Scaling behaviors in the communication network between cities

2009 International Conference on Computational Science and Engineering

Scaling Behaviors in the Communication Network Between Cities

Gautier Krings*†, Francesco Calabrese†, Carlo Ratti† and Vincent D. Blondel*
*Department of Applied Mathematics, Université catholique de Louvain
4 Avenue Georges Lemaitre, B-1348 Louvain-la-Neuve, Belgium
email: gautier.krings@uclouvain.be
†SENSEable City Laboratory, Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge MA 02139, USA
email: fcalabre@mit.edu

***Abstract*—We analyze the anonymous communication patterns of 2.5 million customers of a Belgian mobile phone operator. With these communications, we construct a social network of customers, that we call microscopic network. Grouping customers together by billing address city, we obtain a network of cities, which we call the macroscopic network, built from 571 towns and cities in Belgium. We show that the macroscopic network has both a degree distribution and edge weight distribution with lognormal characteristics. We find that inter-city communications can be characterized by a gravity model: the intensity of communication between two cities is proportional to the product of the two populations divided by the square of the distance between the cities. Furthermore, we observe that intra-urban communications scale superlinearly with city population.**

I. Introduction

In the last decade, social network analysis has attracted increasing interest. This increase of interest is mainly due to the availability of large human-generated datasets, that allow wide-scale analysis of social behavior. Mobile phones have played an important role in this content and various studies of anonymized datasets have allowed significant progress in understanding the structure of our social network [1], our ability to gather in communities [2], the variety of our ways to create connections [3], as well as the dynamics of our contacts with friends [4].
Recently, a geographical component was added to mobile phone networks. This component has been added in two different ways, either by tracking users each time they make a phone call, or by assigning to every user a home address, assuming that his activities are centered in that position. Both ways have given remarkable results, the first method enabled to infer a model of human displacement [5], while the second showed that the probability for two people to know each other decreases with physical distance [6].
Yet, this second way allows another kind of analysis, which has not yet been explored. Based on the location of the customers addresses (typically at ZIP code level), one is able to group those customers together by cities and analyze how these groups interact. Similar analysis has been done based on different datasets and aggregation level [7], [8], [9].
The aim of this work is to analyze a communication network between Belgian cities that is constructed based on individual calling patterns.
From the anonymized mobile phone communications of a Belgian operator, we infer a human social network where nodes are customers and edges are weighted by the total time of call between them. We call this network the micro-network. By grouping customers by residential ZIP codes, we also infer a social network of cities, where nodes are cities and edges are weighted by the total time of call between their inhabitants. This network is the macro-network.
The article is structured as follows : in section 2, we present the analyzed data, as well as the definitions of cities we use. In section 3, we construct the two layers of the network, and analyze how volume of calls relate to the size of cities and the impact of distance. Finally, in section 4, we provide some conclusions.

II. Micro network: the communication network of Belgian customers

We consider mobile communications made by more than 3.3 million customers of a Belgian mobile phone operator over a period of 6 months in 2006 [6]. Every customer is identified by a surrogate key and to every customer we associate its corresponding billing address zip code. To construct the micro-network, we have filtered out calls involving other operators (there are three main operators in Belgium), incoming or outgoing, and we have kept only those transactions in which both the calling and receiving individuals are customers of the mobile phone company. To eliminate “accidental calls”, we have kept links between two customers i and j only if there are at least six calls in both directions during the 6 months time interval[1]. The resulting network is represented by a weighted graph $G(V,E)$ of 2.5 million nodes and of 38 million links. The link between customers i and j is weighted by their communication intensity, the total communication time in seconds l_{ij}
To analyze the relation between this social network and

[1]The limit of 6 is the one chosen by the authors of [6]. A limit is needed so as to remove strongly connected hubs, such as call centers, from the network and to remove connections that are not human to human connections. The authors of [6] report that their observations are robust with respect to modifications of the bound of 6 calls.

978-0-7695-3823-5/09 $26.00 © 2009 IEEE
DOI 10.1109/CSE.2009.272

geographical positioning, we assign customers to cities based on their billing address zip code. Belgium is a country of approximately 10.5 million inhabitants, with a high population density of 344 inhab./km^2. The Belgian National Institute of Statistics (NIS) [10] divides this population in 571 cities (cities, towns and villages), whose sizes show an overall lognormal population distribution with approximate parameters $\mu = 4.05$ and $\sigma = 0.37$.[2]

The analyzed communication network provides information for the operator's customers rather than for the entire population. However, the number of customers for each city follows the same lognormal distribution as the total population, this may suggest that our dataset is not structurally biased by particular user-groups and market shares. This is also confirmed by having the rank of city population sizes, (Fig. 1), match the rank of customers. By population of a city, we will furthermore mean the number of customers living in this city.

