keywords: social networks; tie strength; temporal patterns

1 Introduction

Social networks are dynamic objects, they grow and change over time through the addition of new ties or the removal of old ones, leading to an ongoing appearance and disappearance of interactions in the underlying social structure [1, 2]. Identifying the different mechanisms by which ties form or decay is a fundamental and challenging question of individual human behavior. But also it can unravel the processes behind group, community and network dynamics that shape our social fabric. And in turn, how network evolution impacts important processes in our society like cooperation [3], disease spreading [4] or information diffusion [5–7]. On the other hand, understanding tie persistence may shed light on the circumstances under which an observed interaction can actually be considered a genuine social relationship [8, 9]. This will lead to predict its presence and future potential strength in the different processes happening in social networks.

Most of the understanding on the dynamics of tie formation and decay comes from the determination of microscopic factors governing tie formation and persistence [10]. Special attention has been given to endogenous factors, i.e. those properties that can be extrapolated from the network itself to predict future tie behavior. Intensity of previous interactions, reciprocity, network proximity, triadic closure or the existence of common friends are not only predictors of tie formation [11], but also of its persistence in the future [8, 12]. In the context of Granovetter's theory of *strength of weak ties*, strong ties are those which are more likely to persist, since they are structurally embedded (common friends) and are more intense (number of interactions). On the other hand bridges between communities are weak and, as Burt found in [13], they are more likely to decay in the future. Intensity and embeddedness are thus commonly acknowledged as properties behind a strong and/or persistent tie.

Despite these findings, we still have not a comprehensive understanding of what are the main properties of human interaction that make social ties to persist. This is largely due to the lack of quality data: although some online social networks have explicit mechanisms to '*unfollow*' (Twitter) [14] or '*unfriending*' (Facebook) [15] other users, access to structural or intensity data in those platforms is limited. On the other hand, most studies infer tie decay from absence of tie activity in large databases [8, 12]. This is a potential problem since, given the large burstiness of human interaction [6, 16], large inactivity periods could be mistaken as tie decay events. Thus, although previous studies agree on the general importance of tie structural embeddedness, intensity or reciprocity to predict its future persistence [8, 12], they still provide an incomplete picture of what are the main properties that make ties persistent. As it was done in the problem of tie prediction, can we build efficient models based on endogenous properties of ties to predict if a social relationship is bound to decay?

In this paper we address those questions by studying tie persistence in human communication using a large longitudinal database of 19 month of mobile phone calls. The large duration of the database allows us to accurately assess the presence of a tie by using the method introduced by Miritello *et al.* [17] which splits the observation period in different time windows and uses each of them to characterize and assess tie presence. But more importantly, having a detailed and large longitudinal database for human communication allows us to characterize better the patterns of communication within a tie and see if temporal properties of human interaction are predictors of tie persistence. Although simple temporal properties have been considered before in the problem of tie prediction [18] and strength estimation [12, 19], here we show that the tie persistence is also encoded in the bursty patterns of communication between people. Furthermore, by building a highly accurate predictive model based on different tie features (structural, intensity, intimacy and temporal) we are able to show that temporal properties are indeed as important as intensity and much more than structural properties in predicting tie persistence. Our results show that it is possible to build simple predictive models of network evolution based only on the temporal and intensity properties of the human interaction.

2 Measuring the strength of a tie

We study a sample of 100,000 ties drawn randomly from the *Call Detail Records* (CDR) of 20 million people from a single mobile phone operator over a period of 19 months. As in [17] we divide the time interval in three periods: the 7 months in the middle Ω define

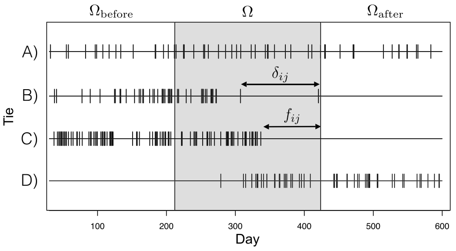


Figure 1 Detecting tie decay and strength. Definition of observation periods and examples of call activity for 4 given ties. Any vertical segment is a call between the users in a particular tie. Our 19 months database is divided in three periods, where the 7 months in the middle Ω is our observation period where all the tie features will be measured. The period Ω_{after} is used to asses if ties are persistent, i.e. if there is activity in the tie. For example, ties **(A)** and **(D)** are persistent, while ties **(B)** and **(C)** are said to have decayed in Ω_{after}. All ties have similar values of number of calls in the observation period with $w_{ij} \in [30, 40]$. We also show specific examples of one inter-event time δ_{ij} (tie **(B)**) and freshness f_{ij} (tie **(C)**).

our observation and measurement period for the ties. We only select 60,592 ties in which there are at least 5 calls in Ω between users, and among those calls there has been at least one call in each direction. We only consider ties which have been observed at least for 50 days, to prevent very short ties. As in [17], the first and last periods of 6 months Ω_{before} and Ω_{after} are used to assess whether the tie has formed and/or decayed. In our particular case and since there is no explicit information about whether social interactions stop, we will say that the tie between user i and j has decayed if there are no calls between them in Ω_{after}. This functional definition of the existence of a tie underestimates the possibility of having another call after those 6 months, but as it was shown in [17], only 3% of ties contain such long inter-event times δ_{ij} between calls (see Figure 1), which shows that our method is subject only to a small error. It is important to understand that since activity within ties is bursty, large inter-events between interactions are likely and thus they might be mistaken as tie decay. In particular, in our database we find that the average time between calls in a tie is $\overline{\delta}_{ij} = 14$ days (with a standard deviation of 18 days), and thus we might get spurious effects if Ω_{after} is of the order of a month, as interactions may fall outside the Ω_{after} period. See the *Methods* section for further description of the mobile phone dataset. We have also considered another (smaller) database of Facebook communication through wall posts. Since the results on both databases are similar we discuss here only the mobile phone database and refer to the *Methods* section for further details about the Facebook database analysis.

To characterize the strength of the tie we will find those features that can anticipate its persistence. Thus, we will implicitly identify strong relationships with persistency, while weak ties are those more likely to decay. This *dynamical definition of strength* is then a much more functional form of describing its utility in present and future social processes and operationalizes Granovetter's idea that strong ties are those which are more likely to persist. To describe which tie features are related with its *dynamical strength (persistence)*, we will also follow Granovetter's notion of *static strength* of an interpersonal tie [20]: 'the strength of a tie is a combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie'. Within that framework, we define four categories of tie features: intensity, temporal, structural and intimacy features, and we will try to characterize which ties are the strongest (more persistent) according to these variables. Intensity, frequency and intimacy features will refer to properties of the communication patterns between users, while structural variables are those derived by understanding how the tie is embedded in the rest of the social network. Given the nature of our data, our features will be constructed solely taking into account the information about call events between users. Our working assumption is that there is enough information in those events to predict the persistence of the tie.

Some of the variables are adapted from previous works both in tie formation and decay prediction [12, 15, 17, 21], but others are introduced for the first time in this work. Specifically we introduce a number of variables that take into account the temporal patterns of the communication between users [1, 17]. Contrary to the static and aggregated version of relationships and networks, ties and networks are always evolving: not only communication between users is highly bursty and correlated in time [6, 7], but also the dynamical strategies by which users create and destroy ties are very different [17, 22]. The hypothesis we investigate in this paper is whether those patterns convey information about the fate of a social relationship. For example, if the periodicity or burstiness of how two people communicate or if they are involved in very fast social creation and destruction of ties can inform us about the persistence of social ties.

