Structure and tie strengths in a mobile communication network

Jari Saramäki JARI.SARAMAKI@TKK.FI
Dept. of Biomedical Engineering and Computational Science, Helsinki University of Technology, P.O. Box 9203, FI-02015 TKK, Finland

Jukka-Pekka Onnela JP.ONNELA@PHYSICS.OX.AC.UK
Physics Department, Clarendon Laboratory, Oxford University, Oxford OX1 3PU, United Kingdom

Keywords: complex network, social network, network analysis, call record

Abstract

We examine the communication patterns of millions of anonymized mobile phone users. Based on call records, we construct a communication network where vertices are subscribers and edge weights are defined as aggregated duration of calls, reflecting the strengths of social ties between callers. We observe a coupling between tie strengths and network topology: at the "local" level, strong ties are associated with densely connected network neighbourhoods, providing the first large-scale confirmation of the Granovetter hypothesis. Based on fragmentation analysis, weak ties are seen to play an important role at the network level, accounting for global connectivity. The observed coupling is shown to significantly slow down the spreading of random information, resulting in dynamic trapping of information in communities.

1. Introduction

Uncovering the structure of large-scale social networks has traditionally been constrained by lack of data. In the questionnaire-based approach, the number of individuals is limited by practical reasons, and the nature and strength of social ties are based on subjective recollection. However, modern electronic databases provide an unprecedented opportunity to map out interactions among a large number of individuals. It is evident that in this approach there is a tradeoff: the number of individuals can be very large, but the scope of social interactions to be investigated is more narrow, such as communication via email (Eckmann et al., 2004; Dodds et al., 2003), instant messaging (Leskovec & Horvitz, 2008), and telephone (Aiello et al., 2000). Nevertheless, these interactions can be viewed as a proxy for the underlying social network. Here, we present results on a large communication network derived from call records, published earlier in (Onnela et al., 2007b; Onnela et al., 2007a). We focus on the coupling between network topology and tie strengths, defined as aggregated call minutes. Although our approach follows the tradition of complex network science, which is largely based on statistical physics (Boccaletti et al., 2006), the research questions we investigate come from network sociology: how do tie strengths correlate with the surrounding network structure? What roles do weak and strong ties play in structure and function of social networks?

1.1. Source data

Our studies are based on a call record database containing all mobile telephone calls between 7 million people during a period of 18 weeks. We focus on private subscriptions only. The data originates from an operator having a market share of 20 % in an undisclosed European country. To translate the log data into a network representation, we connect two users with a link if there had been at least one reciprocated pair of phone calls between them; the reciprocity requirement eliminates single events such as wrong numbers, telemarketing from a private subscription etc. The resulting mobile call graph (MGC) has $N = 4.6 \times 10^6$ vertices and $L = 7.0 \times 10^6$ edges; 84 % of vertices belong to a single connected cluster. We define the tie strength w_{ij}, i.e. edge weight, as the aggregated duration of calls between two users

Preliminary work. Under review by the International Workshop on Mining and Learning with Graphs (MLG). Do not distribute.

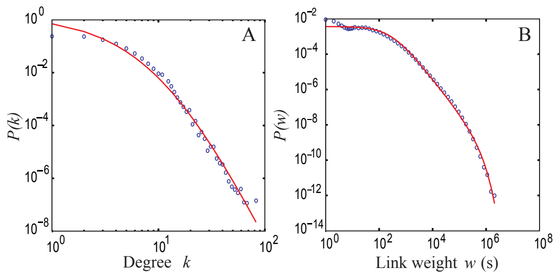

Figure 1. Vertex degree distribution (A) and edge weight distribution (B) for the data. The solid curves denote fitted functions of the form $P(x) = a\,(x + x_0)^{-\gamma_x} \exp(-x/x_c)$, where $\gamma_k = 8.4$ and $\gamma_w = 1.9$.

i and j. Statistics for some quantities are shown in Table 1 below. The probability density distributions for both vertex degrees and edge weights are broad (Fig. 1) and can be relatively well fitted with $P(x) = a\,(x + x_0)^{-\gamma_x} \exp(-x/x_c)$, i.e. a power law with an exponential cutoff. However, the degree distribution exponent is relatively large, $\gamma_k = 8.4$. This means that although the distribution can be approximated as scale-free, the highest observed degrees are relatively limited ($O(10^2)$), and thus the role of hubs is less pronounced.