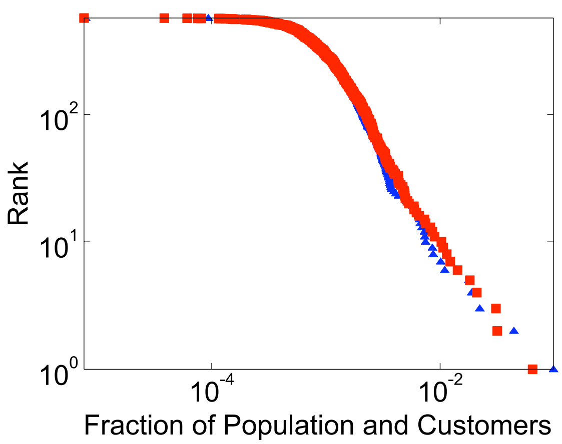

Fig. 1. Ranks of city population sizes (blue triangles) and number of customers (red squares) follow similar distributions.

III. Macro network: the communication network of Belgian cities

By aggregating the individual communications at city level, we obtain a network of 571 cities in Belgium.

In this macro-network, each node represents a group of nodes in the micro-network. To each node A of the macro-network, we associate a set $s_A = \{i_1, i_2, \dots\}$ that contains the nodes of the micro-network that belong to it. As each node of the micro-network is assigned to one and only one city, the sets $s_{A_1}, s_{A_2}, \dots$ are disjoint and their union is the set of all nodes V.

$$s_{A_m} \cap s_{A_n} = \emptyset \qquad \forall A_m \neq A_n,$$
$$s_{A_1} \cup s_{A_2} \cup \cdots \cup s_{A_N} = V$$

[2]The lognormal distribution of city size is consistent with similar data on US cities [11].

We can then define for each node A of the macro-network its mass M_A as $|s_A|$, namely the number of nodes of the micro-network that belong to it.

We define the intensity of interaction between the the nodes A and B by (Fig. 2):

$$L_{AB} = \sum_{i \in A,\, j \in B} l_{ij}.$$

Finally, we also define three measures of node strength, namely the incoming strength L_{*A}, the outgoing strength L_{A*} and the internal strength L_{AA} as follows:

$$L_{*A} = \sum_{i \notin A, j \in A} l_{ij}, \qquad L_{A*} = \sum_{i \in A, j \notin A} l_{ij},$$
$$L_{AA} = \sum_{i \in A, j \in A} l_{ij}.$$

In network analysis, classical clustering is based on the network structure, like in community detection, where strongly connected nodes are aggregated together in communities [12]. Here, we use a different clustering method, where the criteria of aggregation does not depend on the network structure, but on geographical proximity. Although this kind of networks is widely met in daily life, few analysis has been done on the properties of macroscopic networks, and about the influence of the microscopic level on the macroscopic one. Examples of these aggregations can be found in various contexts, spanning from economics to transport [13], [14].

A. Degree and Edge weight distributions

Analyzing the macro-network, we observe properties that are sharply different from typical social network characteristics. For example, the degree distribution is relatively narrow and spans from approximatively 100 to 570. It also presents a lognormal distribution (Figure 2), in opposition to the power law traditionnaly found in social network analysis.

The same lognormal distribution is also observed for the distribution of the intensities L_{AB} (Figure 3). The interest of this lognormal distribution is that, contrary to power laws, the values are equally distributed around the average, while the average of a power law distribution is usually not representative of the distribution.

B. Scaling behaviors in the intensity of interaction

The incoming and outgoing strengths represent the total time spent on the phone by customers of the city with the rest of the country. These values are strongly correlated with the mass M_A of the node, as defined previously.

As shown on Figure 4, both incoming and outgoing strength scale linearly with the mass ($L_{*A}, L_{A*} = kM_A^{\beta}$, $\beta = 0.96$, confidence interval: [0.93,0.99], $R^2 = 0.87$). Also, incoming and outgoing strength are strongly symmetric ($L_{A*} \approx L_{*A}, \quad \forall A$), that is, calls in one direction find always match in the opposite direction.