2.1 Intensity features

The first group of variables describe the amount of communication between users. Stronger relations imply a more frequent relationship which we can quantify by the number of calls w_{ij} between users. This variable is highly heterogeneous in our database in a similar way as other similar works in the literature [23] (see Figure 5). Specifically we find that the average number of calls is $\overline{w_{ij}} = 76$ while it varies from a minimum of 5 and a maximum of 2468 calls per tie. To take into account this heterogeneity, the rest of the variables we will consider are calculated with respect to that level of activity per tie. For example, instead of considering the total duration of calls per tie we will consider the average duration d_{ij}. On the other hand, several works have found that if the tie is highly reciprocal, the relationship is stronger and thus is less likely to decay [8, 12, 24]. Our database contains information about which user initiates the call so we can measure $w_{\overrightarrow{ij}}$, the number of calls between i to j initiated by i. Using this, we define the level of reciprocity in between users i and j as

$$r_{ij} = \left| \frac{w_{\overrightarrow{ij}}}{w_{ij}} - \frac{1}{2} \right|. \tag{1}$$

Note that this variable take values between 0 and 1/2. When user i initiates most of the calls in the tie, then $w_{\overrightarrow{ij}} \simeq w_{ij}$ and $r_{ij} \simeq 1/2$. On the contrary, when the number of calls from i to j is equal to the number of calls from i to j, we have that $w_{\overrightarrow{ij}} \simeq w_{ij}/2$ and then $r_{ij} = 0$. Thus larger values of r_{ij} indicate less reciprocity.

2.2 Structural features

Formation and decay of a tie is also related with the social structure around it. People tend to form groups and in particular, people tend to form relationships with friends of friends (triadic closure) which leads to high clustering around a tie [10]. This is the reasoning behind Granovetter's influential 'strength of weak ties' argument which implies that not also structural embedded ties are more likely to arise in a social network but they are also more persistent, a result corroborated by Burt in different works [13, 25]. Although there are many metrics to quantify embeddedness of a tie within the social network, we will use the topological overlap o_{ij} defined as the fraction of neighbors of i and j which are commonly shared [23]. Specifically,

$$o_{ij} = \frac{|n_i \cap n_j|}{|n_i \cup n_j|}, \tag{2}$$

where n_i and n_j are respectively the set of neighbors of nodes i and j and $|n_i|$ indicates the number of them. Note that, this variable takes values between 0 and 1, because if i and j have no common neighbors, then o_{ij} will take value 0. On the contrary, if i and j call to the same circle of id's o_{ij} will take value 1. The topological overlap is then a variable measuring the (normalized) number of 'common friends' between two nodes.

The topological overlap is a particular way to measure the structural information around a tie. Another metric we will consider is the level of social connectivity around a tie. In particular, if k_i and k_j are the number of neighbors of i and j we will construct the geometric mean of connectivity $k_{ij} = \sqrt{k_i k_j}$. This variable is introduced to take into account the effect of the different importance of a tie for the users involved in the relationship. If k_{ij} is small, the tie between i and j is important for both or one of them, while if k_{ij} is large, then it is just another tie among the many they have. Variations of structural connectivity around a tie have been considered in other works studying tie strength and dynamics [12, 19].

2.3 Intimacy features

Following Granovetter's hypothesis of a strong tie, the intimacy (mutual confidence) between two nodes could provide a better characterization of the tie and allow a more accurate prediction of its dynamics. As opposed to other studies in social networks [19] our mobile phone database does not contain any information about the context and content of the call. Thus we quantify the mutual confidence by the day or hour when the calls are made. Specifically, we consider the fraction of calls within a tie that are made after 8 pm and during the weekend, μ_{ij}^{int}. As was shown recently, calls made in the evening and at night are typically focused on a small number of emotionally intense relationship [26] and thus, quantifying the amount of communication happening at that time of the day can give us a proxy for intimacy.

On the other hand, demographic differences between users have an impact in tie dynamics. For example, the temporal communication patterns formed by groups of males or females are different [27], and those patterns can be associated with the different preference strategies of both sexes across the lifespan [28]. To quantify those relationship preferences, we consider the age and gender difference between the users participating in a tie. Age difference age_{ij} is measured as the absolute value of the difference in years while gender difference is a dichotomous variable where $gender_{ij} = 1$ if both users have same gender and $gender_{ij} = 0$ if they are different.

2.4 Temporal features

Finally we characterize the temporal patterns within and around the tie. Since communication within the tie is very heterogeneous (see Figure 1), we want to understand whether that heterogeneity might reveal something about the persistence of the tie. The first variable we consider is the *freshness* of the tie f_{ij}, i.e. the time since the last call between i and j at the end of Ω [12, 19]. Since activity within ties is very heterogeneous, we consider the relative *freshness* as the relative time elapsed from the last call compared to the typical time between calls in the tie $\hat{f}_{ij} = f_{ij}/\overline{\delta}_{ij}$ where $\overline{\delta}_{ij}$ is the average inter-event time between calls.

At the same time we also consider the age of the tie as the time of the first call between users in our database $t_{ij}^{\min}$ measured in days.

Another feature we consider is the burstiness of the communication patterns. The hypothesis we want to test is whether more regular communication patterns could reflect stronger/more persistent ties. For example, strong relationships like family and close friends require constant communication and thus they might have more regular patterns than acquaintances (see [29] and references therein). Although there are many ways to characterize burstiness of events [30], we will use two simple metrics. The first one is the coefficient of variation of the inter-event times $cv_{ij} = \sigma_{ij}/\overline{\delta}_{ij}$, where $\overline{\delta}_{ij}$ is the average inter-event time between two calls and σ_{ij} is their standard deviation [2]. If $cv_{ij} \gg 1$ then communication is very bursty, with large untypical periods of time in which users didn't communicate (see for example tie B in Figure 1), while if $cv_{ij} \ll 1$, communication was very regular, happening almost at the same time intervals (see tie A in Figure 1). The value $cv_{ij} = 1$ correspond to the Poissonian homogenoeus case in which inter-event times are distributed randomly along the Ω period [30]. Another way to characterize the burstiness is to quantify how many communication events happened in bursts or rapid consecutive successions of calls (we will call them *chats*) [6, 31]. To do that we calculate the fraction of calls μ_{ij}^{chats} that happened only with 5 minutes difference between them.

Finally, another reason why a tie decays is simply because users involved in the tie have very different dynamical social strategies. As was found in [17] humans constantly create and destroy ties and they have different strategies to do that. While some individuals create and destroy a lot of ties (*explorers*), others tend to maintain their social circle (*keepers*). If both users in a tie are explorers, the probability for the tie to decay is high. To measure how dynamical are the strategies of users in a tie we consider a_i, the number of ties created by user i in period Ω. As in [17] we say that a tie is created in Ω if there is no call between users in Ω_{before}. The ratio between the number of created ties and the total number of ties $a_i/k_i \in [0,1]$ describe how frequent user i changes her social neighborhood. If $a_i/k_i \simeq 1$ it means that most of the ties of user i where created during Ω (i.e. the user *social explorer*), while if $a_i/k_i \ll 1$ most of the ties are stable (*social keeper*). To characterize how dynamical are the strategies of both i and j we consider the geometrical mean

$$a_{ij} = \sqrt{\frac{a_i}{k_i} \cdot \frac{a_j}{k_j}}. \tag{3}$$

If both i and j are explorers, $a_{ij} \simeq 1$ and the tie is more likely to decay since it connects users with highly dynamical social strategies, while if they are both keepers, $a_{ij} \simeq 0$ and the tie most likely will persist.

