Persistence, Topology and Sociodemographics of a Mobile Phone Network

César A. Hidalgo[1†], C. Rodriguez-Sickert[2]

[1] *Center for Complex Network Research, Department of Physics, University of Notre Dame, Notre Dame, IN, 46556.*

[2] *Department of Sociology, Pontifícia Universidad Católica de Chile, Vicuña Mackenna 4860, Macul, Santiago, Chile.*

[†] *to whom correspondence should be addressed: chidalgo@nd.edu*

Abstract

The lack of longitudinal data has prevented the study of dynamic network properties. Here, we use cellular phone billing data to analyze the coupling between network structure and dynamics. Using 10 panels we find that the persistence of social ties, defined as the probability of observing a tie when looking at a panel, is coupled to the network's structure and sociodemographics. Standard hypothesis on the stability of ties are confirmed. The coupling between the topology and persistence is strong enough (R^2 >0.39) to accurately predict ties a year in advance.

PACS: 89.75.Hc 89.65.-s 87.23.Ge

I. INTRODUCTION

Social networks evolve as agents create or sever links with others. These dynamical properties reshape the network and ultimately determine the social outcome of collective decisions or contagion processes. The study of network dynamics has focused largely on growth, stemming from either random processes [1, 2] or game theoretical approaches[3], making growth responsible for network characteristics. However, research may focus on the dynamic properties of an observed tie. Granovetter [4] claimed that the amount of time spent, together with aspects such as emotional intensity and intimacy determine the strength of a tie. In his definition, larger commitments and, thus, the stability of a tie were inextricably linked to the concept of tie strength.

In the last decade several empirical studies have attempted to fill this gap. Morgan et. al.[5] showed that the stability of ties is better described using a *core-periphery* structure, with some ties always present, forming the core of someone's social network, and others rarely there defining its periphery. Burt [6] showed that social ties decay as a power-law meaning that relationships are subject to the *liability of newness*: Links are more prone to decay just after they have been formed. However, conserving them may prove advantageous. People good at keeping ties tend to be better evaluated at work[7]. Decay is slowed in dense social groups [6, 8] but it does not necessarily follow a smooth process. *Status transitions* such as re-entering college, marriage or divorce, promote turnover as it was observed by Suitor et. al.[9] and Wellman et. al.[10].

Despite the above studies, there is a lack of standardized methods to analyze panels of network data in all realms of network science. The example presented here belongs to social sciences, where phenomena such as opinion formation and collective decision are affected by network structure and therefore by its changes. Here we introduce a systematic way to characterize the simplest aspect of social network dynamics using billing data from a mobile phone network. We are interested in the relationship between stability, sociodemographics and local topology. To analyze stability, we introduce a formal definition of *persistence* representing the probability of observing a tie on a network panel and show that this dependent variable is heavily correlated with the topology of the network and reveals age and gender mixing patterns.

The paper is organized as follows. On the next section we introduce the data set used. Then we explain how we constructed our dependent variables and discuss their advantages and limitations. We continue by introducing the relevant theory for our study and follow with our results at the tie and node levels. We finish by showing the predictive power of the correlations found and summarize our results on the conclusion.

II. DATA, MEASUREMENTS AND THEORY

A. Data

Automatically collected records have been proposed as a source of reliable data about personal connections[11]. Email data has been used to study social processes such as tie formation[12] and social structure[13]. Nowadays social ties are also expressed through the WWW. For examples Adamic and Glance[14] studied how political opinion is distributed over the blog citation network while Holme et al. [15] described how an online dating community grow. Communication records overcome problems of survey data such as subjective biases on the respondents and the intrinsic limitations of ego-centered networks such as their unreliability measuring degree and clustering [9, 16].

Our study is based on mobile phone calls. The data consists of 7'948'890 voice calls between 1'950'426 users of a service provider holding approximately 25% of an industrialized country's market. The data consist of ten panels collected along one year (from April 15th 2004 to March 31st 2005), each containing 15 days of mobile phone calls. Calls to or from members serviced by a different provider are not part of our data. We also limited ourselves to agents that made or received at least one call in each panel to avoid dropouts or new subscribers. We hereafter assume that at high service penetration levels (in our case more than 100%), agents serviced by a particular provider are equivalent to a random sample. This data set is the one recently used by Palla et al.[17] and Onella et al.[18] to study the dynamics of social groups and the robustness of the social network respectively.

It is not our intention to claim that cellular phone communications fully capture social exchange. A social network is expressed through a host of interactions ranging from e-mails to sexual contacts, cellular phone calls are just one of the ways in which a social tie is expressed. People in close social contact tend to express ties through multiple interaction channels[19], such as email, cell-phone communications and instant messaging. However, there are arguments favoring the use of cellular phone calls as a relevant proxy for social networks. The strongest argument is based on recent experiments at MIT showing that the network extracted from cell-phone interactions correlates significantly with to the one obtained from survey data[20]. In addition, Castells et al.[21] identifies a number of underlying social process supported by the mobile communication networks, namely: the construction of a peer group through networked sociability; the emergence of collective identity; strengthening of individual identity and the formation of fashion. Recent literature has also focused on the social changes induced by these new forms of communication technologies. Social changes induced by communication technologies change the social network they connect. The world has changed from a local community to a global one in which social groups are not only based on neighborhoods or "bowling group" associations[22], but on social connections between spatially distant agents giving rise to the concept of "connected presence"[23, 24]. This recently became manifest in Spain, when citizens empowered by cellular phones organized to sanction the incumbent government in the aftermath of the Atocha Station 3/11 attacks [25]. There are also some technical aspects that favor the use of a mobile phone network as a proxy for social interactions. Mobile phone numbers are unlisted, thus knowing them reveals some sort of social connection between caller and callee. Also, cellular phones are the most widespread information technology with a penetration larger than 40% worldwide and 130% in developed countries (the number of cellular phones sold/assigned exceeds the population). During the same time period internet penetration was just over 13% worldwide and 51% for developed countries (MDGS indicators U.N. http://mdgs.un.org/unsd/mdg)[26]. Making it currently, the most complete method to study social interactions at the population scale.

