Urban Gravity: a Model for Intercity Telecommunication Flows

Gautier Krings[1,2]**, Francesco Calabrese**[2]**, Carlo Ratti**[2] **and Vincent D Blondel**[1]

[1] Université catholique de Louvain (UCL), Department of Applied Mathematics -
4 Avenue Georges Lemaitre, B-1348 Louvain-la-Neuve, Belgium
[2] SENSEable City Laboratory, Massachusetts Institute of Technology -
77 Massachusetts Avenue, Cambridge, MA 02139, USA

E-mail: `gautier.krings@uclouvain.be`

Abstract. We analyze the anonymous communication patterns of 2.5 million customers of a Belgian mobile phone operator. Grouping customers by billing address, we build a social network of cities, that consists of communications between 571 cities in Belgium. We show that inter-city communication intensity is characterized by a gravity model: the communication intensity between two cities is proportional to the product of their sizes divided by the square of their distance.

PACS numbers: 89.75.Da, 89.75.Fb, 89.65.Ef

Recent research has shown that certain characteristics of cities grow in different ways in relation to population size. While some characteristics are directly proportional to a city's population size, instead, other features - such as productivity or energy consumption - are not linear but exhibit superlinear or sublinear dependence to population size [1]. Interestingly, some of these features have strong similarities with those found in biological cells - an observation that has led to the creation of a metaphor where cities are seen as living entities [2]. Interactions between cities, such as passenger transport flows and phone messages, have also been related to population and distance [3, 4]. Meanwhile, in socio-economic networks, interactions between entities such as cities or countries have led to models remembering Newton's gravity law, where the sizes of the entities play the role of mass [5]. Road and airline networks between cities have also been studied [6, 7], and in the case of road networks, it appears that the strength of interaction also follows a gravity law.

While these results have provided a better understanding of the way cities interact, a finer analysis at human level was until now difficult because of a lack of data. Recently, however, telephone communication data has opened up a new way of analyzing cities at both a fine and aggregate level (whereby as Gottman as already as 1957 noted [8]: "the density of the flow of telephone calls is a fairly good measure of the relationships binding together the economic interests of the region"). Several large datasets of email and phone calls have recently become available. By using these as a proxy for social networks, they have enabled the study of human connections and behaviors [9, 10, 11, 12, 13]. The use of geographical information makes it possible to go one step further in the study of individual and group interactions. For example, Lambiotte *et al.* use a mobile phone dataset to show that the probability for a call between two people decreases by the square of their distance [14]. However, while the structure of complex networks has already been widely studied [15, 16, 17, 18], to date, contributions have not yet analyzed large-scale features of social networks where people are aggregated based on their geographical proximity.

In this work, we study anonymized mobile phone communications from a Belgian operator and derive a model of interaction between cities. Grouping customers together by billing address, we create a two-level network, containing both a microscopic network of human-to-human interactions, and a macroscopic network of interactions between cities.

The data that we consider consists of the communications made by more than 2.5 million customers of a Belgian mobile phone operator over a period of 6 months in 2006 [14]. Every customer is identified by a surrogate key and to every customer we associate their corresponding billing address zip code. In order to construct the communication 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. In order 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. The resulting network is composed of 2.5 million nodes and 38 million links. To the link between the customers i and j we associate a communication intensity by computing the total communication time in seconds l_{ij} between i and j.

In order to analyze the relationship between this social network and geographical positioning, we associate 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./km². The Belgian National Institute of Statistics (NIS) [19] divides this population into 571 cities (cities, towns and villages), whose sizes show an overall lognormal population distribution with approximate parameters $\mu = 4.05$ and $\sigma = 0.37$‡.