Table 1 summarizes the features considered to assess the dynamical strength of persistent ties. Before constructing our models and because of the large heterogeneity found in connectivity, activity and burstiness across ties in social networks, we scale and normalize our variables before using them in a model. For example, we consider $\log w_{ij}$ instead of w_{ij} since the distribution of number of calls per tie is heavy skewed in mobile phone databases [23]. On the other hand burstiness within ties make variables like cv_{ij} or $\hat{f}_{ij}$ also very heavy-tailed across our dataset. Thus we also use a logarithmic scaling for them. Although they are logarithmically scaled, in the rest of the paper we denote them by its original name for sake of clarity, unless were numerical values are given (for example in Figure 3). Finally,

Table 1 Features of ties between user *i* and *j* considered to characterize the dynamical strength (persistence) of the ties, including their (normalized) complexity measured in computational time

Type	Feature	Description	Computational complexity
Intensity	w_{ij}	Total number of calls	1.00
Intensity	d_{ij}	Average duration of calls	1.00
Intensity	r_{ij}	Reciprocity of calls	1.12
Structural	o_{ij}	Topological overlap	1.82
Structural	k_{ij}	Connectivity diversity	1.33
Intimacy	μ_{ij}^{int}	Fraction of calls after 8 am and weekends	1.05
Intimacy	age_{ij}	Age difference in years	0.18
Intimacy	$gender_{ij}$	Gender difference	0.15
Temporal	$\hat{f}_{ij}$	Relative freshness	1.01
Temporal	t_{ij}^{min}	Age of tie (in days).	1.01
Temporal	cv_{ij}	Inter-event time coefficient of variation	1.11
Temporal	$\mu_{ij}^{\mathrm{chats}}$	Fraction of consecutive calls (5 mins.)	1.31
Temporal	a_{ij}	Users' Activity diversity	1.21

since the correlation between the variables is small, we keep all features in our analysis excepting t_{ij}^{min} which is moderately correlated with w_{ij} (see *Methods* section to learn about the preprocessing and selection of variables).

3 Results

A simple inspection of how persistence depends on some tie features corroborates some results found in the literature. For example, as Burt found in [13] we observe that weak ties with small topological overlap have a higher probability to decay (see Figure 2(A)), i.e. bridges are more likely to decay while persistent ties are those embedded within communities. Note that this effect can amount to a 50% change in probability from ties with no overlap $o_{ij} = 0$ to the largest overlap observed in the database $o_{ij} \simeq 0.5$. The same happens for tie age: the older the tie, the more persistent it is as we can see if Figure 2(D). Similarly to [19] we find that the time since the last communication also reveals how likely it is to observe activity in the tie again: most recent activity implies that the tie will persist in the future (see Figure 2(B)). Finally, we find that some temporal features are strongly correlated with tie persistence. For example in Figure 2(C) we find the interesting result that more bursty communication within a social tie is correlated with tie decay.

Although these individual results demonstrate the potential predictive power of our tie features, to get a complete picture of tie persistence we build a predictive model of tie decay based on *all* the features introduced in the last section. We define two different prediction models depending on the reference frame used to characterize tie strength features. In the first one (*Model 1*) we used a fixed reference frame for all ties, namely we try to predict if the tie decays in Ω_{after} by observing its features along Ω. Although this is the traditional setting for tie persistence prediction, the features calculated during Ω might be impacted by the fact that the tie decayed early in the interval Ω (see for example tie C in Figure 1). If this happens, variables like the number of calls, their duration, or the structural overlap are going to be naturally smaller just because the tie decayed earlier, making it difficult to disentangle what part of the prediction power comes from properties of the tie before or after it decays. For this reason we will build another predicting model *Model 2* in which we will only consider those ties that have a call within the last two weeks of Ω. This way we will use a relative reference frame in which we want to understand what properties of an

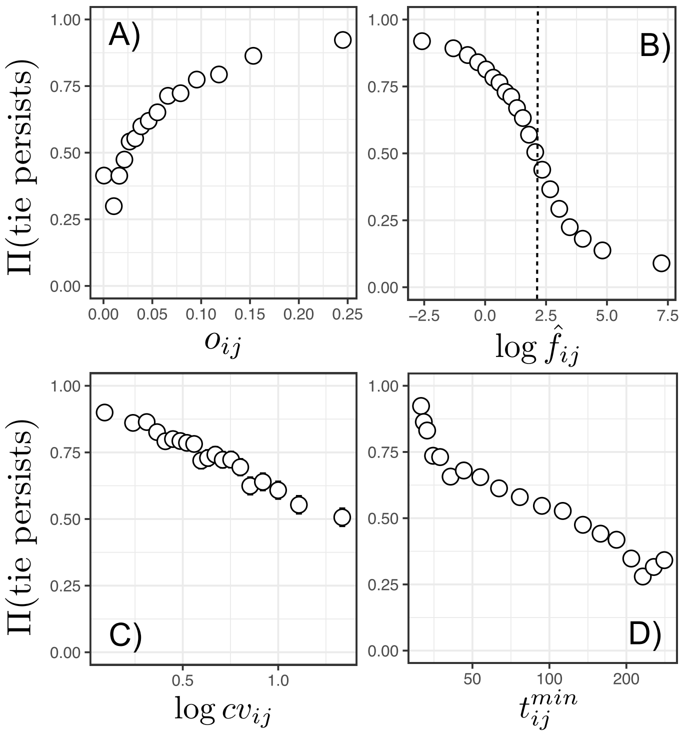


Figure 2 Impact of some features on tie persistence. Conditional probability of persistence as a function of the different variables of the tie: **(A)** topological embeddedness, **(B)** relative freshness, **(C)** coefficient of variation and **(D)** time of the first call in our database. In **(C)** only ties with $w_{ij} \in [20, 50]$ are considered. The vertical line in **(B)** indicates the critical relative freshness $\hat{f}_{ij} = 8.33$, where $\Pi = 1/2$. Error bars are showed when they are bigger than symbol size.

existing tie have more impact in its immediate future stability. Both models are important to understand the dynamics of a tie, its stability, and in general, the evolution of networks. But *Model 2* might give a more direct understanding of what defines a strong social relationship without requiring a long time interval to observe if there is a significant decay in the activity of the tie.

To predict tie persistence we build a classification model using simple logistic regression (LogR) models where the positive class is tie persistence, that is, that we observe at least a communication event in Ω_{after}. We use a train dataset using 75% of our ties and 10-fold cross validation to fit the probability for a tie to persist using the inverse logit function

$$\Pi(\text{tie } ij \text{ persists}) = \frac{1}{1 + e^{-\beta_0 - \sum_{l=1}^{n} \beta_l x_l}}, \tag{4}$$

where x_l are the features introduced in the last section and β_l are the coefficients obtained in the fit. Note that positive values of β_l indicate that the variable x_l has a positive effect in the persistence of the tie: larger values of x_l increase the probability for the tie to persist. The performance of the model (see Table 2) is measured using the rest 25% of our ties, achieving values around 0.8 for its accuracy, sensitivity and specificity, showing the good balance of our model detecting both classes (persistent and decaying ties). Details of how the predicting model was constructed can be found in the *Methods* section.