B. The concept of Persistence

We formalize persistence as the number of panels in which a link appears over the total number of panels available,

$$P_{ij} = \frac{\sum_T A_{ij}(T)}{N}, \tag{1}$$

where $A_{ij}(T)$ is 1 if nodes i and j communicated on panel T and 0 otherwise. Persistence is the probability of observing a tie when observing a panel of a certain duration. If we consider panels with a duration comparable to the one of links, ($\sim$ minutes in the case of phone calls), our definition of persistence just gives the number of times a tie appeared. Whereas, when we consider panels with a duration considerably longer than the typical duration of a link, our definition of persistence captures the temporal stability of a link on a larger temporal scale. Our data set consists of 10 panels summarizing 15 days of voice call activity. Thus, it measures persistence at a monthly to yearly time scale.

We illustrate our definition of persistence using four different panels of a five node network

(Fig. 1 a). In this example, the link between nodes 2 and 4 is present in all panels while the one between nodes 1 and 2 is present only in half of them. We say that the persistence of the link between nodes 2 and 4 is 4/4 while the persistence of the link connecting nodes 1 and 2 is 2/4. Persistence is a change of representation in which we summarize all panels in a single network and represent the dynamics through links carrying its persistence value (Fig. 1 b).

Our measure of persistence weakly increases with the number of times a link is observed, persistence indicates stability, as understood in previous studies [5, 8]. However, given that we measure whether the link is observed in $N > 2$ panels, it will not describe a link dichotomously as stable or unstable, but will give the degree of stability $1/N \geq p \geq 1$. For our measure, it is possible that a link, which involves many calls on a single panel is regarded as unstable. Thus, our measure rewards those links expressed consistently in many panels.

Persistence is a tie attribute that can be defined for a particular node as the average persistence of all its ties. We denote this as *perseverence* and define it as

$$P_i = \frac{1}{k_i} \sum_j P_{ij} \tag{2}$$

where k_i is the degree, or number of connections of the i^{th} node. We will use this quantity to study what characterizes nodes carrying persistent ties.

Our definition has limitations. One could claim we are unfairly punishing newly formed links. An alternative strategy would be to consider only the links involved in the first panel, however an exercise in this line showed us that there is a strong selection bias towards stable links when we consider such an option. Our definition also does not differentiate between links active half of the time or those active during a particular half of the year. We do not propose our measure as the ultimate way to reduce a set of network panels into a weighted network, but as a simple way to do so for massive amounts of data.

C. Social Theory

Longitudinal questions can be approached from a socio-demographic perspective, which explains systematic differences in tie decay based on the type of support exchanged along the tie or demographic features such as gender, age, race or socio-economic status. Alternatively, longitudinal questions can be approached using a topological perspective, in which we study the correlations between the structure of the social network and tie persistence or decay. Here, we combine these two approaches in an attempt to understand how do social structures and characteristics coevolve.

1. Sociodemographics

Gender and Age

We can interpret our age and gender results in the context of the sociology of the family, where several publications have showed that females tend to be more involved in elderly support [27–29] and kin-keeping tasks[30]. Thus we should expect more persistent contacts between females and elders.

Suitor and Keeton[31] show that in status transitions such as re-entering college, marriage or divorce, turnover is more likely to occur. In particular, as argued by Glaeser et al.[32], assuming complementarity between investment in human capital and investment in social capital, one would expect more intensive networking in the early stages of professional careers. In parallel, the allocation of time between labor and leisure should complement the relationship between sociodemographics and persistence: less time dedicated to participate in the labor market gives more time to keep links alive. In this line, one might expect persistence to increase in females to the extent that they present lower levels of participation and, for both males and females, an increase in persistence after retirement.

Assortative Mixing and Homophyly

In the sociological literature, ties connecting actors who share specific personal attributes are referred to as homophilous ties. Actors carrying homophilous ties should be able to maintain them over time[33]. This argument riddles back to Lazarsfeld & Merton[34] who claim that homophilous network ties are more sustainable over time: "Birds of a feather flock together" . The assortative mixing hypothesis, in network theory, involves the idea of homophyly (people with similar social characteristics are more likely to create ties between them), but also consider the possibility that similarity on topological attributes could also lead to preferential attachment and drive the emergence of communities[11, 35, 36]. Based on this one should expect a link between assortative mixing and persistence.

2. Local Topology

Clustering, Topological Overlap and Embededness

One of the most prominent studies on how structural network properties influence the survival of organizations focuses on relational embeddedness [37][38]. Embeddedness can be seen as an exchange system with unique opportunities relative to markets. As the studies show, firms that are highly embedded in networks of relationships have higher chances of survival than do firms with only arm's-length market relationships. At the triadic level, based on arguments of balance theory, the social influence among triadic friendship ties might also be a factor, that influences tie sustainability. Holland & Leinhardt [39] showed in their study that if a chooses b as a friend and b chooses c as a friend, then a will tend to choose c as a friend as well. As argued by Hrushka & Henrich [40], an underlying process of cooperation might explain this tendency: the success of conditionally cooperative strategies relies on their cliquishness –a propensity to defect with strangers if they already have an adequate number of partners. This process can be reinforced to the extent that c can offer an exogenous instance (e.g., c's birthday) for a and b to cultivate their friendship. Thus, high levels of topological overlap should be associated with higher persistence. One could also expect that the risk of splitting connections (specially when a tie is broken as the result of a conflict between the actors) also deters tie decay.