The same analysis can be made for internal strength, which are the time of communication between customers of the same city

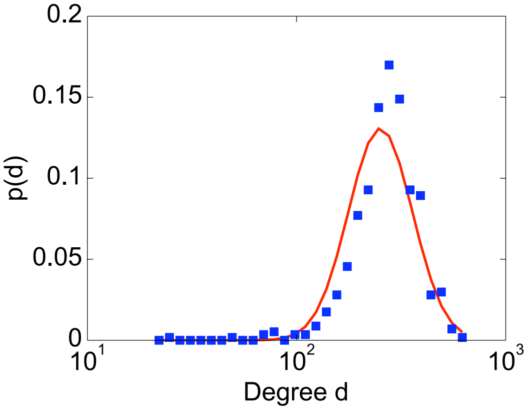

Fig. 2. Degree distribution of the macroscopic network, self edges are not considered. The red curve shows the best lognormal fit, with parameters $\mu = 2.4$ and $\sigma = 0.15$

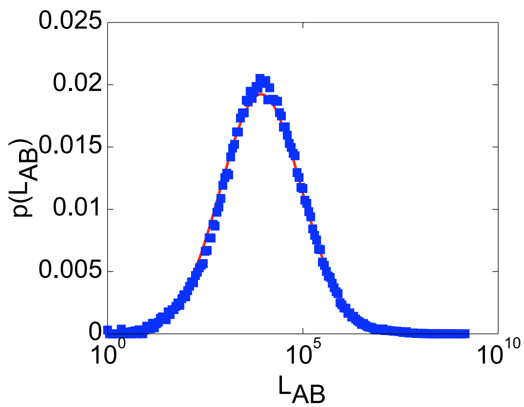

Fig. 3. Edge weight distribution of the macroscopic network, self edges are not considered. The red curve shows the lognormal best fit, with parameters $\mu = 3.93$ and $\sigma = 1.03$

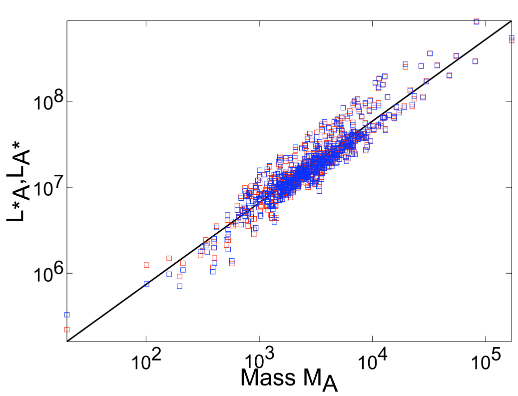

Fig. 4. Incoming and outgoing strengths (L_{*A} and L_{A*}) in function of the node mass M_A. Black line represents the best linear fit, with slope 0.96.

(see Figure 5). In this case, we observe a superlinear relationship ($\beta = 1.17$, confidence interval: [1.14,1.19], $R^2 = 0.95$),

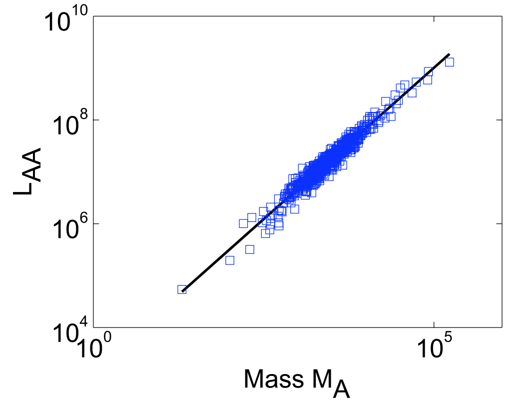

Fig. 5. Internal strength L_{AA} in function of the node mass M_A. Black line represents the best linear fit, with slope 1.17.

meaning that the amount of internal communication grows faster than population.
This is in line with previous results [9]. In particular, they showed that several factors linked with cities' dynamism scaled superlinearly with city size, leading the authors to the conclusion that innovation rate and pace of life increase superlinearly with city size.

C. Gravity model for intercity communication

Based on the observations we made, let us try to develop a model of communication between cities.
Our first result is that the flow of communication from and to a city is proportional to its mass: $L_{A*}, L_{*A} \approx kM_A$. Secondly, these flows are strongly symmetrical: $L_{A*} \approx L_{*A}$.
We finally tested the impact of distance on intercity communication, and found that the intensity of the edges in the macro-network decreases with $\frac{1}{d^2}$, where d is the distance between the centroids of the cities.
These three results lead naturally to a gravitational model, which is similar to models developed for economical exchange [15], traffic flow [14] and other socio-economical networks [8], [16]. We tried to fit the intensity of interaction between cities with a model

$$L_{AB} \propto \frac{M_A M_B}{d_{AB}^2}.$$

To ensure the validity of our results, we plotted the estimated intensity given by the gravity model versus the observed intensity. As shown on Figure 6, the results match particularly well for pairs of cities A and B that have a large value for L_{AB}. Let us finally observe that this gravity model is consistent with the results presented in [6] that described the probability of connection between customers based on their distance. One can check that the intensity of communication between two customers that make a link does not vary with the distance between them. Our gravity model can thus be derived using the model described in [6], and by hypothesizing a homogeneous distribution of people around

