XXXV CONFERENZA ITALIANA DI SCIENZE REGIONALI

THE "DEATH OF DISTANCE" IN THE AGE OF URBAN SENSING

Michele Ferretti[1]

ABSTRACT

The paper presents a novel approach to the "Death of Distance - Flat World" hypothesis, based on the use of innovative data sources gathered from User-Generated (UG) data, and techniques of applied statistics. Despite in fact the numerous examples present in the literature addressing the same topic, few have insofar dealt specifically with the possibilities offered by the contemporary massive production of data, and even less have applied modern analytical algorithms to such data. On the contrary, the present research aimed at deriving new insights regarding the interaction of economic phenomena with and within the spatial dimension, by explicitly looking at the hidden patterns of mobility and interaction resulting from mobile phone sensors. The techniques applied to investigate the multiple dimensions in which those interactions unfold outline also the potential and the convenience of such a methodological approach in respect to urban analysis at the local and regional scale.

[1] Politecnico di Milano, Piazza Leonardo da Vinci 32, 20100, Milan, Italy, e-mail: michele.ferretti@mail.polimi.it

1. Introduction

Despite a certain line of thought mainly developed across the end of the 90's and the beginning of the 00's, which claimed for a supposed "Death of Distance" (Cairncross, 2001), arguing for a "Flatter World" (Friedman, 2005) in respect to social and economic phenomena, hard evidence still seems to stand up for the contrary (McCann 2008; Rodriguez-Pose et al. 2008). The same spread of ICT and new communication technologies that supposedly would have put an "End to Geography" (Bates, 1995), has indeed provided plenty of data for successive studies that assessed the relevance of geographic dimension in shaping nowadays interactions between and within urban environments (Tranos et al. 2012). The impressive rise of new technologies, able to remotely sense data, crowd-sourced in real time over fine-grained spatial and temporal scales (the so-called "Big Picture" (Nelson, 2013) have conducted leading commentators, such as (Batty, 2012), to openly talk of the emergence of new theoretical paradigms, and of the need for new theories and models able to make a good use of tomorrow's "wired" cities scenario. Among these new sources of data, certainly mobile communication technologies play a dominant role, given the scale and the pervasiveness of their spread, particularly in some regions of the developing world (Sutherland, 2011). African countries for instance, structurally lacking most of the resource, in financial and governance capacity terms, to implement and use a traditional data collection infrastructure (Manfredini et al., 2012), could greatly benefit from the use of an already present ubiquitous information network of cheap sensors. The resulting combination of this two elements, mainly the high availability at low cost of data from non-conventional sources, in an extremely needy context for overtaking such informational gap, has been the rationale at the base of UN's "Big Data for Development" white paper (United Nations, 2012). A document under the umbrella of the "Global Pulse" initiative, which recognizes the strategic opportunities represented by the use of new digital data sources, for conceiving and implementing social and economic growth policies in the developing countries. Within this framework the French mobile phone carrier Orange Mobile, released in the spring of 2013, a dataset for an open data analysis challenge: the "Data4Development". Participants were asked to devise innovative ways to address society development questions through the use of mobile phone data. Despite the sound results submitted for the competition, some gaps worth exploring can be found, mainly the fact that the contributions so far produced have been mostly focused on: (i) developing location-aware mobile applications, (ii) devising scenarios for planning transport systems, and (iii) disease- diffusion modelling. Scarce evidence, on the other hand, has been provided about local and regional economic patterns of interaction. Further, few have explicitly addressed the effects of geographical distance on the formation and maintenance of relationships among economic actors, and even less have been able to give meaningful

insights regarding the supposedly homogeneity and overall “flatness” of the world (Tranos et al. 2012). In particular in respect to one of the main flattening forces, ICTs diffusion (McCann, 2008), and on the effects it causes on urban environments, as well as on cities’ development and growth at large. It seems advisable to hence put this rich dataset to good use, trying to address the just mentioned theoretical gaps. Given the former premises, the present article is hence positioned at the convergence of different research paths: it aims to give a contribution in pushing further the understanding of socio-economic processes mediated through space, and on their effects on the urban milieu, To reach this goals, it proposes to use, for its analyses, data coming from the exact same technologies that the “World is Flat” hypothesis theorized as one of the pillars of the Globalization 3.0 homogeneous landscape (Friedman, 2005). By doing so, it further intends to dwell greatly in the new theoretical openings offered by the Big Data paradigm (United Nations, 2012), hopefully providing an additional example of the usefulness and richness of applying computer science methodologies, and data, to the service of urban research.

First
To what extent does ICT play a role in effectively overcoming geographic distance?

Second
What is the emerging picture of urban realities in the Ivory Coast, and in the use of their spaces and places, as portrayed by user-generated mobile phone data?

2. Data

The main dataset employed in the thesis has been provided in the spring of 2013 by the French mobile phone carrier Orange for the “Data for Development (D4D)” open data analysis competition. The data were provided to various institutions, NGOs and research groups, among which the “Laboratorio Analisi Dati e Cartografia (LADeC)” spatial analysis laboratory at Politecnico di Milano. In line with the United Nation “Global Pulse” program (United Nations, 2012), Orange’s initiative has been undertaken to probe the potential inherent in user generated data collected from mobile devices, as a valid substitute of traditional data sources in fostering research and assisting development-related policies (Blondel et al., 2013). The dataset, gathered in the Ivory Coast, Africa, in the period December 1st 2011 - April 28th 2012, consist of anonymized Call Detail Records (CDR) of over 2.5 billion calls and SMS exchange among 5 million of Ivorian customers.
A first pre-processing phase, conducted by the Parisian Orange Labs, saw the formatting and anonymizing procedure of the data. Further, two additional steps have been taken: the first has been the sample’s homogenization by the means of removal of any new subscriber or re-signer customer, within the observation period. Additionally, also incoming and outgoing

calls has been paired in order to eliminate double counts (Blondel, Esch, Chan, & Clerot, 2012). Further steps have also been taken to protect both Orange's interests, as well as its customer's privacy, as discussed later. The resulting dataset, as given for the competition, consists of four sub-datasets, each provided as textual TAB delimited files (.TSV). Of those four, only the first one (SET1) has been employed for the analyses and is hereafter presented more in depth.