Burt[7], on the other hand, shows that bridges, defined as the links that upon removal increased the number of components of the network, are more prone to decay. It can be argued that in contrast with links with high levels of topological overlap, bridges lack exogenous instances to cultivate the relationship making them more fragile.

Agents' Degree and Core-Periphery Structure

Cummings & Higgins[41] showed that people who are at the core of a network and are therefore more embedded in the overall network structure, are more likely to keep their ties stable compared to those at the periphery of a network. This core-periphery structure of the overall network adds to a higher closeness of the individual actor and with that more opportunities to interact.

If assuming a fixed amount of time to the maintenance of links, agents would face a trade-off between the number of links and the investment in maintaining them (i.e., a negative relationship should be expected between degree and average persistence. However, whether higher degree actors will have larger cores or not will depend on the magnitude of two opposite effects: a lower degree agent can concentrate on the investment process associated with maintaining the link, but higher degree agents have more opportunities to choose among mutually beneficial relationships.

Bidirectional ties and Reciprocity

Preexisting friendship ties, which indicate a positive content of the tie, have shown that people are more likely to stay connected. This also includes reciprocated contacts that are revisited because two actors feel emotionally close[42]. In the context of social exchange theory, which assumes that people tend to keep a balanced exchange of support, Ikkink & van Tilburg[43]used the social support networks of 2,057 older adults to show that surviving ties tend to be the ones in which more support was exchanged. The norm of reciprocity[44] states that people prefer a relationship where they receive and give a more or less equal amount of support[45], helping to avoid feelings of exploitation or indebtedness. Discrepancies between giving and receiving support can threaten a relationship[46] because individuals make an effort to maintain them partly due to what they get from them[47].

This process has been emphasized in the recent analytical literature on social capital. According to Putnam[22], networks of organized reciprocity give the grounds for social efficiency and social cohesion in the modern world. Thus, one might expect that persistence and reciprocity co-evolve in time, and together represent an expression of long-term mutually cooperative relationships[48].

III. RESULTS

A. Global Analysis of the Persistence of Ties

Figure 2 a shows the persistence histogram for the voice call network indicating that it has a core periphery structure similar to that found by Morgan *et. al.*. This is clear from the bimodal distribution in which ties tend to be either persistent or not. Persistent ties define the *core* of an agents' social network ($P \geq 5/10$) while non-persistent ties define the *periphery* ($P < 5/10$). The decay of ties as a function of time is well approximated by a power-function (Fig. 2 b), in agreement with the 4 year study performed by Burt[6].

We find that the survival probability of a tie follows $\sim t^{-\alpha}$ with $\alpha = 0.25 \pm 0.07$ which is not far from the value found by Burt[7] for bankers kin relationships ($\alpha \sim 0.47$). This heavy tailed distribution evidences a core-periphery hypothesis and supports the idea that ties are subject to the liability of newness. The exponent found shows that social ties are very reluctant to disappear compared to a random process in which their decay would be exponential.

B. Local Topology, Sociodemographics and the Persistence of Ties: Bivariate Analysis

Figure 3 a shows a fragment of the mobile call network extracted by considering all connections up to 3 links away of a randomly chosen agent. Although this example shows less than the 0.0008% of our network, it visually summarizes the correlations between persistence, perseverance and the topological attributes of the mobile call network. In particular, we find that these temporal attributes correlate with topological variables such as degree k_i, average reciprocity r (fraction of ties containing both, incoming and outgoing calls) and clustering coefficient C_i defined as:

$$C_i = \frac{2\triangle}{k_i(k_i-1)}. \tag{3}$$

where $\triangle$ is the number of triads in which the node is involved. Figure 3 b shows a histogram of persistence split into 9 different degree categories revealing that persistent links represent a large fraction of the connections for low degree nodes while transient links are more common for large degree nodes. However, the number of persistent ties grows as a function of degree, meaning that although in average, the persistence of high degree nodes is lower, in absolute terms their core is larger.

Figure 3 d shows the distribution of persistence divided by clustering coefficient categories, indicating that highly clustered nodes tend to have relatively large cores. In the core periphery context, this means that persevering nodes are located in dense parts of the social network (Fig. 3 a I) while those in sparser parts tend to have non-persistent ties acting as bridges which interruptedly connect different parts of the network (Fig. 3 a II). Finally, we split the distribution of persistence by reciprocity (figure 3 e) and observe that nodes with more reciprocated ties tend to be more persistent. Thus reciprocal ties connect to members from an agent's social core.

Sociodemographic quantities also relate to persistence. Figure 4 a shows the distribution of persistence for males and females, revealing that females appear to be slightly, but significantly more persistent (p-value ANOVA :$< 10^{-30}$). For age we look at the mixing matrix defined by

$$F(a,b) = \frac{< P(a,b) > - < P >}{< P >} \tag{4}$$

were $< P(a,b) >$ is the average persistence of links connecting nodes of ages a and b and $< P >$ is the average persistence of the network. Persistent links tend to connect agents with ones from their same age group and with those one generation apart. In the sociology of the family several publications have shown that females are more involved in supporting family elders ([27–29] and performing kin-keeping tasks for the family[30], thus we should expect more persistent contacts between females and elders. In order to investigate the kin-keeping hypothesis, we constructed the age and gender mixing matrices, finding that females tend to have a larger number of stable connections with people from a different generation than males (Fig. 4 c).