Table 1. Statistics for key quantities in the MGC. Degree = number of edges of a vertex, edge weight = total call duration along edge, call mins per user = total call duration along all edges of a vertex.

Quantity	Mean	Std dev
degree	3.3	2.5
edge weight	41 min	206 min
call mins per user	135 min	386 min

2. Local coupling between tie strength and topology

The "strength of weak ties" hypothesis (Granovetter, 1973) states that the strength of a tie between i and j increases with the overlap of their friendship circles; for weak ties this means that they typically act as connectors between different communities or circles of friendship. In order to quantitatively study this in our data set, we define the relative topological overlap of the neighbourhood of two users i and j as

$$O_{ij} = n_{ij} / \left((k_i - 1) + (k_j - 1) - n_{ij}\right), \qquad (1)$$

where n_{ij} is the number of common network neighbours of i and j, and k_i (k_j) denotes the degree, i.e. the number of edges, of vertex i (vertex j). If i and j have no common acquaintances, then $O_{ij} = 0$; if all their acquaintances are common to both, $O_{ij} = 1$. Fig. 2 shows the average overlap $\langle O \rangle$ as a function of (cumulative) edge weight, i.e. the percentage of links with weights smaller than w. It is clearly seen that for 95% of the edges, the stronger the tie between two subscribers, the more their network neighbourhoods overlap. This result is consistent with the weak ties hypothesis and its first verification in large, societal-level networks; tie strength is at least in part driven by the surrounding network topology. Analysis of weight-related properties (*intensities*) of fully connected subgraphs (cliques) in the network confirms the above findings: clique intensities are higher than in reference networks where the weights have been shuffled to remove correlations. See (Onnela et al., 2007a) for more details.

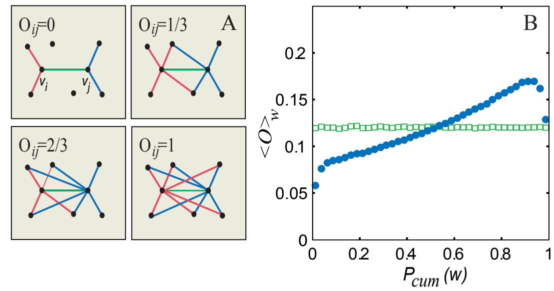

Figure 2. (A) Illustration of the overlap between two vertices i and j. (B) Average overlap as a function of cumulative tie strength (solid blue circles), showing that increasing tie strength implies increasingly overlapping circles of friendship. Open circles display a reference case where all tie strengths have been randomly shuffled, and no strength dependence is seen.

3. Global role of tie strengths

In order to explore the global, network-wide implications of the above-mentioned local relationship, we investigate the network's robustness with respect to the removal of either weak or strong ties. Here, the control parameter f is the fraction of removed links, and the order parameter $R_{gc}(f)$ measures the relative size of the giant component, i.e. the largest connected cluster of vertices, after link removal. Fig. 3 (A) illustrates the result: removing ties in rank order from weakest to strongest leads to the network's disintegration at $f_w \approx 0.8$, whereas removing the strong ties first shrinks the network but will not precipitously break it apart. The precise point of disintegration can be determined by monitoring the *suscep-*