Table 2 Regression results for the tie persistence using generalized linear models for the two prediction models

Feature	Model 1			Model 2	
	Full	Simplified	Simplified′	Full	Simplified
w_{ij}	1.491***	1.180***		2.096***	2.229***
	(0.021)	(0.015)		(0.060)	(0.057)
d_{ij}	0.105***			0.109**	
	(0.012)			(0.034)	
r_{ij}	−0.094			−0.044**	
	(0.012)			(0.032)	
o_{ij}	0.241***			0.286***	
	(0.017)			(0.046)	
k_{ij}	0.039**			−0.026	
	(0.012)			(0.033)	
μ_{ij}^{int}	0.084***			0.151**	
	(0.012)			(0.032)	
age_{ij}	0.001			−0.021	
	(0.012)			(0.034)	
$gender_{ij}$	0.079***			0.035*	
	(0.012)			(0.033)	
$\hat{f}_{ij}$	−1.102***	−0.660***	−0.611***		
	(0.016)	(0.008)	(0.007)		
cv_{ij}	−0.362***			−0.653***	−0.759***
	(0.014)			(0.037)	(0.034)
μ_{ij}^{chats}	0.029*			0.036	
	(0.014)			(0.038)	
a_{ij}	−0.310***			−0.316***	
	(0.013)			(0.036)	
Constant	0.681***	−2.364***	1.053***	0.779***	0.748***
	(0.014)	(0.042)	(0.014)	(0.039)	(0.034)
Number of points	45,444	45,444	45,444	6684	8268
AUC	0.864	0.847	0.755	0.875	0.866

Performance	Model 1			Model 2	
	Full	Simplified	Simplified′	Full	Simplified
Accuracy	0.802	0.766	0.721	0.803	0.796
Sensitivity	0.828	0.777	0.652	0.815	0.808
Specificity	0.767	0.753	0.781	0.787	0.760

Coefficients are shown with uncertainties (standard errors) in parentheses. Model *Full* include all the features described in the text, while model *Simplified/Simplified′* only includes the most important two/one feature(s). *Note:* *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

The results for the different models are presented in Table 2, where we can see that, as expected, variables like the number of calls w_{ij}, mean duration d_{ij} or topological overlap o_{ij} have a positive effect in tie persistence [8, 12]: the larger they are the more likely the tie will persist in the future. Interestingly, the same happens with gender difference: ties between individuals with equal gender are more persistent than those between persons of different gender, a reflection of the same-gender homophily previously found in the most stable relationships [28]. However, other well-studied variables like reciprocity, connectivity levels or age difference seem not to be important for tie persistence.

Temporal variables play a major role in the models. Specifically, in *Model 1* the persistence of the tie is highly determined by the (relative) freshness $\hat{f}_{ij}$, i.e. how much time has passed since the last communication between users: as we can see, the coefficient is negative, which means that larger times since the last communication mean smaller probability for the tie to persist. Other temporal variables like the coefficient of variation and number

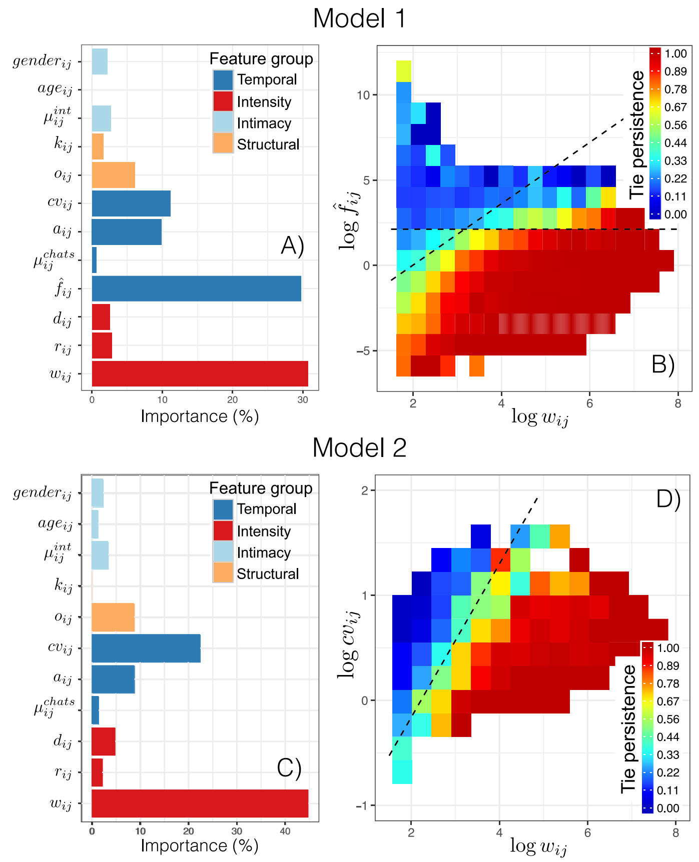


Figure 3 Feature importance in tie persistence. (A) Importance of the variables in the full models in Table 2. Importance is measured as the normalized % of the t-statistics for each model parameter. **(B)** Density plot of the average persistence of ties as a function of the two most important variables in **(A)**, namely, normalized freshness and total number of calls for model 1 and coefficient of variation and total number of calls for model 2. The dashed line shows the solutions of equation $\Pi = 1/2$ for the simplified models in Table 2.

of chats have some impact on the persistence of the tie. For example, larger number of rapid consecutive calls (larger μ_{ij}^{chats}) or more regular patterns (smaller cv_{ij}) yield to better stability of ties, an interesting result showing that high frequency patterns of communication between users also encode some information about how strong the tie is. Finally, the coefficient for a_{ij} is negative, i.e, if users participating in the tie have more *explorer* behavior, the tie has lower probability to persist.

However, not all the variables have equal importance in the persistence model. All together, temporal variables are the most important variables in the model: they amount to around ∼51% of the importance in our predictive model (see Figure 3), while intensity variables giving ∼36% of the importance and finally structural and intimacy variables representing less than ∼10% (each) of the model importance. The relative small importance of well studied properties like the topological overlap o_{ij} could be due to the Granovetter effect, i.e. because o_{ij} and w_{ij} are moderately correlated, the former will have less importance in the model since its effect is already included in w_{ij}. As we can see in Figure 3 it is remarkable that just two variables (number of calls w_{ij} and relative freshness $\hat{f}_{ij}$ or coefficient of variation cv_{ij}) have most of the importance in the model to the point that a simplified model based on only those two variables achieve similar levels of performance (see Table 2) to the full model. In the case of *Model 1*, actually, just the number of calls and the relative freshness achieve a high accuracy (77%), a result that can be shown graphically in Figure 3 where the diagonal dashed line corresponds to the $\Pi = 1/2$ probability. Interestingly, similar level of accuracy is found for the really simple model based on just the relative freshness (horizontal line in Figure 3). In that case $\Pi = 1/2$ corresponds to a critical relative freshness of $\hat{f}_{ij} = 8.33$ so ties with larger/smaller values have less/more than 50% probability to persist. This result shows that ties in which the natural rhythm of their communication is halted have higher probability to decay. Specifically we find this happens when the last interaction between users happened at least 8.33 times their typical inter-event time. As an example, if two users typically called themselves each day in the past and more than 2 weeks have elapsed since their last communication, the tie might have decayed.