C. Local Topology, Sociodemographics and the Persistence of Ties: Multivariate Analysis

In the previous section we presented a bivariate analysis in which we analyzed the effect on tie persistence for a host of Independent Variables (IVs) and found that persistence depends monotonically on all IVs (degree, clustering coefficient, reciprocity, gender), except for Age, which is characterized by a tri-diagonal mixing pattern. In this section we conduct a multivariate analysis to explore the impact of each variable on persistence. Given the number of agents considered, any correlation of the order of 10^{-6} or larger is significant at the 99% confidence level. Thus we concentrate on the magnitude of their contributions to the variance and on the direction/magnitude of changes associated with each IV, rather than their significance.

We analyzed how persistence depended on tie attributes. As we introduced in the theory section, we would like to see if assortative mixing explains the variance in persistence. For this purpose, we constructed variables equal to the absolute difference between node attributes, to see if persistent links are more likely to connect nodes with similar characteristics. We also performed the same analysis using the products between the attributes found at both ends of a link and found a slightly smaller $R^2 = 0.397\ v/s\ R^2 = 0.392$ and the same structure in the correlation matrix, thus we continue our discussion based on the differences between them. Table 1 shows the correlation matrix between persistence and our assortative mixing IVs defined as the difference in clustering ΔC, degree Δk, reciprocity Δr, gender (same gender = 0, different gender = 1) and age ΔAge. We confirm the homophyly hypothesis for the available sociodemographic variables: age and gender; and remind that most crucial variables such as race, education, income, ideology and values are not part of our data. We believe that such variables are the most important ones when testing a homophily hypothesis. In contrast, assortative mixing does not have the expected effect and persistence is slightly larger in links connecting nodes with different topological characteristics. We also studied the reciprocity of the links (0 non-reciprocal, 1 reciprocal) and the topological overlap between nodes connecting links defined as:

$$T.O.(i,j) = \frac{(\text{Number of Common Neighbors})^2}{k_i \cdot k_j}. \quad (5)$$

as possible predictors of persistence. Although individually the difference in degree, clustering and age appears to explain a small fraction of the variance, when we control for other IVs we see that the reciprocity of ties accounts for most of the variance in persistence followed by topological overlap. However, it is important to stress that whereas reciprocity explains most of the variance, the magnitude of the change in persistence associated with topological overlap is greater than the one associated with reciprocity. Thus a link with someone sharing most neighbors is a better predictor of persistence than the reciprocal nature of the link. However, this link will likely be reciprocated adding to the explanatory power of reciprocity.

In the studied network we do observe a tendency for links to connect people from the same age, degree and different clustering. However, our observations show that the observed homophyly does not explain a significant fraction of the variance in the persistence of a tie. It is very likely that these patterns come from people with similar characteristics sharing work or social environments and that although at this level they appear to be similar, other underlying factors drive the persistence of their interactions. Reciprocity appears to be a good indicator for the existence of these other factors, therefore it is able to explain more persistence than homophyly.

Table 2 shows the correlations between perseverance P_i and all the studied IVs. We controlled for redundancy using a linear, log-linear and double logarithmic models, but kept the linear one because of its marginally higher R^2 (0.4943). When we look at the partial correlation coefficients we find that most correlations vanish and the biggest contribution to perseverance is given by the average reciprocity (r), followed by degree (k). The clustering coefficient (C), which appeared as the strongest predictor in the bivariate case, explains just 6% of the variance. This is because they are formed by reciprocal ties and large degree nodes tend to have fewer of them.

D. Predictability

How well can we predict the stability of ties starting from a single panel? As mentioned before, persistence is a time-like, vertical variable and is not constrained to correlate with space-like, horizontal variables. Space-like variables can be naturally constrained[49, 50] and thus it is important to take into account their correlations to unveil their real contribution to the persistence of ties. Can we use this information to predict which ties persist in time? To answer this question we looked at our first data panel and use different criteria to predict which ties will be stable. We then looked at the fraction of these ties appearing after 1, 3, 6, 9 and 12 months and gauged the accuracy of our predictions by measuring their Positive Predictive Value (P.P.V.) defined as:

$$\text{P.P.V.} = \frac{TP}{TP + FP}, \tag{6}$$

where TP is the number of true positives and FP is the number of false positives. When the prediction is based only on the reciprocity of ties observed in the first panel, the P.P.V. ranges from 70% after one month to 43% after a year (Fig 5 a) compared to picking a random set of existing ties which gives a P.P.V. of 35% after a month and 20% after a year.

We can improve our predictive power by using a more stringent criterion. If we consider all reciprocal links that also have a topological overlap larger than $T.O. \geq 0.01$ we improve the P.P.V. by 5% while an even more stringent criterion based on a $T.O \geq 0.1$, gives us an extra percent that allows us to predict with a P.P.V. larger than 50% after one year. This increase in accuracy, however, comes with a cost, reducing the number of links predicted to be persistent. Thus the sensitivity defined as:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{7}$$

where FN is the number of false negatives, decreases with the stringency of the criteria but increases with time (Fig. 5 b).

Reciprocity appears to be the best predictor of persistence, however, it is not the only one. The fact that the variance explained by other IVs was redundant with that explained by reciprocity allows us to use these IV's as alternative predictor of ties. Figure 5 a shows the P.P.V. obtained when we use only the topological overlap as our predictive criteria. In this case we see that although the accuracy is lower, it is still significantly better than random, for links with $T.O. > 0.1$ and $T.O. > 0.01$. and has a similar trade off with sensitivity (Fig. 5 b). Thus the redundancy observed in the system can be turned into a predictive advantage and in the absence of information about the reciprocity of links we can use redundant measures to make good educated guesses about the existence of future ties.