The analyzed communication network provides information for the operator's
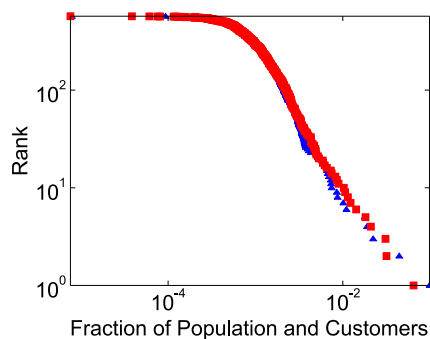

Figure 1. Ranks of city population sizes (blue triangles) and number of customers (red squares) follow similar distributions.

customers rather than for the entire population. However, the number of customers present in each city follows the same lognormal distribution as the total population and so this suggests that our dataset is not structurally biased by particular user-groups and market shares. This is also confirmed by the ranks of city population sizes that match with those of customers, as shown in Fig. 1. In the rest of this article, when we use the term *population* of a city, we are refering to the number of customers that have the corresponding ZIP code billing address, even if they are not connected to any other customer in the microscopic network of human-to-human communications defined above. These customers are still taken into account for the population size, since their presence is of interest for the normalization of the communication data.

By aggregating the individual communications at a city level, we obtain a network of 571 cities in Belgium. We define the intensity of interaction between the cities A and B by (Fig. 2 (a)):

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

The distribution of intensity is narrow and it appears to follow a lognormal distribution (Fig. 2 (b)). A similar lognormal distribution is found for the degree distribution. This

‡ The lognormal distribution of city size is consistent with similar data on US cities [20].
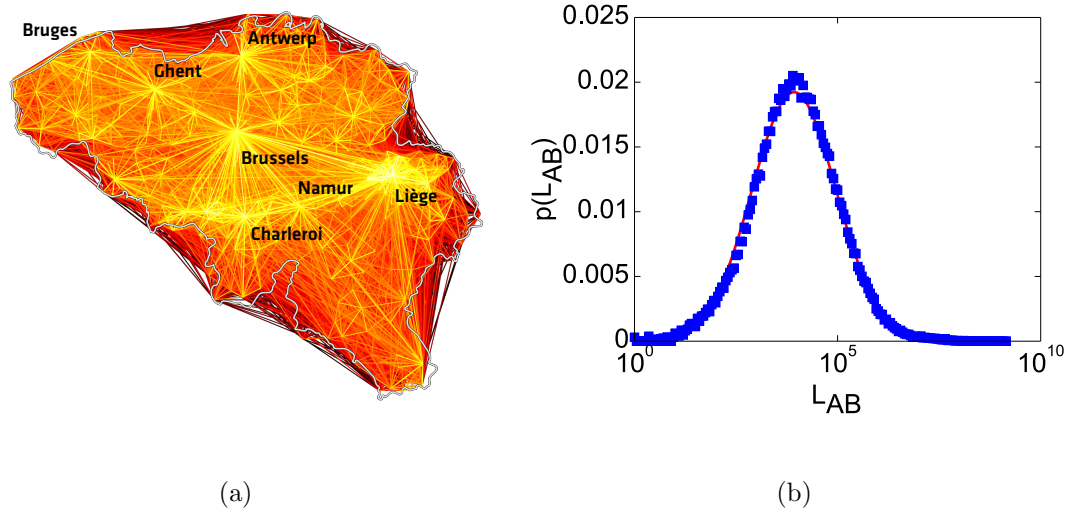
(a)   (b)

Figure 2. (a) Illustration of the macroscopic communication network (only the top 30% of the links (having the strongest intensity) are represented). Colors indicate the intensity of communication between the cities: bright colors indicate a strong intensity. (b) Intensity distribution of the macroscopic network, self edges are not considered. The red curve shows the lognormal best fit, with parameters $\mu = 3.93$ and $\sigma = 1.03$

lognormal intensity distribution is sharply different from what is typically observed in social networks but is consistent with observations in other macroscopic networks, such as the intensity of trade between countries, obtained by aggregating the individual trades made by agents [21].

Many studies have been made on human-to-human communications but few analyses are available on how these communications, once aggregated at the city level, are reliant on the properties of that city. In the following, we model the communication intensity between cities as a function of the population sizes and of the distance between them.

First, we analyze how communication flowing into and out of cities, scale with population size. For doing this, for each of the 571 cities we compare the total incoming (L_{*A}) and outgoing (L_{A*}) communication intensities, as defined below, to the population sizes of these cities.