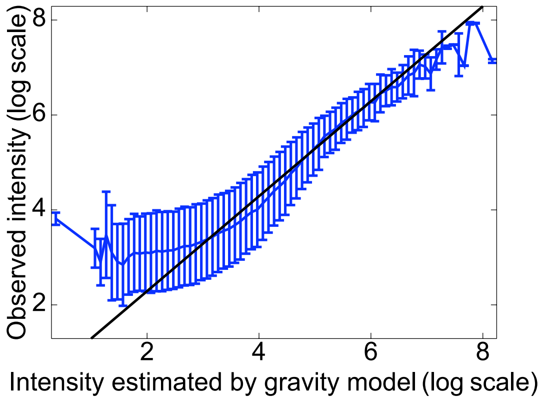

Fig. 6. Communications intensity between pairs of cities versus the ration $\frac{M_A M_B}{d_{AB}^2}$. The black line shows the gravitational law.

the centroid of each city.

IV. Conclusion

We have introduced the definition of microscopic and macroscopic networks. Micro-networks are typical human-to-human social networks of interaction. Macro-networks are made from the aggregation of a micro-network, and in our case the aggregation is not based on network structure, in contrast as usual network aggregation methods, like community analysis. In our study the aggregation criteria is a geographical criteria, namely the billing address ZIP code.

We analyze a communication network made of 2.5 million customers of a Belgian mobile operator as proxy for a social network, and use the total time of communication as a measure of the strength of social interaction.

We discover that calls internal to cities show power law properties and scale superlinearly with population. This result is in agreement with previous research that showed that innovation rate in a city scales superlinearly with city size. Secondly, each pair of cities communicates following a gravitional model which is in line with previous results from the literature of economical and transport networks.

This analysis is exploratory and attempts to open up new research directions on how people organize their social network in function of the city in which they live, and more generally, on how geography influences social networks.

References

[1] J.P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, M.A. de Menezes, K. Kaski, A.L. Barabasi, and J. Kertesz. Analysis of a large-scale weighted network of one-to-one human communication. *New Journal of Physics*, 9(6):179, 2007.

[2] V.D. Blondel, J.L. Guillaume, R. Lambiotte, and E.L.J.S. Mech. Fast unfolding of communities in large networks. *J. Stat. Mech*, page P10008, 2008.

[3] Mukund Seshadri, Sridhar Machiraju, Ashwin Sridharan, Jean Bolot, Christos Faloutsos, and Jure Leskove. Mobile call graphs: beyond power-law and lognormal distributions. In *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 596–604, New York, NY, USA, 2008. ACM.

[4] C.A. Hidalgo and C. Rodriguez-Sickert. The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications*, 387(12):3017–3024, 2008.

[5] M.C. González, C.A. Hidalgo, and A.L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[6] R. Lambiotte, V.D. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Van Dooren. Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, 387(21):5317–5325, 2008.

[7] D. Brockmann and F. Theis. Money Circulation, Trackable Items, and the Emergence of Universal Human Mobility Patterns. *IEEE Pervasive Computing*, 7(4):28–35, 2008.

[8] K. Bhattacharya, G. Mukherjee, J. Saramaki, K. Kaski, and SS Manna. The International Trade Network: weighted network analysis and modelling. *J. Stat. Mech*, page P02002, 2008.

[9] L. Bettencourt, J. Lobo, D. Helbing, C. Kuhnert, and G.B. West. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104(17):7301, 2007.

[10] Belgian national institute of statistics. http://www.statbel.fgov.be/home_en.asp.

[11] J. Eeckhout. Gibrat's Law for (All) Cities. *American Economic Review*, 94(5):1429–1451, 2004.

[12] MEJ Newman. Detecting community structure in networks. *The European Physical Journal B-Condensed Matter*, 38(2):321–330, 2004.

[13] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101(11):3747–3752, 2004.

[14] W.S. Jung, F. Wang, and H.E. Stanley. Gravity model in the Korean highway. *EPL-Europhysics Letters*, 81(4):48005–48005, 2008.

[15] J. Tinbergen. An Analysis of World Trade Flows. *Shaping the World Economy, Tinbergen, ed. New York: Twentieth Century Fund*, 1962.

[16] G. Olsson. *Distance and Human Interaction: A Review and Bibliography*. Regional Science Research Institute, 1965.