IV. DISCUSSION

We have defined and measured the persistence of ties in a one year period using 10 panels of data summarizing the activity of all voice calls carried by a mobile phone carrier from an industrialized country. We showed that the persistence of ties and perseverance of nodes depend on topological (degree, clustering, reciprocity and topological overlap) and sociodemographic variables (age and gender). In our study, topological variables explain most of the variance, so the temporal stability of an agent's tie is better explained by its social network's structure than by its age or gender. At the node level perseverance is a behavioral attribute, thus it is not surprising that the local structure of the social network, that it is likely also a result of social behavior, predicts the persistence of ties.

Detailed measures of the social structure at the population level have been hard to obtain in the past. However, the increasing availability of data like the one used in this study will allow testing the effects of structural, sociodemographic and cultural attributes on the behavior of individuals. We have pointed out how several results presented here agree with previous results found using more traditional methods. A secondary objective of this work was to show that quantitative sociological measurements can be performed at the population level using mobile phone data, which appears to be an adequate proxy due to high penetration and the personal nature of the uncovered relations. The results obtained with these methods answer questions of a different nature than the ones obtained using survey data. We believe, however, that both methods can work complementarily in favor of a more thorough understanding of our evolving society.

The relationships shown here demonstrate that the temporal dynamics of social interactions are intrinsically coupled to social network structure in such a way that the existence of a tie can be predicted, with a respectable accuracy, using a simple criterion. The dynamic properties of static networks used in other studies can be approximated using the results presented here, encouraging the use of structural measures to explain human behavior.

V. ACKNOWLEDGMENTS

C.A. Hidalgo was supported by the Kellogg Institute at Notre Dame and acknowledges support from NSF grant ITR DMR-0426737, IIS-0513650 and the James Mc Donald Foundation 220020084. C. Rodríguez-Sickert acknowledges Sam Bowles and the S.F.I. Special acknowledgments to A.-L. Barabási for providing the source data and discussing earlier versions of the manuscript.

[1] Newman, M., A.-L. Barabási and D. Watts (2006) The Structure and Dynamics of Networks. Princeton University Press.

[2] Barabási A.-L. and R. Albert (1999) Science 286: 509-12.

[3] Jackson M.O. and A. Wolinsky (1996) Journal of Economic Theory 71(1):4474.

[4] Granovetter M.S. (1973) The American Journal of Sociology 78(6): 1360-80

[5] Morgan David L., M.B. Neal and P. Carder (1996) Social Networks 19:9-25

[6] Burt Ronald S. (2000) Social Networks 22:1-28.

[7] Burt Ronald S. (2002) Social Networks 24:333-363

[8] Feld Scott L. (1997) Social Networks 19:91-95.

[9] Suitor J.J, S. Keeton (1997) Social Networks 19:51-62

[10] Wellman Barry, R.Y. Wong, D. Tindall and N. Nazer (1997) Social Networks 19:27-50

[11] Newman M.E.J. (2003) Phys. Rev. E. 67:026126 .

[12] Kossinets G. and D.J. Watts (2006) Science 311:88-90

[13] Ebel H. L.-I. Mielsch and S. Bornholdt (2002) Phys. Rev. E 66:035103

[14] Adamic, L, and N. Glance (2005) Proceedings of the 3rd international workshop on Link discovery 36-43.

[15] Holme P. C.R. Edling and F. Liljeros (2004) Social Networks 26:155-174

[16] Marsden (1990) Anual Review of Sociology 16: 435-63.

[17] Gergely Palla, Albert-László Barabási and Tamás Vicsek Nature 446, 664-667 (2007)

[18] Onnela J.-P., J. Saramki, J. Hyvnen, G. Szab, D. Lazer, K. Kaski, J. Kertsz, A.-L. Barabsi (2007), P.N.A.S. 104(18) 7332

[19] Haythornthwaite C (2005) Information, Communication, & Society 8:125-147.

[20] Eagle N, A. Pentland, and D. Lazer (2007), Inferring Social Network Structure using Mobile Phone Data, PNAS, (in submission)

[21] Castells, M., M.Fernandez-Ardevol, J.Linchuan and A. Sey (2007) The Mobile Communication Society, MIT Press.

[22] Putnam Robert D. (1995) Journal of Democracy 6:65-78

[23] Licoppe Christian and Z. Smoreda (2005) Social Networks, 27:317-335.

[24] Wellman, Barry and Bernie Hogan with Kristen Berg, Jeffrey Boase, Juan-Antonio Carrasco, Rochelle Ct, Jennifer Kayahara, Tracy L. M. Kennedy and Phuoc Tran. (2006) Connected Lives: The Project. In Purcell, P. ed The Networked Neighbourhood. (Springer, London). Pp. 157-211.

[25] Suárez, S. (2006) "Mobile Democracy: Text Messages, Voter Turnout and the 2004 Spanish Elections", Representation (forthcoming).

[26] MDGS indicators, United Nations. http://mdgs.un.org/unsd/mdg/

[27] Spitze G. & Logan J. (1990) Journal of Marriage and the Family 52:420-430

[28] Townsend P. (1968) Old people in three industrial societies. Atherton Press; New York.

[29] Brody, E. (1981) Gerontologist 21:471-479

[30] Rosenthal C. (1985) Journal of Marriage and the Family 47:965-974

[31] Suitor J.J, B. Wellman and D.L. Morgan (1997) Social Networks 19:1-7

[32] Glaeser, E., D. Laibson, J. Scheinkman and C. Soutter (2000) Quarterly Journal of Economics, 115 (3): 811-46.