2.1 SET1 - Antenna-to-Antenna traffic dataset

The first sub-dataset is composed by the calling traffic activity recorded as connections between Orange' s antennae. It is composed by 5 variables:

- *Date & Hour = t*he timestamp of the recorded call. In the format YYYY-MM-DD hh:mm:ss
- *Originating Antenna ID* = the unique identifier of the antenna in which the call originated from.
- *Terminating Antenna ID* = the unique identifier of the antenna in which the call terminated to.
- *Number of Voice Calls* = the total number of calls aggregated by hour.
- *Duration of Voice Calls* = the sum of calling minutes within each hour time-slot.

The time resolution is equal to one hour. Consequently all calls between any pair of antennas have been aggregated accordingly. Multiple time slots calls are recorded only in the first time slot in which they appear. SET1 spans over the entire observation period, but consists only of communication between Orange customers (communication with customers of other providers have been removed). The spatial resolution is equal to the antenna's coverage radius, quantifying on average within hundreds meters. To permit the geographic tracing of the calling patterns, each antenna has been provided with a unique identifier ID, and the relative geographic location. Other ancillary dataset have been further utilized, either in their given format or after some pre-processing operations, to integrate the "D4D" dataset with additional information regarding demography trends, socio-economic processes, and the Ivorian physical landscape, as suggested in (Blondel et al., 2013).

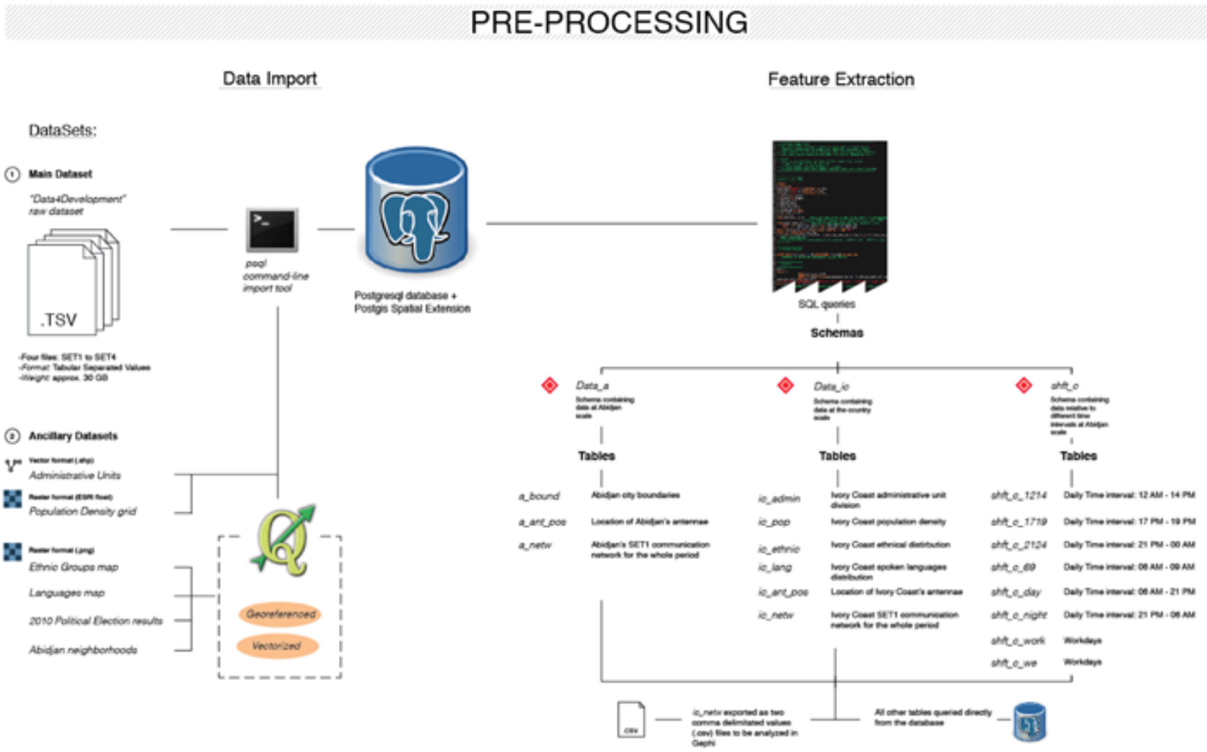
Among the suggested datasets, the following might have been of particular interest:

- *The OpenStreetMap set*, which provides a de- tailed and scalable source of information regarding the physical features of the Ivory Coast (in particular its infrastructure system)
- *The D4D Boundaries set*, comprising the administrative boundaries at the subnational level