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

As shown on Figure 3 (a), both incoming and outgoing inter-city communication intensities scale linearly with city size ($L_{A*}, L_{*A} = kP_A^{\beta}$, $\beta = 0.96$, confidence interval: [0.93 0.99], $R^2 = 0.87$). Also, incoming and outgoing communications are strongly symmetric ($L_{A*} \approx L_{*A}, \quad \forall A$), that is, calls in one direction always find a match in the opposite direction.

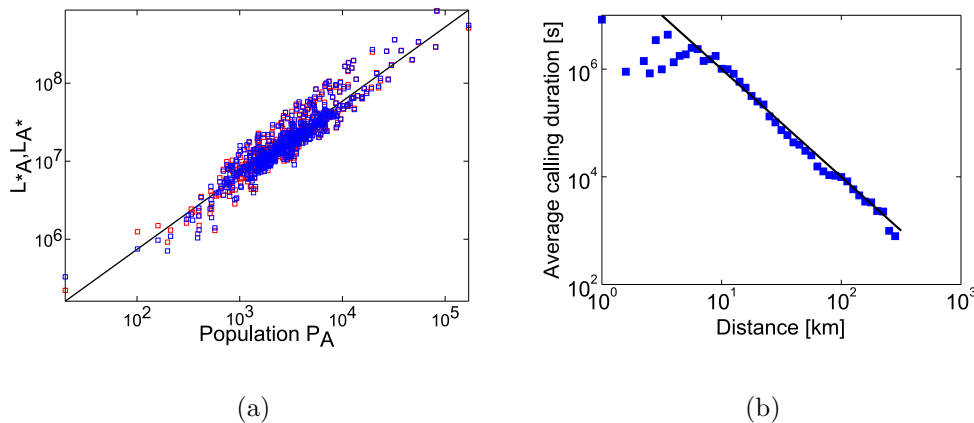Another parameter that influences communication intensity between cities is distance.

Figure 3. (a) Relation between the outgoing intensity L_{A*} (blue), the incoming intensity L_{*A} (red) and the city population size P_A. (b) Dependence of the average communication intensity between pairs of cities and the average distance separating them. The black line shows a $\frac{1}{d^2}$ decrease.

It seems reasonable to expect that the intensity of communication between two cities will decay when distance increases and other parameters are kept unchanged. This idea is supported by several studies that suggest gravity-like models for car traffic [6], trade [21] or economic activity [22]. In all these cases, the intensity at distance d is proportional to $\frac{1}{d^2}$. For the case of communication, a similar model has been presented in [3], but with an intensity evolving like $\frac{1}{d}$, though the authors acknowledge that their model does not seem to fit well with the data.

To test the influence of distance on the Belgian network, we measure it as the distance between the centroids of each cities area. The communication intensity between two cities is then compared with the distance between them, where a power law decrease, with an exponent close to -2 is shown (see Fig. 3 (b)). The power law fits very well for inter-city distances larger than 10 km.

This result suggests that the communication between cities is ruled by the following gravity model, which is symmetric, scales linearly with the population sizes and decreases with d^2:

$$L_{AB} = K \frac{P_A P_B}{d_{AB}^2},$$

there, the scaling constant K is the gravity constant for a timespan of 6 months of calling activity.

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 4, the results match particularly well for pairs of cities A and B that have a large estimated intensity. The gravitational constant K can be estimated with a simple best fit of the data. If, over the 6 months of data, the intensity of communication is constant, we obtain the general gravity constant $G = 1.07 \cdot 10^{-2} \frac{s}{day}$. This constant enables us to estimate
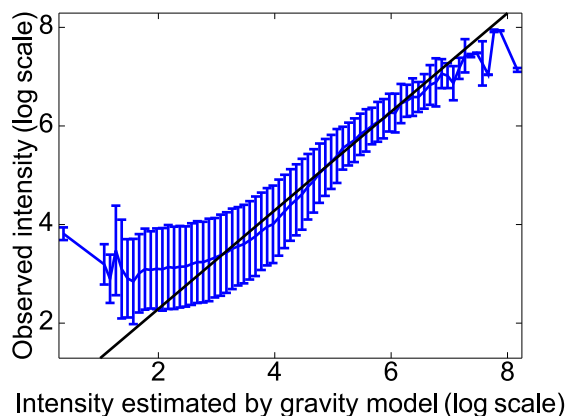
Figure 4. Communication intensity between pairs of cities versus the ratio $\frac{P_A P_B}{d_{AB}^2}$. The black line shows the gravitational law.

the intensity of communication between any pair of Belgian cities, based on population, distance and duration of the considered period. Let us finally observe that this gravity model is consistent with the results presented in [14] 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 much with the distance between them (see Figure 5), so, the distance decay observed in Figure 3 (b), does not result from a weaker intensity of communication between customers, but from a smaller fraction of customers communicating with each other.