In the case of *Model 2* we also find that intensity and temporal properties are the most important variables to explain tie persistence giving respectively $\sim 51\%$ and $\sim 32\%$ of the importance of the model, as we can see in Figure 3. But also we can explain most of its accuracy by a simplified model in which only the number of calls and the coefficient of variation are considered, see diagonal dashed line in Figure 3. The strong importance of cv_{ij} in the model signals a very interesting fact: for a given level of activity w_{ij}, ties which are more bursty (high cv_{ij}) have more probability to decay. This finding suggest that special attention paid by users to maintain a periodic communication might be an indication of a stronger and more persistent relationship, while highly bursty and heterogeneous call patterns might be a sign of an informal or casual relationships that could decay in the near future.

Another dimension controlling the effectiveness of the different variables in a predictive model is their complexity. While some of the variables are easy to compute for a given dataset, other features like topological overlap o_{ij} or users activity diversity a_{ij} are very complex, i.e. they need larger computational time. Table 1 shows the computational time (in seconds) of our own code to compute each tie feature normalized to the time it takes to compute w_{ij}. Although the actual times could depend on the different code implementation, our results agree with the expected result that metrics that require to compute next neighbors' properties are very costly. For example, structural features like topological overlap or social connectivity take up to 1.82 times the total number of calls. On the other hand, temporal features are cheaper to compute. This result, together with the low predictive power of traditionally considered variables like o_{ij} or r_{ij} shows that temporal features could be much more efficient to detect and predict future tie persistence in a social network.

4 Discussion

Human behavior display very different temporal patterns due to many constrains like circadian rhythms, cognitive limits or finite capacity to perform tasks [1, 32]. Since most of those constrains are common to human nature, those patterns show also a large degree of universality across individuals. Interestingly, deviations from universal rhythms can inform us about changes of behavior related to, for example, unemployment [33], health conditions [34], or crowd events [35, 36]. Along this line, our research also shows that future network dynamics is encoded in the relative properties of the temporal patterns of communication between individuals and that those temporal properties have more predicting power than structural, intensity or intimacy features of the communication. Specifically, we find that if tie activity is not observed for more than 8 times its typical inter-event time, the tie has a great probability to decay, a result that indicates that each tie has a natural rhythm and that when communication is halted for a long time it will probably decay. More importantly, although recent research has found that burstiness affects a large number of human activities and some explanations have been given to explain its universality [16], our results show that relative burstiness could be also related to the weakness of ties and that those ties that show excessive burstiness might decay in the future. Since burstiness in ties slows down information spreading [6], we have found that more bursty ties are not only weaker to transmit information, but also they are more prone to disappear, making them extremely fragile for the structural and functional processes happening in social networks.

Our analysis reveals that there is a large entanglement between the different time scales present in social networks and that analyses based on pure structural static features of human relationships might give a partial and biased description on the evolution of our communities, groups and societies [1, 37]. For example, short time scales (minutes, time between calls in a tie) seem to foresee the decay of ties in the future (month time scale). More importantly, it seems that temporal properties of ties are better and more efficient descriptions of tie persistence than structural features, which will allow faster and simpler detection of changing events in the topology of social networks. In fact we find that structural features like topological overlap play a minor role in our model. This is probably the result of the moderate correlation between the strength and embeddedness in social networks (the Granovetter effect [20]), but also shows that a better picture of strong/persistent ties can be obtained just by looking at temporal and intensity features of social relationships. Our results are in line with recent measures of strength of social ties in social media [19] where structural variables account only for 4.5% of tie strength. The same small impact of common friends was found in detecting tie persistence [12]. This body of research and our results seem to imply that, although structural features are very important (and probably the only) predictors of future formation of a tie [11], once the tie is formed its strength or persistence is immediately encoded into the intensity and temporal features of the interaction. Thus, structural features are important in the tie prediction problem, while temporal properties might be more efficient in the persistence problem.

Finally, a possible explanation of our results might be in the way people share their attention and time over their relationships, giving more frequent and more regular attention to stronger ties than to the weak ones. As we know, humans are bounded by time, money or cognitive limits and they make decisions to share their time across tasks (including the social ones) causing irregular (bursty) activity. Our findings show that strong and persistent ties suffer less from those bursty patterns, indicating that those ties might have different weight in evaluating how to share our time [22, 38]. We hope our results will help future research to identify better what is the origin of the temporal signs of strong and/or weak ties in social networks.

5 Methods

5.1 Mobile phone data

As in [17] the data used in this study has been obtained from the Call Detail Records (CDR) database of a unique mobile phone operator in a single country. We focused exclusively on voice calls records, filtering out short text messages, multimedia messages and operator calls. Each subscription is anonymized such that it is not possible to recover personal information of the users. We filtered out all the incoming or outgoing calls that involve other operators due to the partial access we have to the activity of other providers. To avoid business-like subscriptions, which usually appear as users with a huge number of connections and calls never returned, we only retain ties which are reciprocated, which leads to the removal of about the 50% of the total links in our database. This restriction also eliminates calls to wrong numbers, telemarketing-type calls, customer service lines, etc. But it might eliminate genuine social interactions in which calls are not reciprocated. However, given that the observation window is 7 months long, the probability that there is not a reciprocated call in a genuine social connection in such a long window is very low. Within this approach, we neglect the directionality of links and consider a call from user i to user j equivalent to a call from j to i.

To disentangle the dynamics of ties creation/removal from their call activity, we use the first 6 months to determine if ties have being created (crucial to determine the a_{ij} variable) and the last 6 months to assess the persistence of the tie. Since we are interested only in tie dynamics between individuals, we have to take into account the problem of subscription and churn of users in our database. For example, subscription of a new user and its communication with other users in our database results into formation of many new ties for the new subscriber. The same would happen for the decay of ties of a subscribe that churns from the company. To mitigate this problem, we only keep active users in our data set: in particular, we only consider those users who are involved (as calling or as called party) at least in one communication event in each of the three subintervals in the 19 months and also if they are present in the database at least one month before Ω and are still active one month after Ω. This latter filter prevents spurious effects in the analysis of tie dynamics just because individuals subscribe/unsubscribe just before/after Ω; for example, we could have observed an apparent rapid growth of their social network at the beginning of the observation window or a fast dissolution at its end [5]. These results in the removal of about the 17% of nodes and the 37% of reciprocated links within Ω. In our analysis we have considered 100,000 random ties from the remaining reciprocated links of the mobile phone graph that have some activity in Ω. Finally, in our modeling we have only consider the 60,592 ties which are sufficiently active (more than 5 communication events in Ω) that have a duration of more than 50 days to prevent very short ties.

5.2 Prediction models

To predict tie decay/persistence we have used a simple logistic regression model where the positive class is that the tie persists, that is, that we observe at least a communication event in Ω_{after}. Since the fraction of ties that decay is small (only 20% in our sample) our classification problem is slightly unbalanced, which might cause problems when training our algorithm. To palliate this problem we use the SMOTE algorithm [39] to generate synthetic cases for the minority class (decay) so that the number of ties that persist and decay is around 50%. We split our new dataset into a train and test samples which contain respectively 75% and 25% of the ties and use 10 fold cross-validation to train the model with Area Under the Curve (AUC) as the performance metric. Final performance of the model is evaluated using the 25% test sample of the data.