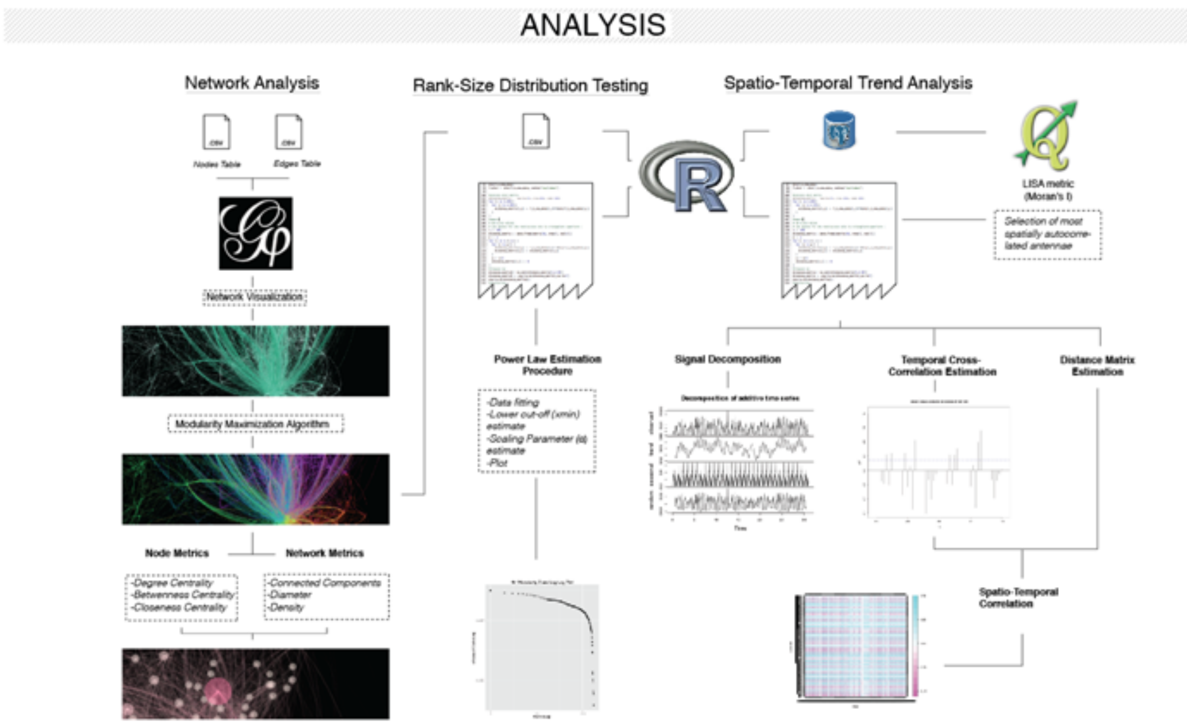
- *The Afripop.org set*, which provides an accurate reconstruction of the Population Density, as derived from high resolution satellite imagery, combined with land cover maps
- *The Ethnologues.org* ethnicities mapping, useful to verify the influence of linguistic and ethnic differences on proximity relations, in the vein of (Losada, Benito, Morales, & Creixell, 2013) and (Andris et al. 2013).

3. Methodology

The proposed methodological approach can be briefly sketched as composed by a pre-processing phase, followed by three main analytical steps (see below Figure 1). The pre-processing consisted roughly in the transformation of the raw data into more manageable and processable formats, and in their dimensional reduction into smaller subsets to ease the computational effort related to their processing. This phase required the import of the whole dataset into a suitable analytical framework, mainly a spatially enabled database system (PostgreSQL), where the data have been sliced into proper subsets, and from which they have been accessed and queried in later steps. Following this phase, three main analytical steps have been taken, mainly a visualization and network analysis of the SET1 communication network, a test phase aimed at investigating the presence of a Power-Law distribution in the calls data, and a final analysis of the calling signals profile through the time dimension, aimed at investigating a possible correlation with the spatial dimension of the phenomena rooted behind the calls' signals profiles.



(a) Pre-Processing Phase Diagram



(b) Analysis Phase Diagram

Figure 1 (a, b): Data Analysis Overview

3.1 Network Analysis

Various analytical tools, derived from the tradition of Social Network Analysis (SNA) have been applied on the SET1 Communication Network, in order to explore its characteristics and dynamics. Both the analytical as well as the visualization part have been carried out within Gephi network analysis software.

3.1.1 Modularity Maximization Algorithm

The first passage in the analysis of the Communication Network derived from SET1 data in the previous phase consisted in its partition into smaller sub-graphs. This served two purposes, mainly to gain more easily a deeper initial insight of the dynamics at play within the graph at large, and to orient the further analytical steps towards more significative areas of the country/graph.
Similarly to what done by (Amini, Kung, Kang, Sobolevsky, & Ratti, 2013), the Modularity Maximization Algorithm (also known as Louvain Algorithm) has been employed to partition the SET1 network. This tool has in fact proved to be the right instrument for detecting communities in a communication graph derived from mobile phone data (Dinh et al. 2013). Communities detection is already far from being a trivial task, since even a relatively banal operation can quickly escalate into a truly demanding computational problem (Dinh et al. 2013). In addition, this class of algorithms have had to endure in recent years an even more pressing demand for efficiency, due to the increased dimension of the available network datasets. Yet, despite such challenges, the Modularity Maximization approach defined by (M. E. J. Newman, 2006) has been recognized to be more computationally efficient, even on large scale graphs, in respect to other methodologies (Blondel et al. 2008).
More formally modularity can be described as a network property, a scalar value between -1 and 1 that measures the density of links inside communities as compared to links between communities (Girvan et al. 2002):

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{i,j} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

In this formalization, modularity Q is a function of a distance δ of two parameters c_i and c_j, representing the communities to which each vertex *i(j)* is assigned. This is multiplied to the sum of the differences between the weight of the edges between node *i* and node j, $\sum A_{i,j}$, and the weigthed products of the edges attached to nodes i and j (respectively represented by $k_i = \sum A_{i,j}$ and $k_j = \sum A_{j,i}$) over the term $m = \frac{1}{2} \sum A_{i,j}$. The distance function δ assumes values equals to 1 if $u = v$ and 0 otherwise (Girvan et al. 2002). What this entails, is that the

algorithm operates with two steps, one to compute the similarity distance within a community, and a second to aggregate the found sub-communities into new ones. The iteration of such procedure until no separation is further possible leads to the partition of the original graph into new clusters of sub-communities, called modularity classes, where the inter-node connectivity is higher than the intra-node's.

3.1.2 Social Network Analysis metrics

The sub-networks resulting from the initial partition have been further investigated by the means of traditional metrics belonging to SNA, and aimed at assessing the different networks' properties, through the account of the role and importance of each individual node in the network:

i. *Network Diameter* = The largest distance path to be found in a network (or graph).
ii. *Average Path Length* = The mean length (in terms of number of steps) for all the possible combinations of node pairs in a network.
iii. *Average Degree* = The mean degree of all the nodes in the network
iv. *Density* = The ratio of edges present in the graph, over the maximum number of edges theoretically possible to establish among the nodes.
v. *Strongly Connected Components* =A partition of a direct the graph in which all the nodes present the characteristics of being mutually directly reachable.
vi. *Weakly Connected Components* = In a directed graph, by exclusion from the above statement, a set of a weakly connected component is the maximal group of nodes that are mutually reachable by violating the edge directions.
vii. *Degree Centrality (Bonachich's Centrality)* = this metric returns the relative importance of each node in the graph, giving a first account of the urban hierarchy structure. In general the Degree K_i of node i is defined as the total number of its connections. Bonachich's variant takes into ac- count also the number of connection of each 1st order neighbour.