[33] McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001) Annual Review of Sociology 27(1): 415-439.

[34] Lazarsfeld, P. F. & Merton, R. K. (1954) in M. Berger (Ed.), Freedom and Control in Modern Society: 18-66. New York: Van Nostrand.

[35] Newman M.E.J. (2002) Phys. Rev. Lett. 89(20) 208701

[36] Newman M.E.J. and Juyong Park (2002) Phys. Rev. E. 68:036122

[37] Uzzi, B. (1999) American Sociological Review, 64(4): 481-505.

[38] Granovetter, M.S. (1985) American Journal of Sociology 91:481-510

[39] Holland, P. W. & S. Leinhardt (1970) American Journal of Sociology, 76(3): 492-513.

[40] Hrushka, D. & J. Henrich (2006). Journal of Theoretical Biology 239: 1–15.

[41] Cummings Jonathon N., M.C. Higgins (2006) Social Networks 28:38-55

[42] Fischer, C. (1982). To dwell among friends. Chicago: University of Chicago Press.

[43] Ikkink K.K. and T. van Tilburg (1999) Social Networks 21:131–146

[44] Gouldner A.W. (1960) American Sociological Review 25:161–179

[45] Antonucci T.C. and J.S. Jackson (1990) Social Support: An Interactional View. Wiley, New York, 173–198

[46] Neufeld A. and M.J. Harrison (1995) Qualitative Health Research 5:348–365

[47] Rusbult C.E. and A.P. Buunk (1993) Journal of Social and Personal Relationships 10:175–204

[48] Zimmermann, MG et al (2000) "Cooperation in an Adaptive Network" in Application of Simulations to Social Sciences. Eds. G. Ballot and G. Weisbuch, Hermes Science Publications , 283-297

[49] Newman. M.E.J. and J. Park (2004) Phys. Rev. E 70:066117

[50] Vázquez A., R. Dobrin, D. Sergi, J.-P. Eckmann, Z. N. Oltvai, and A.-L. Barabási. (2004) PNAS 101:17940-17945

VI. TABLES

PCC	ΔC	Δk	Δr	$\Delta Gender$	ΔAge	R	T.O.	P
ΔC	1	0.023	0.15	-0.005	0.019	0.11	0.23	0.15
Δk		1	0.020	0.0092	0.059	-0.13	-0.19	-0.16
Δr			1	-0.001	0.0046	-0.068	-0.073	0.033
$\Delta Gender$				1	0.10	-0.056	-0.046	-0.081
ΔAge					1	-0.1025	-0.0501	-0.1206
R						1	0.2964	0.5886
T.O.							1	0.3537
Linear Regression Coefficients	0.09	0.002	0.15	-0.03	-0.002	0.35	0.56	
p^2_{ri}	0.0027	0.0032	0.0070	0.0018	0.0035	0.26	0.034	P

TABLE I: Pearson correlation coefficients between the seven tie IVs and Persistence. The bottom rows shows the partial correlation between these and persistence and the regression coefficients respectively.

PCC	C	k	r	Age	Gender	π
C	1	-0.51	0.49	-0.02	0.16	0.64
k		1	-0.34	0.04	-0.15	-0.45
r			1	-0.02	0.03	0.62
Age				1	-0.01	-0.04
Gender					1	0.14
Linear Regression Coefficients	0.0598	-0.0122	0.3626	0.0015	0.009	
Partial Correlation (p^2_{ri})	0.062	0.11	0.27	-0.00014	0.0056	π

TABLE II: Pearson correlation coefficients between the five node IVs and Persistence. The bottom rows show the partial correlation coefficients between and the linear regression coefficients respectively.

VII. FIGURES

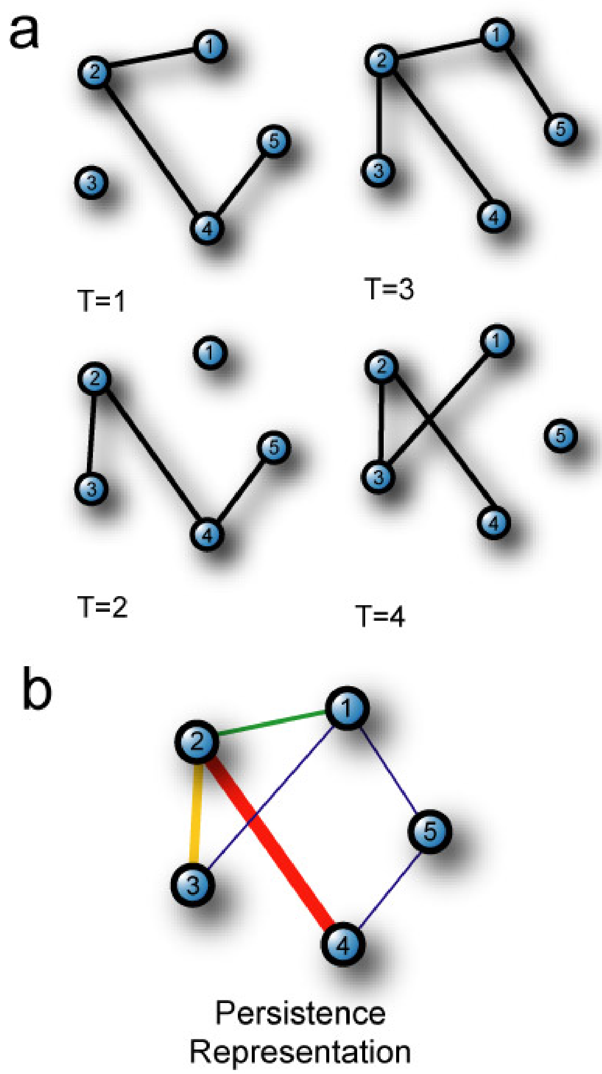

FIG. 1: Figure 1: Definition of Persistence. a. Four panels of a five node network in which not all links are equally persistent. b. Persistence representation of the four panels presented in a.