To test that our results are not due to the particular algorithm used to predict tie persistence, we have also used other prediction models for this two-classes classification prob-

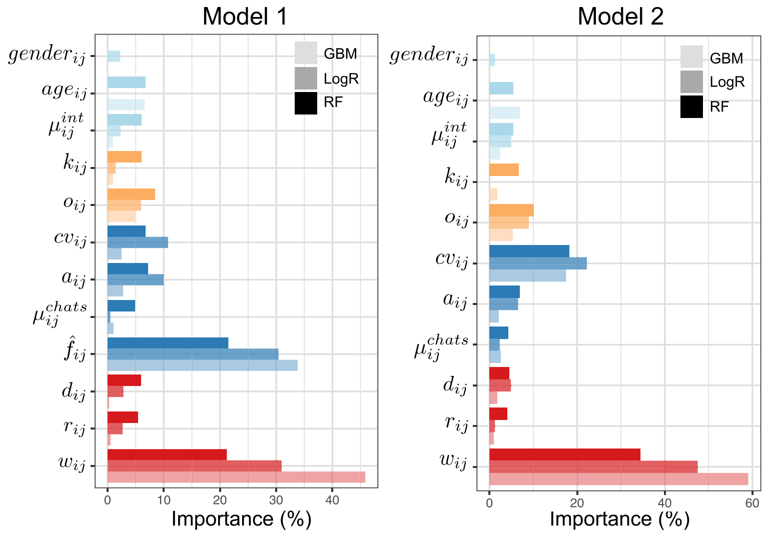


Figure 4 Comparison with other algorithms. Comparison of the importance of the variables for the two models using different algorithms, namely Random Forests (RF), Logistic Regression (LogR) and Generalized Boosted Regression Models (GBM). As we can see models agree on the relative importance of individual and group variables. Color of the variables correspond to the intimacy (light blue), structural (orange), temporal (blue) and intensity (red) groups.

lem. Specifically we have used Random Forests (RF) and Generalized Boosted Regression Models (GBM) [40]. As we can see in Figure 4 results are very similar for the different importance of variables. However accuracy is bigger in RF (90% in Model 1 and 87% in Model 2) and GBM (83% for Model 1 and Model 2) when compared with the logistic regression (LogR). This comparison shows that our results do not depend on the actual algorithm used to build the predictive algorithm and that the importance of temporal variables is a genuine finding in our data.

Finally, we have also tested the sensibility of our results on the threshold in the number of calls used to consider the ties. Figure 3 shows already that the effect of variables like relative freshness $\hat{f}_{ij}$ and coefficient of variation cv_{ij} is important even for large values of w_{ij}. To further support this observation, we have trained models 1 and 2 using different thresholds for w_{ij}. Results are presented in Table 3, where we can see that the performance and relative importance of the variables is maintained for different thresholds.

5.3 Normalization and selection of tie features

In the logistic regression classifier is common to implement some kind of normalization of variables through transformations. This is specially important when variables have highly skewed distributions as is typically found in variables describing human activity and behavior. In our case variables like the intensity w_{ij}, average duration d_{ij}, relative freshness $\hat{f}_{ij}$, time since the first call $t_{ij}^{\min}$ and coefficient of variation cv_{ij} are heavy-tailed distributed and thus we have log-transformed them before using them in our models. As we can see in Figure 5, after this transformation, the histogram of the main variables used in our models is more homogeneous.

Finally, the variables constructed might be all relevant to our predicting model, but they can carry redundant information about the ties, i.e., they can be highly correlated. It is well known that correlated variables can diminish the predicting power of the model and

Table 3 Regression results for tie persistence using generalized linear models for the two prediction models and different thresholds for the number of calls w_{ij}

Feature	Model 1			Model 2		
	$w_{ij} > 5$	$w_{ij} > 10$	$w_{ij} > 20$	$w_{ij} > 5$	$w_{ij} > 10$	$w_{ij} > 20$
w_{ij}	1.491***	1.440***	1.331***	2.096***	1.793***	1.367***
	(0.021)	(0.029)	(0.045)	(0.060)	(0.072)	(0.093)
d_{ij}	0.105***	0.049**	0.036	0.109**	0.131**	0.115
	(0.012)	(0.019)	(0.032)	(0.034)	(0.046)	(0.067)
r_{ij}	–0.094	–0.085***	–0.127***	–0.044**	–0.143***	–0.142*
	(0.010)	(0.018)	(0.030)	(0.032)	(0.041)	(0.064)
o_{ij}	0.241***	0.215***	0.245***	0.286***	0.340***	0.413***
	(0.017)	(0.023)	(0.038)	(0.046)	(0.061)	(0.082)
k_{ij}	0.039**	0.035	0.031	–0.026	0.032	–0.019
	(0.010)	(0.018)	(0.029)	(0.033)	(0.045)	(0.066)
μ_{ij}^{int}	0.084***	0.109***	0.101***	0.151**	0.127**	0.147*
	(0.012)	(0.017)	(0.029)	(0.032)	(0.041)	(0.062)
age_{ij}	0.001	0.031	–0.008	–0.021	0.003	–0.036
	(0.012)	(0.018)	(0.031)	(0.034)	(0.044)	(0.064)
$gender_{ij}$	0.079***	0.121***	0.189***	0.035*	0.097*	0.239***
	(0.012)	(0.018)	(0.031)	(0.033)	(0.043)	(0.064)
$\hat{f}_{ij}$	–1.102***	–1.414***	–1.931***			
	(0.015)	(0.024)	(0.044)			
cv_{ij}	–0.362***	–0.422***	–0.539***	–0.653***	–0.724***	–0.853***
	(0.014)	(0.020)	(0.035)	(0.037)	(0.048)	(0.074))
μ_{ij}^{chats}	0.029*	–0.001	0.017	0.036	–0.036	0.046
	(0.014)	(0.019)	(0.033)	(0.038)	(0.048)	(0.069)
a_{ij}	–0.310***	–0.282***	–0.307***	–0.316***	–0.290***	–0.269***
	(0.013)	(0.019)	(0.034)	(0.036)	(0.048)	(0.071)
Constant	0.681***	0.743***	0.852***	0.779***	0.708***	0.590***
	(0.014)	(0.021)	(0.035)	(0.039)	(0.050)	(0.070)
Number of points	45,444	23,400	9744	6684	3822	1638
AUC	0.864	0.884	0.919	0.875	0.867	0.845

Performance	Model 1			Model 2		
	$w_{ij} > 5$	$w_{ij} > 10$	$w_{ij} > 20$	$w_{ij} > 5$	$w_{ij} > 10$	$w_{ij} > 20$
Accuracy	0.802	0.808	0.844	0.803	0.808	0.778
Sensitivity	0.828	0.818	0.859	0.815	0.828	0.814
Specificity	0.767	0.794	0.824	0.787	0.782	0.731

Coefficients are shown with uncertainties (standard errors) in parentheses. *Note:* *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

thus we must understand the explanatory power between them first in order to construct a statistical significant model. This process which is known as selection of variables will be addressed qualitatively in this section using the correlation matrix between them. As we can see in Figure 6 most of the variables we have selected are highly uncorrelated, with correlation coefficients below $\rho = 0.2$. As expected, we can see a moderate relationship between number of calls and topological overlap, i.e. the Granovetter effect [20, 23] ($\rho = 0.32 \pm 0.01$). Larger correlation is found for the variable t_{ij}^{min} with w_{ij} ($\rho = 0.41 \pm 0.01$) and thus we discard it in our models. We keep the rest of variables since correlation coefficients remain below $\rho = 0.4$.