3.2 Power-Law Distribution Test

The theoretical relevance in assessing the presence of a power-law distribution (Zipf, 1949) for the study and comprehension of urban phenomena needs no introduction. Suffice to briefly review here its main traits. In the most disparate scientific fields, from linguistics to thermodynamics the observed distribution of a social (or natural) phenomena in relation to its frequency or magnitude has been proven to follow a particular trend, best approximated by a power law distribution, a particular distribution characterized by invariance of scale (Zipf

1949; Gabaix 1999; Chen et al. 2004; Gabaix et al. 2004; Newman 2005; Nitsch 2005; Clauset et al. 2009; Dittmar 2009; O'Connor 2009; Reggiani et al. 2012).
From a mathematical point of view, a quantity x follows a power-law distribution if it is drawn from a probability distribution

$$\mathbb{P} \propto (x)^{-\alpha}$$

with α being an exponential parameter of constant value, known as scaling parameter (Clauset et al., 2009). Regarding city-size distribution, this scaling parameter is equal to 1 in the case of a perfect urban hierarchy, with cities distributing proportionally, starting from the major urban hub of rank 1 (Reggiani & Nijkamp, 2012).
To adapt the former equation to the case of cities, the Law of Zipf holds if city sizes are distributed such that the probability of drawing a city with population size S greater than some threshold N is

$$\mathbb{P}(S > N) = \alpha N^{-\beta}$$

which can be compared to the equation

$$R = \alpha S^{-\beta}$$

with R size rank of a city being inversely proportional to its size S in terms of population. The latter equation can be written in a logarithmic form as

$$\log(R) = \log(\alpha) - \beta\log(S)$$

Plotting the data expressed in the latter form on a log-log scale allows for a more immediate comprehension of the phenomena (Dittmar, 2009).
Having thus established the relevance of the study of Zipf's Law for understanding the urban dimension, and in particular the urban hierarchy within a country, a less straightforward task is to define the correct methodological approach to assess its validity in the dataset hereafter analysed. According in fact to (Clauset, Newman, & Moore, 2004) the detection of power laws is complicated by two factors, mainly overcoming the bias introduced by the heavy-tails' fluctuations (occurring mostly within the distribution's upper-tail, in the presence, for instance, of larger-than-the-average cities), and by the difficulty in the estimation of the correct range over which the power-law actually holds. Two issues that, on the other hand, are not considered by the current methodologies, such as least square fitting. The approach employed in the current work follows the recommendations of (Clauset et al., 2009).
The nodes files exported from Gephi software as Comma-Separated-Values (.CSV) file have been imported in the R statistical analysis environment and stored in a data frame containing the information over the frequency of communication among the antennae for each of the six

modularity classes found in the previous step of the analysis. On such matrix a maximum-likelihood fitting method has been performed in order to estimate the two parameters recommended: the x_{min}, describing the starting point of power-law distribution, and the scaling parameter $alpha$. A goodness-of-fit test, based on the Kolmogorov-Smirnov statistic and likelihood ratio have been conducted for each of the six estimates via a bootstrap procedure of a 1000 iterations, which resulted in an easily interpretable p-value statistic (Clauset et al., 2009).

3.3 Spatio-Temporal Trend Analysis

Having so far visualized and analysed the interactions among and within urban centres through the lenses of Social Network Analysis tools, the last phase dealt with the analysis of the spatial and temporal dimension of such interactions. To assess the spatial variance in the data, and in particular their local distribution in space, two indicators have been chosen, mainly *Getis-Ord Gi** and *Local Moran's I*. These two metrics, firstly proposed by (Anselin, 1995), focus on the identification of local patterns of spatial association. They belong in fact to a class of metrics known as LISA - Local Indicators of Spatial Autocorrelation. Formally, they are function of two parameters, a variable y_i, observed at each location i, and the $y_{j,i}$ values observed in the neighbourhood J_i of I.

$$L_i = f\left(y_i, y_{j,i}\right)$$

Such class of indicators is suited for analysing the local dimension, which is extremely relevant to the thesis analysis. In fact, in respect to a global statistic, which might hide large spatial lags of auto-correlated values, the LISA metrics take into account the values in a neighbourhood contiguous to a feature, by calculating firstly a matrix of Spatial Weights, and lately using it to weight the relative "importance" of each value found in the surrounding neighbourhood.

i. *Getis-Ord G**

   Data: SET1 gridded density of calls (minutes)

   Aim: The metrics express fully the proportion of all x values in the study area accounted for by the neighbours of location i. It is in fact essentially a ratio of the sum of the values in the spatial neighbourhood of each feature (each grid's cell in this case) to the global total, without including also the feature value (contrary to Moran's I). It used therefore to investigate the presence of high or low spots of values within the dataset (hot-spot analysis)

ii. *Anselin Local Moran's I*

Data: SET1 gridded density of calls (minutes)