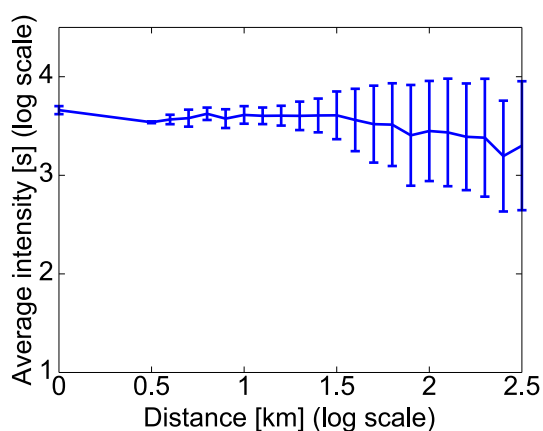The gravity model for inter-city communication intensity is analogous to other models

Figure 5. Average intensity of communication between pairs of nodes, if they make a link, versus the distance separating them.

of economic and transportation networks, but has to be considered cautiously as it might be biased by the nature of the data. First of all, Belgium is a small country with a specific demography and two main language communities. Secondly, we note that our study relies on the definition of census areas, as defined by the Belgian National Institute

of Statistics. It is certainly possible that changes to this definition might have non-trivial implications for the shape of the network. Finally, we only consider the customers of one operator for whom the billing address zip code is available, which means that these customers have a contractual plan. However, since the observed results suggest clear behavioral influences, it seems unlikely that these results reflect biased or corrupt data. Also, the lognormal shape of city sizes and edge intensities, as well as the power law behavior of calls, is consistent with studies made in other countries using different types of networks.

This work is exploratory, but raises various interesting questions on how people organize their social network depending on the city they live in, as well as, more generally, on the influence of geography on social networks. More significantly, the analysis of social networks, as captured by digital networks like phone calls and email, when overlaid on physical space, could help to improve our understanding of the structure of cities. Or how do they grow into places, where, as Lewis Mumford [23] so eloquently stated: "the diffused rays of many separate beams of life fall into focus, with gains in both social effectiveness and significance"?

Acknowledgements

References

[1] 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.

[2] J.J. Macionis and V.N. Parrillo. *Cities and Urban Life*. Prentice Hall, 2001.

[3] G.K. Zipf. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley Press, 1949.

[4] Wayne K. Davies. Urban connectivity in montana. *The Annals of Regional Science*, 13(2):29–46, 1979.

[5] V. Carrothers. A historical review of the gravity and potential concepts of human relations. *Journal of the American Institute of Planners*, 22:94–102, 1958.

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

[7] 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.

[8] J. Gottman. Megalopolis: the Urbanization of the Northeastern Seaboard of the United States. *Economic Geography*, 33(7):31–40, 1957.

[9] P.S. Dodds, R. Muhamad, and D.J. Watts. An Experimental Study of Search in Global Social Networks. *Science*, 301(5634):827–829, 2003.

[10] J.P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, and A.L. Barabasi. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332, 2007.

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

[12] 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.

[13] 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.

[14] 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.

[15] D.J. Watts. *Small worlds: the dynamics of networks between order and randomness.* Princeton University Press, 1999.

[16] A.L. Barabasi. *Linked: how everything is connected to everything else and what it means for business, science, and everyday life.* Plume Books, 2003.

[17] D.J. Watts and S.H. Strogatz. Collective dynamics of small-world' networks. *Nature*, 393(6684):440–442, 1998.

[18] MEJ Newman. The structure and function of complex networks. *SIAM Rev.*, 45(2):167–256, 2003.

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

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

[21] 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.

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

[23] L. Mumford and B.S. Turner. *The culture of cities.* Secker & Warburg London, 1938.