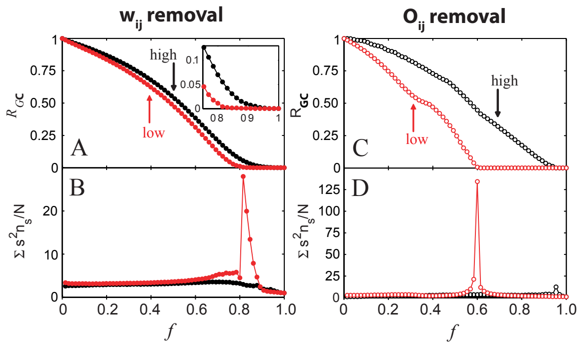

Figure 3. Robustness of the communication network with respect to edge removal. (A, B) Edges are removed on the basis of their weights. The black curves correspond to removing strong edges first, while the red curves denote the opposite, starting from weakest edges. (A) shows the relative size of the largest component as a function of fraction of removed edges; removing weak links first fragments the network at around $f = 0.8$ (see also inset). This is corroborated by the susceptibility peak in panel (B). (C,D) For reference, edges are removed in order of their overlap O_{ij} and the panels again display the relative largest component size and susceptibility. Removing low-overlap edges first rapidly breaks the network apart, reflecting the weight-overlap dependence shown in Fig. 2.

tibility $S = \sum_{s<smax} n_s s^2/N$, where n_s is the number of clusters containing s vertices (Stauffer & Aharony, 1994). When starting with the lowest-weight links, S develops a peak around 0.8, whereas no peak is seen for the reverse order of tie removal (panel B). Hence, one can conclude that weak links are of major importance to the overall connectivity of the network, whereas strong links participate in dense clusters which are interconnected by weaker links. For reference, we also show the response to removing links in ascending/descending order of overlap O (panels C and D). It is clearly seen that removing low-overlap edges rapidly destroys the connectivity at $f \approx 0.6$, whereas removing high-overlap edges only appears to shrink the networks. This is as expected, given that the network contains communities of densely connected vertices: low-overlap-edges then act as bridges between the communities.

4. Discussion

Although the study of social networks has a long history, investigating the relationship between network topology and tie strengths has generally been impossible until recently. Using mobile telephony call records, we have shown that tie strengths correlate with the topology in a meaningful, non-trivial way, confirming the weak ties hypothesis. On the global scale, these correlations give rise to different roles for strong and weak ties: weak ties are the "glue" holding the network together, whereas strong ties are associated with dense network neighbourhoods.

If one assumes that the main function of social networks is transmission of information, the observed correlations play a crucial role. Indeed, simulations of a simple information spreading process indicate (Onnela et al., 2007b) that contrary to expectations, the strength-topology correlations and the network topology limit random spreading as information tends to become "trapped" in high-edge-weight communities, and the weak links between these constrain the flow.

References

Aiello, W., Chung, F., & Lu, L. (2000). A random graph model for massive graphs. *Proceedings of the 32nd ACM Symposium on the Theory of Computing* (pp. 171–180).

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424*, 175–308.

Dodds, P., Muhanad, R., & Watts, D. (2003). An experimental study of search in global social networks. *Science*, *301*, 827–829.

Eckmann, J.-P., Moses, E., & Sergi, D. (2004). Entropy of dialogues creates coherent structures in e-mail traffic. *Proc. Natl. Acad. Sci. (USA)*, *101*, 14333–14337.

Granovetter, M. (1973). The strength of weak ties. *Am. J. Sociol.*, *78*, 1360–1380.

Leskovec, J., & Horvitz, E. (2008). Planetary-scale views on a large instant-messaging network. *Proceedings of WWW 2008, Beijing, China.*

Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., de Menezes, M. A., Kaski, K., Barabási, A.-L., & Kertész, J. (2007a). Analysis of a large-scale weighted network of one-to-one human communication. *New J. Phys.*, *9*, 179.

Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., & Barabási, A.-L. (2007b). Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. (USA)*, *104*, 7332.

Stauffer, D., & Aharony, A. (1994). *Introduction to percolation theory, 2nd ed.* Taylor and Francis, London.