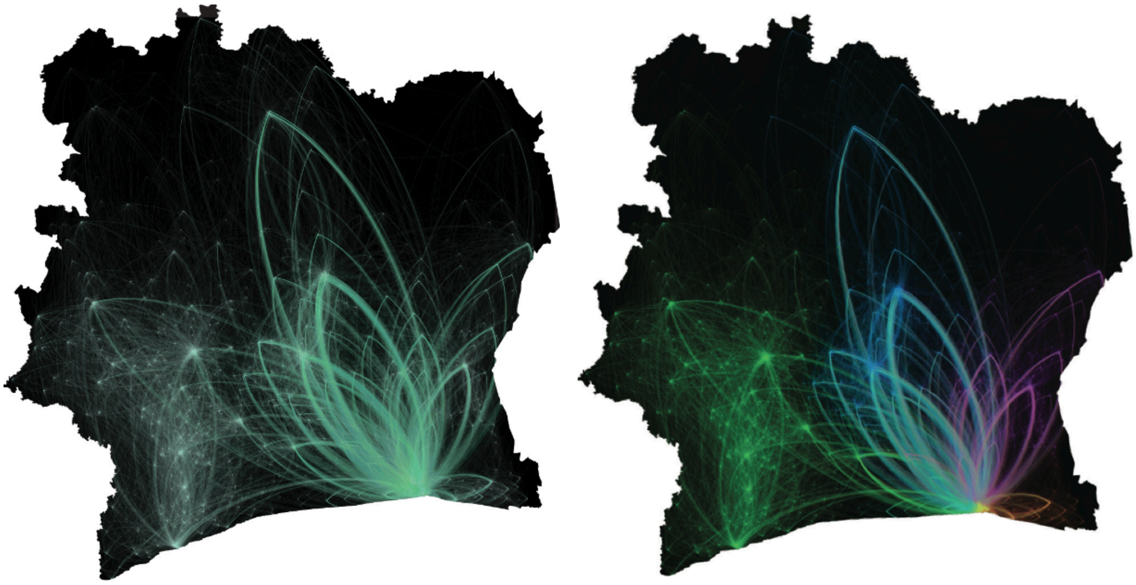
Aim: to inspect the spatial dispersion of values of the variable "Sum of Minutes of Call" against the null hypothesis of spatial randomness.

In order to operate the calculation of these two metrics a first step consisted in spatially joining the layer containing the antennae values with a vector grid (of 500 x 500 meters) extended over Abidjan's city boundaries. The resulting grid thus represented the basis for the calculation of the Weight Matrix, whose values have been weighted according to a rook contiguity criterion. The resulting cells containing hot spots of highly clustered values have served to point out the antennae, which presented the most striking spatial similarity. The respective temporal similarities have also been investigated for each antenna. The last phase of the analysis has therefore consisted in the identification of temporal patterns for a few subsets of highly spatially clustered antennae, identified in the previous steps, mainly the antennae with identifier 27,91 and 436. Their data have been imported in the R environment and aggregated on an hourly base over a week period, intended to be representative of an "average" week. Each of the three antennae has been further decomposed to analyse its spectral signal into its Trend, Seasonal and Random component, of which their autocorrelation has been tested. In the words of (Aguiar et al. 2010), the Temporal Autocorrelation of a signal measures the influence of past observations on the present observations, over the entire time interval of the same series, by shifting the signal spectrum of a certain time lag. It is conceptually similar to Temporal Cross-Correlation, in which the observed profiles belong though to different processes. Each combination of the three antennae has been tested for Temporal Cross-Correlation. Lastly, an attempt to compare the two dimensions, mainly the spatial and the temporal, has been made by calculating the correlation between a matrix of the geographic distance between each possible combination of communicating antennae within Abidjan, with the respective Temporal Cross-Correlation statistic. Given the 293 antennas present in the city the resulting matrix had 293 observations in the rows (spatial dimension) for 293 columns (temporal dimension), for a total of 85849 combinations.

4. Results

The analytical spectrum of techniques employed by the thesis has been specifically conceived to be of such width in order to fully encompass the data variance in its multiple dimensions and under every aspect, considering also the un-conventionality of the data origin. Particularly interesting has been the application of the techniques traditionally employed in Social Network Analysis to the graph derived from the mobile traffic activity logs. Already from the

visualization step is in fact possible to qualitatively interpret some dynamics affecting the country at large (see below Figure 2a).



(a) SET1 Communication Network (b) SET1 Network Partitioned

Figure 2: Network Visualization and Partitioning

4.1 Modularity Maximization Algorithm results

The first striking result is a stark separation of the calls patterns across the country. As correctly pointed out by (Mao, Shuai, Ahn, & Bollen, 2013) this division is certainly partially biased by the different density in the antennae's coverage. But, apart from such difference, it is clear that also the traffic intensity varies greatly among the regions. In fact, even if from the picture seems that apparently the Southeast (a region not entirely deprived of antennae) exhibits a certain level of interactivity, this is certainly not comparable in terms of intensity with that exhibited from the Southwest. This is to remark, as already found by (Liu, Murata, & Wakita, 2013), that regional variance exists even within the more developed part of the country (as will be analysed shortly afterwards).

Further, (Andris & Bettencourt, 2013) correctly underlines the outstanding influence of the economic capital Abidjan, and of its role in attracting the majority of communication fluxes. While is in fact true that the highest concentration of activities and people is located there (Jost & Michon, 2009), the conclusion that they make regarding the urban hierarchy, defined as "still incipient" is worth of further considerations.

From the analysis performed via the MM algorithm as well as through the SNA metrics, it seems in fact that apart from Abidjan's metropolitan area, which acts as the gravity pole of the country, and therefore indeed exhibits the characteristics of a morphologically polycentric region, the same doesn't apply for the smaller and sparser urban agglomerations scattered

throughout the inland. The emergence of six Modularity Classes mirrors six major areas of interactions in the country, which remarkably coincide with some obviously distinguishable regions, in terms of their economic structure and functional dependence.

The First Modularity Class for instance (green sub-graph in Figure 2b) entails the larger part of the Southwest: a region which presents two major cities of regional importance, mainly Daloa in the land-locked department of Haupt-Sassandra, and the port-city of San Pedro (the only other port in addition to Abidjan's to possess a deep-sea harbour). The structure identified by the algorithm correctly presents a web of inter-city edges, linking preferentially the two main centres, but also them with further smaller nodes of local relevance, systematically positioned around the two (like Gagnoa, Divo, Man..). Such spatial arrangement underlines what is perhaps the heritage of the Growth Pole expansion pursued in the 60s (Jedwab, 2010), and in addition underlines also the farness of these areas from the main economic core of the South-East.

On the contrary, the third Class (purple sub graph in Figure 2b) presents an extremely high spatial dependence with the economic capital. It is the graph that includes the regions of Moyien-Comoe and Agneby, with the cities of Agboville and Abengorou, ranked 11th and 14th in terms of city population in 1998 (Denis & Moriconi-Ebrard, 2009). The eastern regions are those that historically present a higher concentration of cocoa plantations, due to the closeness with neighbouring Ghana. The agronomic linkages connecting those fertile regions with the port of Abidjan's and its manufacturing centres can thus explain this high dependence.