5.4 Facebook data

We have also analyzed other communication data to test the independence of our results to the particular mobile phone setting. In particular, we have studied the 90,269 users of the New Orleans Network crawled during December 29th, 2008 and January 3rd, 2009 by

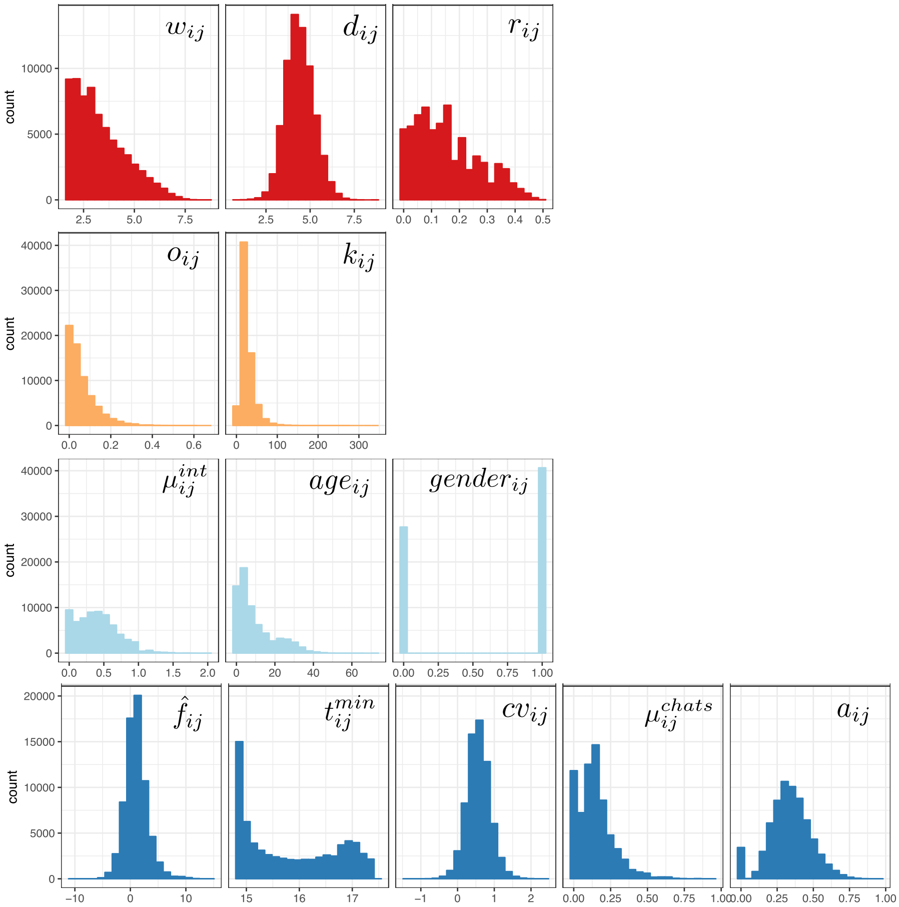
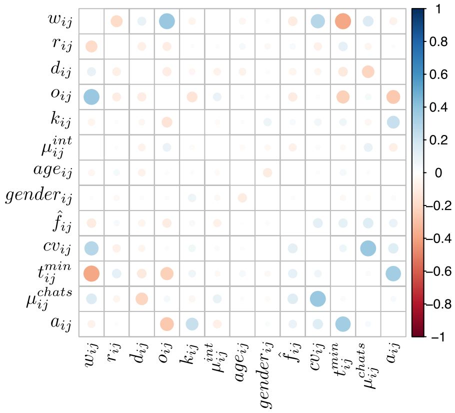


Figure 5 Distribution of features. Histograms of the different features considered in our models. Each row of histograms correspond to a different group of features: intensity, structure, intimacy and temporal features from top to bottom. Note that as mentioned in the text, some variables are log-transformed, specifically w_{ij}, d_{ij}, $\hat{f}_{ij}$, cv_{ij} and t_{ij}^{min}.

Figure 6 Correlation between features. Correlation matrix for the different tie features considered in the model. Each entry shows the Pearson correlation coefficient between two variables. Size is proportional to the absolute value of the correlation coefficient, while color shows also its sign. We only show correlation coefficients which are significantly different from zero (with a 95% confidence interval).



Vismanath *et al.* [41]. The data consists of communication events between users through Facebook wall. Contrary to the mobile phone data, the Facebook data is not steady in time, since the database extends over the early days of Facebook growth and thus it shows

Table 4 Regression results for the tie persistence using generalized linear models for the two prediction models in the Facebook dataset

Feature	**Model 1**		**Model 2**	
	Full	**Simplified**	**Full**	**Simplified**
w_{ij}	0.839***	1.228***	1.709***	2.041***
	(0.044)	(0.066)	(0.180)	(0.105)
r_{ij}	0.118***		0.470***	
	(0.034)		(0.111)	
o_{ij}	0.269***		0.384***	
	(0.035)		(0.111)	
k_{ij}	−0.008		−0.231*	
	(0.031)		(0.099)	
μ_{ij}^{int}	−0.0139		0.006	
	(0.030)		(0.095)	
$\hat{f}_{ij}$	−0.608***	−0.306***		
	(0.0390)	(0.016)		
t_{ij}^{min}	−0.329***		−0.402**	
	(0.039)		(0.128)	
cv_{ij}	−0.286***		−0.229*	−2.394***
	(0.037)		(0.103)	(0.224)
μ_{ij}^{chats}	0.024		−0.047	
	(0.035)		(0.114)	
a_{ij}	0.122***		0.115	
	(0.036)		(0.111)	
Constant	0.435***	1.951***	0.686***	−4.525***
	(0.032)	(0.155)	(0.115)	(0.265)
Number of observations	5466	5466	667	667

Performance	**Model 1**		**Model 2**	
	Full	**Simplified**	**Full**	**Simplified**
Accuracy	0.690	0.688	0.798	0.799
Sensitivity	0.770	0.780	0.814	0.802
Specificity	0.583	0.567	0.776	0.797

Coefficients are shown with uncertainties (standard errors) in parentheses. Model *Full* include all the features described in the text, while model *Simplified* only includes the most important two features. *Note:* *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

a growth in the activity over years, which translates in more wall posts and also more users as a function of time.

To minimize this effect we have chosen only communication events between users that did show any activity in the observation window Ω (the time interval between 1000 and 1212 days in the database) and also which were present 20 days before and after Ω. We do not consider the ties to be reciprocated in order to have more data accessible for our analysis. With this filter our database contains 125×10^3 communication events of $\sim 10^4$ users and 69×10^3 ties. We have considered only 5466 ties which are more active (more than 5 communication events) and build a predictive model similar to the one for the mobile phone data. However, since we do not have information about the age and gender of the users, we have discarded the variables related to their difference. Results of our model for the Facebook data are presented in Table 4 where we can see a qualitative match with the ones for the mobile dataset, although the predictive power of the models is smaller than in that case. Apart from the number of communication events, both the normalized freshness and the coefficient of variation have a similar relevant role in predicting tie persistence. In particular, we find that the critical relative freshness is now $\hat{f}_{ij} = 16.6$, which is double that the one found in the mobile phone calls. This could be a signature of the different rhythm of communication of users on different channels.