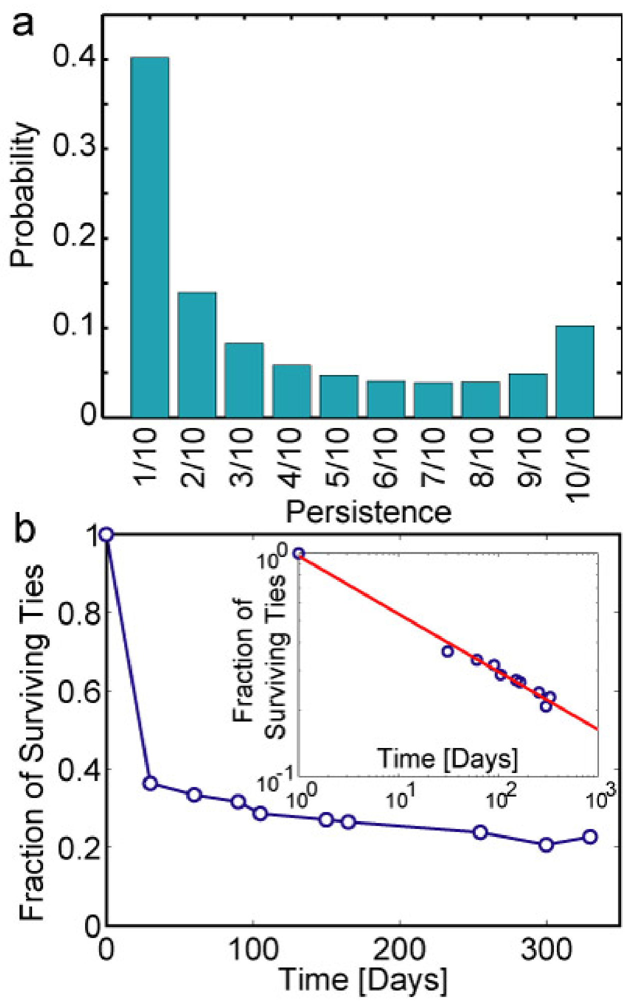

FIG. 2: Figure 2: Basic characterization of the persistence of a cellular phone network. a. Distribution of persistence for all links. b. Fraction of surviving ties as a function of time. The inset shows the same plot in a double logarithmic scale. The continuous line is $t^{-1/4}$.

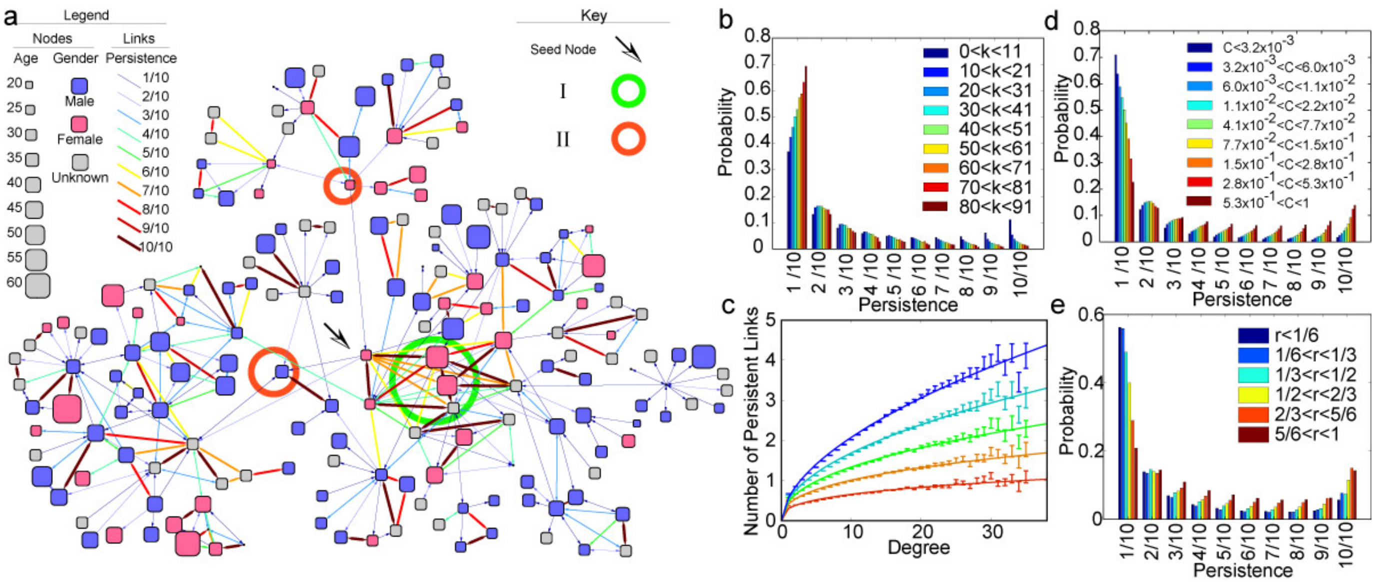

FIG. 3: Figure 3: Topological factors influencing the persistence of ties at the node level. a. A fragment of the network extracted by considering all links in the first three neighborhoods of a randomly chosen node (indicated by the large black arrow). b. Distribution of persistence divided into nine clustering categories. c. Number of persistent links defined as those with a persistence of, from top to bottom: 5/10, 6/10, 7/10, 8/10, 9/10 and 10/10. d. Distribution of persistence divided into nine clustering categories. e. Distribution of persistence divided into five different reciprocity segments.