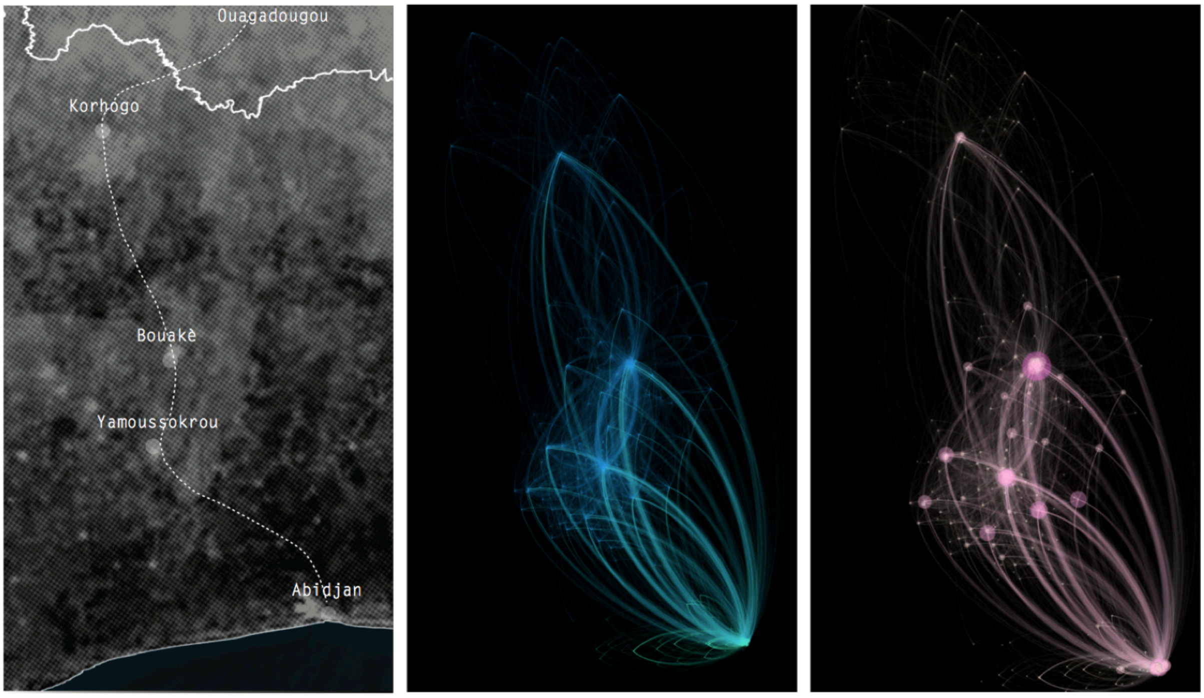


Figure 3: North-South Interaction Vectors

Another interesting pattern is the one linking the two capitals (see Figure 3 above). The blue sub-graph (in the middle in Fig.3) is in fact showing the intensity of traffic between the two economic hubs of the country, mainly Abidjan and the political capital Yamoussoukro. Moreover, what's truly fascinating is that the regions crossed by those links are also those with the highest infrastructural presence. They are in fact crossed by the only railway line of the country. Hence, one partial conclusion might be that the possibility of a physical connection actually strengthens the intensity of communication. Of course also the reverse might be true. In any case the resulting figure is far from what described by (Friedman, 2005), that is, a pointlessness of physical proximity and interaction in the presence of digital connectivity. Rather, the opposite seems in fact to apply. Explaining on the contrary why the only departments, which are one another within physical reach, due to a relatively well-connected transportation, network, exhibit instead such an intensity of communication, would in fact be quite a case.

The classes from four to six are instead less representative of the true urban structure beneath them. They all in fact entail different parts of the larger Abidjan's metropolitan area, roughly distinguishable in an eastern, central and western sub-graphs (respectively the light-blue, yellow and orange sub-graphs in Figure 2b). The reduced representativeness of each of these sub-graphs depends on the fact that these areas lie on the trajectories of the incoming fluxes of phone traffic directed towards Abidjan. In the network topology this translates directly into a higher probability of having a shared node between that part of the sub-graph and the core of the city (an area inside the yellow sub-graph is in fact Abidjan's CBD, Le Plateau). The bias is therefore introduced by a significantly higher percentage of detected links within such parts of the country in respect to the others. Nonetheless, each area can still be thought of being representative of the connection between Abidjan with the centres of Dabou (to the East) and to the former capital of Grand-Bassam (to the West), the two biggest cities in the surroundings. The central yellow sub-graph hence corresponds roughly the city of Abidjan itself.

The emergence of different degrees of interaction between the urban centres is thus clearly visible between but also within each modularity class. The case of the first modularity class is exemplifying, though also at smaller scales the same results can still be found. This has been the case for the abovementioned Abidjan metropolitan area, but is also true for the city of Abidjan itself, clearly partitionable in smaller sub-graphs according to the MM algorithm.

The results seem convincing. The city is in fact divided into four main areas, roughly corresponding to the northern *commune* of Adjamè, to Youpugon to the East, to Plateau and Cocody respectively in the centre and to the West, and to the southern island of Grand-Bassam. But caution should be exercised as this result might in fact be due to a combination of the transitivity property of the algorithm, sensible to the reduced graph size (Blondel et al. 2008), with the particularly network topology in the area, especially dense due to the high

number of antennae. As anticipated by (Mao et al., 2013), the city presents a skewed spatial distribution correlated to the neighbourhood wealth. Of all 396 towers in fact, almost a fourth belong to the rich Cocody, while only a twentieth to the poorer commune of Adjamè.

4.2 Social Network Analysis results

In addition to the qualitative assessment of the network properties, similar results have also been found applying more quantitative methods of Social Network Analysis' tradition to the SET1 graph. In particular the Node Metrics portray once again a structure in which the nodes that stand out are mainly those of regional importance, though declined differently in accordance to their specific role. As such, similarly to what (Andris et al. 2013) found, the city of Yamoussoukro, one of the largest and theoretically also one of the most important cities in the country, presents a smaller ratio of outgoing over incoming calls. This is a sign of a lower position in terms of centrality in the economic and administrative life of the country.