Acknowledgements
We would like to thank Telefónica for providing access to the anonymized data.

Funding
EM acknowledges funding from Ministerio de Economía y Competividad (Spain) through projects FIS2013-47532-C3-3-P and FIS2016-78904-C3-3-P.

Abbreviations
CDR, Call detail record; AUC, Area under the curve; GBM, Generalized Boosted Regression Models; RF, Random Forest; LogR, Logistic Regression.

Availability of data and materials
The CDR data used cannot be shared because is protected by an confidentiality agreement. The table of features for each link can be shared if needed. Feel free to get in contact with the corresponding author in case you need more information.

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

Authors' contributions
All authors contributed equally to this work. All authors read and approved the final manuscript.

Author details
[1]Departamento de Matemáticas & GISC, Universidad Carlos III de Madrid, Avenida de la Universidad 30, Leganés, 28911, Spain. [2]Telefónica Digital, Ronda de la Comunicación s/n, Madrid, 28050, Spain. [3]Telefónica I+D, Valladolid, 47151, Spain.

Publisher's Note

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

Received: 7 June 2017 Accepted: 7 December 2017 Published online: 16 December 2017

References

1. Saramaki J, Moro E (2015) From seconds to months: an overview of multi-scale dynamics of mobile telephone calls. Eur Phys J B 88(6):164
2. Holme P, Saramaki J (2012) Temporal networks. Phys Rep 519(3):97-125
3. Rand DG, Arbesman S, Christakis NA (2011) Dynamic social networks promote cooperation in experiments with humans. Proc Natl Acad Sci 108(48):19193-19198
4. Holme P (2016) Temporal network structures controlling disease spreading. Phys Rev E 94(2):022305
5. Miritello G (2013) Temporal patterns of communication in social networks. Springer, Berlin
6. Miritello G, Moro E, Lara R (2011) Dynamical strength of social ties in information spreading. Phys Rev E 83:045102
7. Karsai M, Kivelä M, Pan RK, Kaski K, Kertész J, Barabási A-L, Saramäki J (2011) Small but slow world: how network topology and burstiness slow down spreading. Phys Rev E 83(2):025102
8. Hidalgo CA, Rodriguez-Sickert C (2008) The dynamics of a mobile phone network. Physica A 387(12):3017-3024
9. Kossinets G, Watts DJ (2006) Empirical analysis of an evolving social network. Science 311(5757):88-90
10. Rivera MT, Soderstrom SB, Uzzi B (2010) Dynamics of dyads in social networks: assortative, relational, and proximity mechanisms. Annu Rev Sociol 36(1):91-115
11. Liben Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. J Am Soc Inf Sci Technol 58(7):1019-1031
12. Raeder T, Lizardo O, Hachen D, Chawla NV (2011) Predictors of short-term decay of cell phone contacts in a large scale communication network. Soc Netw 33(4):245-257
13. Burt RS (2000) Decay functions. Soc Netw 22(1):1-28
14. Kwak H, Moon SB, Lee W (2012) More of a receiver than a giver: why do people unfollow in Twitter? In: Proceedings of the 2012 ICWSM
15. Quercia D, Bodaghi M, Crowcroft J (2012) Loosing "friends" on Facebook. In: Proceedings of the 4th annual ACM web science conference. WebSci'12. ACM, New York, pp 251-254
16. Barabasi A-L (2005) The origin of bursts and heavy tails in human dynamics. Nature 435(7039):207-211
17. Miritello G, Lara R, Cebrian M, Moro E (2013) Limited communication capacity unveils strategies for human interaction. Sci Rep 3:1950
18. Tabourier L, Libert A-S, Lambiotte R (2016) Predicting links in ego-networks using temporal information. EPJ Data Sci 5(1):1
19. Gilbert E, Karahalios K (2009) Predicting tie strength with social media. In: Proceedings of the SIGCHI conference on human factors in computing systems. CHI'09. ACM, New York, pp 211-220
20. Granovetter MS (1973) The strength of weak ties. Am J Sociol 78(6):1360-1380. http://www.jstor.org/stable/2776392
21. Wang P, Xu B, Wu Y, Zhou X (2015) Link prediction in social networks: the state-of-the-art. Sci China Inf Sci 58(1):1-38
22. Saramäki J, Leicht EA, López E, Roberts SG, Reed-Tsochas F, Dunbar RI (2014) Persistence of social signatures in human communication. Proc Natl Acad Sci 111(3):942-947

23. Onnela J-P, Saramaki J, Hyvonen J, Szabo G, Lazer D, Kaski K, Kertesz J, Barabasi A-L (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci USA 104:7332-7336
24. Hallinan MT (1978) The process of friendship formation. Soc Netw 1(2):193-210
25. Burt RS (2002) Bridge decay. Soc Netw 24(4):333-363
26. Aledavood T, López E, Roberts SG, Reed-Tsochas F, Moro E, Dunbar RI, Saramäki J (2015) Daily rhythms in mobile telephone communication. PLoS ONE 10(9):e0138098
27. Onnela J-P, Waber BN, Pentland A, Schnorf S, Lazer D (2014) Using sociometers to quantify social interaction patterns. Sci Rep 4:5604
28. Palchykov V, Kaski K, Kertész J, Barabási A-L, Dunbar RI (2012) Sex differences in intimate relationships. Sci Rep 2:370
29. Roberts SG, Dunbar RI (2011) Communication in social networks: effects of kinship, network size, and emotional closeness. Pers Relatsh 18(3):439-452
30. Goh K-I, Barabási A-L (2008) Burstiness and memory in complex systems. Europhys Lett 81(4):48002
31. Karsai M, Kaski K, Barabási A-L, Kertész J (2012) Universal features of correlated bursty behaviour. Sci Rep 2:397
32. Aledavood T, Lehmann S, Saramaki J (2015) Digital daily cycles of individuals. Front Phys 3(118):15602
33. Llorente A (2015) Social media fingerprints of unemployment. PLoS ONE 10(5):e0128692
34. Madan A, Cebrian M, Moturu S, Farrahi K et al (2012) Sensing the "health state" of a community. IEEE Pervasive Comput 11(4):36-45
35. Botta F, del Genio CI (2017) Analysis of the communities of an urban mobile phone network. PLoS ONE 12(3):e0174198
36. Dong Y, Pinelli F, Gkoufas Y, Nabi Z, Calabrese F, Chawla NV (2015) Inferring unusual crowd events from mobile phone call detail records. In: Joint European conference on machine learning and knowledge discovery in databases, pp 474-492. Springer, Berlin
37. Ubaldi E, Vezzani A, Karsai M, Perra N, Burioni R (2017) Burstiness and tie activation strategies in time-varying social networks. Sci Rep 7:46225
38. Weng L, Karsai M, Perra N, Menczer F, Flammini A (2015) Attention on weak ties in social and communication networks. arXiv:150502399
39. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321-357
40. Friedman J, Hastie T, Tibshirani R (2001) The elements of statistical learning, vol 1. Springer, Berlin
41. Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in Facebook. In: Proceedings of the 2Nd ACM workshop on online social networks. WOSN'09. ACM, New York, pp 37-42