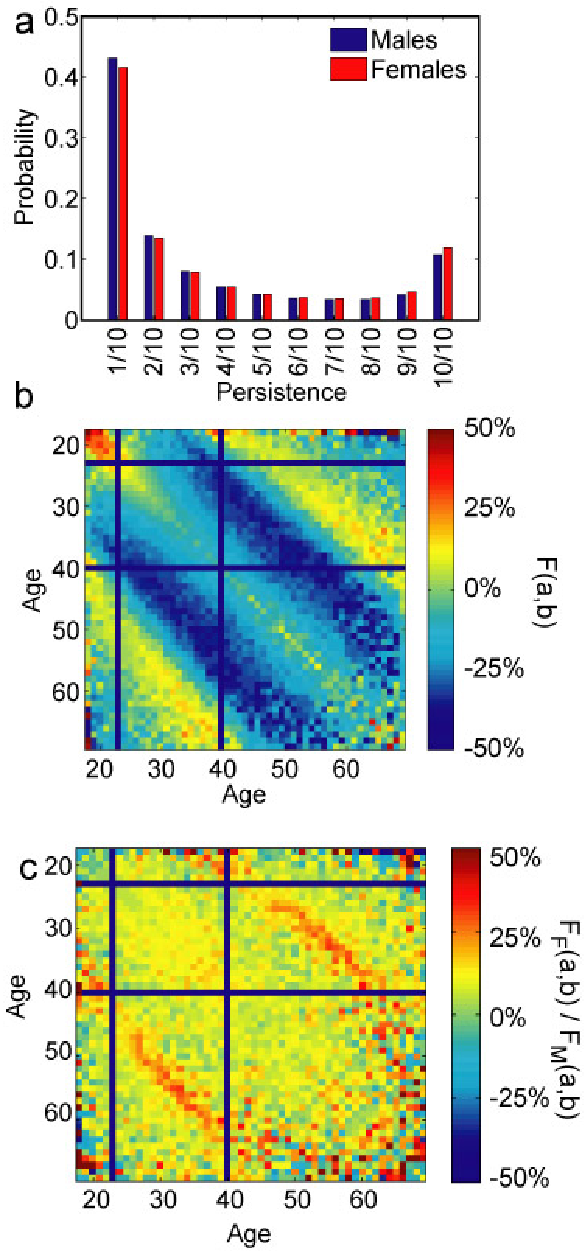

FIG. 4: Figure 4: Socio-demographical factors influencing the persistence of ties. a. Distribution of persistence for males and females. b. Increase of the persistence of links connecting nodes of given age relative to the average persistence of the network ($F(a, b)$). c. Relative increase in the persistence age mixing matrix of females ($F_F(a, b)$) relative to the one of males ($F_M(a, b)$).

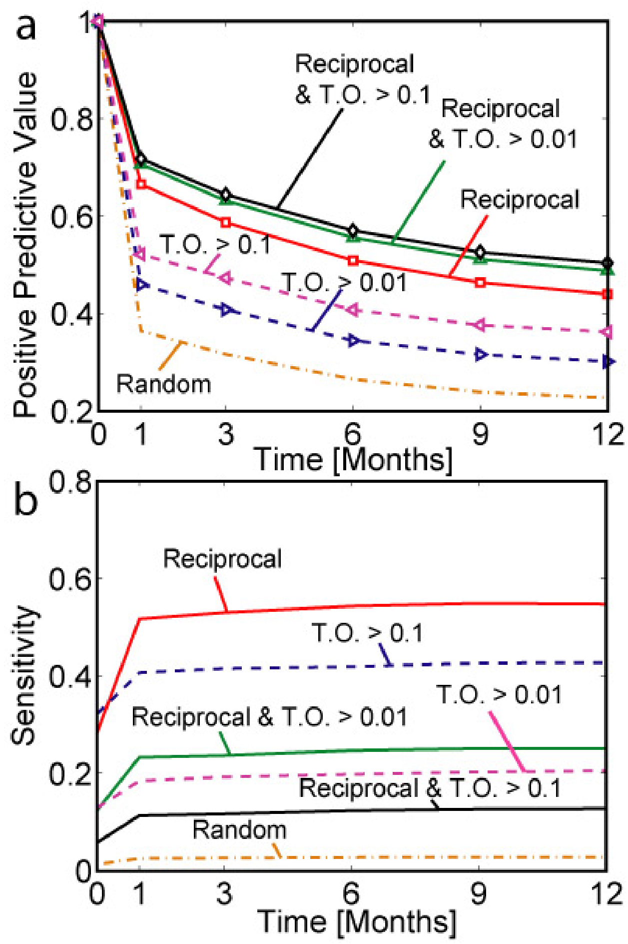

FIG. 5: Figure 5: Predictability of the temporal stability of ties. a. Fraction of ties accurately predicted to exist based on our first panel, by randomly choosing ties (orange), choosing reciprocal ties (red), reciprocal ties with a $T.O. > 0.01$ (green), reciprocal ties with a $T.O. > 0.01$, ties with a $T.O. > 0.01$ (blue) and a $T.O > 0.1$ (purple). b. Sensitivity of the predictive methods presented in a. Same color scheme.