	Weighted		Not Weighted	
Resolution	Num. Classes	Modularity	Num. Classes	Modularity
0.3	50	0.46	124	0.216
0.5	45	0.507	64	0.298
0.7	40	0.528	38	0.350
1	39	0.534	33	0.379
1.2	34	0.537	32	0.364

Table 1: Mod. Max. Algorithm Test Results

The Network Metrics' overview (see Table 1 below) further confirms what presented so far: Modularity Class (Md) 1, encompassing the largest area, correctly presents the highest value of Average Path Length while at the same time being the sparsest, with the lowest Density value of 0.168. Exactly the opposite of Abidjan's Md-4, which, in addition to being the densest (with a value of 0.662) has also the highest Average Degree and the lowest Average Path Length. All signs of an environment which is not only characterized by a high concentration of activities, but that presents also a remarkably thick and short network of connections, denoting a high vitality and diversity of activities. It is in fact the Md class with the lowest Modularity score, which remarks its extreme compactness and density. As expected, and for the reasons stated before, also the Md classes nearby Abidjan, Md-4 and Md-6, present similar values between themselves and in respect to Md-4. The present though some differences: Md-4, located eastern of Abidjan, and roughly corresponding to the city of Dabou, presents a denser pattern of interaction with the core city if compared to Md-5. Its lower Modularity can therefore depend on it being a region dominated mostly by one centre, but might also be a sign of its functional dependence to Abidjan. In addition, its closer

location to the main economic pole of the country, if compared to Md-6 main cities, might account for a shorter Average Path Length. On the other hand Md-6 is formed by a greater variety of smaller centres ranked on lower positions of the urban hierarchy. The higher value of Modularity can thus be explained in this sense. Lastly, Md-2 and Md-3 present the most similar values of all classes and across all metrics. This is partially dependent on the structure of the network, since both classes present in fact a topology characterized by an alternation of long and short edges, representing the main connections of the respective regional hubs with Abidjan as well as interactions with local nodes, coherently with their economic structure which is partially dependent on the Ivorian main city.

4.3 Power-Law Distribution Test results

Apart from the results presented so far, dealing with the existence of a certain degree of functional poly-centricity, another path undertaken to assess the degree of mutual interaction between an within urban spaces, has been the test for the presence of a power-law in the distribution of the call frequency of each sub-graph resulted from the MM algorithm (see Table 2 and Figure 4): Of the two estimated distribution parameters, x_{min} and α, the latter is the most interesting in relation to the thesis ‘aim. This scaling coefficient assumes ranges from 1 (in case of a perfect urban hierarchy) to larger values (in case of high urban heterogeneity). According to the estimated values, every area in the country (illustrated by its Md class) should therefore present a rich mixture of nodes characterized by a high diversity of behaviour in relation to the frequency of call.

Modularity Class	Scaling Coefficient	xmin
1	1,898172	426
2	1,872458	681
3	1,868046	462
4	2,572523	3511
5	1,957543	677
6	3,326649	13734

Table 2: Power-Law Parameters Estimation

In practice though, from a successive validity test on the estimated parameters, performed through a bootstrap procedure of over a thousand iteration for each Md, the result show that only Md3 truly presents a power-law distribution, at a confidence interval of 5%. Therefore probably the frequency of contacts within a particular Modularity Class will be better approximated by another distribution, like the lognormal, as hypothesised by (Clauset et al., 2004). Not much can be said in this regard at the current state of the analysis.

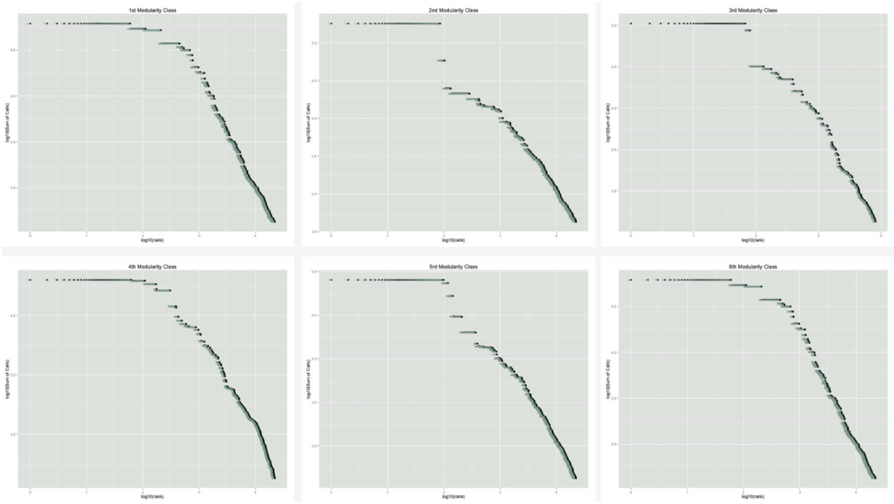


Figure 4: Call Frequency Power Law Distribution for each Modularity Class

Similarly, also the Spatio-Temporal analysis had not brought particularly interesting results. The LISA methodologies permitted to identify clusters of highly spatially auto-correlated antennae, whose temporal behaviour has been further analysed. While a clear variation in the daily presence of people and use of mobile phones is present, few assumptions can be inferred regarding the causes at the root of such behaviour. In fact, while recurring trends are commonly found both in the auto-correlated as well as cross-correlated signals, the difficulty in deriving a meaningful interpretation is a major drawback. The most that can be said, in particular in respect to the temporal similarity of spatially close antennae is that they present a low correlation whose period is in average equal to 4 or 8 hours. But to fully unfold the process that generated such behaviour a more in depth analysis is required.

The last step of the analysis involved the test for similarity in respect to both spatial as well as temporal behaviour of the antennae. To reach this aim, a correlation matrix between each couple of antennae's distance and their cross-correlation has been tested. As a result there's clearly no sign of any correlation among any matched pairs of antennae's. This un-satisfying result might depend on a series of factors, among which the most important might have been the inappropriate temporal scale. The high temporal variance expressed in the autocorrelation and cross-correlation plots of the previous step, suggests in fact the need to test the correlation of geographic distance against a smaller temporal interval, to better control the variability of the latter. This might be an interesting goal to pursue in the future.

5. Discussion

To conclude, to the research questions presented in the opening paragraphs are hereafter re-proposed, together with a short summary of the main results obtained:

I. *To what extent does ICT play a role in effectively overcoming geographic distance?* As already established by many, among which (Morgan 2004; McCann 2008; Tranos et al. 2012) there is certainly a definite degree of truth in the Death of Distance/Flat World framework of (Cairncross, 2001; Friedman, 2005). Coming from the same perspective, the current paper aims though to shed further light on the extent in which such hypothesis work. Information and Communication Technologies can in fact be though as acting in some respect as "flattening forces", since they manage to somehow bring closer distant location, and connect far away people. But that's not the whole story. They in fact also operate in the opposite sense, favouring the concentration of not standardizable information in a few spatially defined amount of places, where the exchange of such information is facilitated by a set of condition locally present (Sassen, 2010). This can be seen, for example, in the extraordinary primacy of the Ivorian economic capital, Abidjan, not so much in respect to population size, but rather according to its centrality from a functional point of view. The massive weight of this node in the national network is clearly undeniable, as well as its influential position of centrality and its role in attracting fluxes of goods, people and information. This behaviour can also be found among cities positioned in the lower ranks of the hierarchy, as well as in more peripheral areas, where the same centripetal action is still present to some extent. As the majority of communication fluxes tends to converge towards regional and provincial hubs. The gradual dualism of ICT's effects on human interaction over geographic distance is therefore once again confirmed.

II. *What is the emerging picture of urban realities in the Ivory Coast, and in the use of their spaces and places, as portrayed by User-Generated mobile phone data?* In accordance to the findings of (Mao et al., 2013), the Communication Network derived from SET1 portrays an outstandingly diversified sequence of characteristic spaces, which morph completely into new forms if looked at from different spatial scales and different points of views, as well as according to the different temporal rhythms of urban life (Jensen, 2009). The great variety of cultures, ethnicities, languages, economies, and their mutual interaction, is fully reflected in the data. Once again, the introductions of User Generated data analysis in the field of Urban Research have widened the disciplinary field to the understanding of urban life in new, fine-grained dimensions. The convergence of different analytical methodologies have proved to be a successful path towards a greater understanding of urbanity at large, in particular due to the solid methodologies offered by the scientific fields of spatial statistics and network analysis.

References

Aguiar, M. J. D., & Janssen, M. (2010). *Spatial and Temporal Analysis in Ecology : A Primer* (p. 101). Guelph, CA.

Amini, A., Kung, K., Kang, C., Sobolevsky, S., & Ratti, C. (2013). The Differing Tribal and Infrastructural Influences on Mobility in Developing and Industrialized Regions. In V. D. Blondel, N. de Cordes, A. Decuyper, P. Deville, J. Raguenez, & Z. Smoreda (Eds.), *Mobile Phone Data for Development* (pp. 330–340). NetMob Conference 2013.

Andris, C., & Bettencourt, L. M. A. (2013). Development, Information and Social Connectivity in Cote d'Ivoire. In V. D. Blondel, N. de Cordes, A. Decuyper, P. Deville, J. Raguenez, & Z. Smoreda (Eds.), *Mobile Phone Data for Development* (pp. 55–68).

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, *27*(2), 93–115.

Bates, S. (1995). *The End of Geography symposium: theories and metaphors of cyberspace*. Retrieved April 17, 2013, fromhttp://pespmc1.vub.ac.be/Cybspasy/SBates.html

Batty, M. (2012). Smart cities, big data. *Environment and Planning B: Planning and Design*, *39*(2), 191–193.

Blondel, V. D., de Cordes, N., Decuyper, A., Deville, P., Raguenez, J., & Smoreda, Z. (2013). Mobile Phone Data for Development - Analysis of mobile phone datasets for the development of Ivory Coast. In V. D. Blondel, N. de Cordes, A. Decuyper, P. Deville, J. Raguenez, & Z. Smoreda (Eds.), *Mobile Phone Data for Development* (pp. 7–16). NetMob Conference 2013.

Blondel, V. D., Esch, M., Chan, C., & Clerot, F. (2012). Data for Development: the D4D Challenge on Mobile Phone Data. *CoRR*, 1–10.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: The Differing*, *2008*(10), P10008. Physics and Society; Statistical Mechanics; Computers and Society; Data Structures and Algorithms.

Cairncross, F. (2001). *The death of distance: how the communication revolution is changing our lives*. *Harvard Business School Press* (p. 303). Boston, MA (USA): Harvard Business Press.

Chen, Y., & Zhou, Y. (2004). Multi-fractal measures of city-size distributions based on the three-parameter Zipf model. *Chaos, Solitons & Fractals*, *22*(4), 793–805.

Clauset, A., Newman, M., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, *70*(6), 066111. Statistical Mechanics; Disordered Systems and Neural Networks.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM Review*, *51*(4), 661–703.

Denis, E., & Moriconi-Ebrard, F. (2009). *Africapolis - COUNTRY FACTSHEET* (p. 213). Paris, FR.

Dinh, T. N., & Thai, M. T. (2013). Community Detection in Scale-Free Networks: Approximation Algorithms for Maximizing Modularity. *IEEE Journal on Selected Areas in Communications*, *31*(6), 997–1006.

Dittmar, J. (2009). Cities, Institutions, and Growth: The Emergence of Zipf's Law. *Manuscript, University of California, Berkeley*, 1–51. Retrieved from http://iga.ucdavis.edu/Research/all-uc-spring-09-papers/dittmar paper.PDF

Friedman, T. (2005). The World is Flat: A Brief History of the Twenty-First Century. New York: Farrar, Straus and Giroux (1st editio., p. 488). New York, USA: Farrar, Straus and Giroux.

Gabaix, X. (1999). Zipf's Law and the Growth of Cities. The American Economic Review, Vol.89(2), 129–132. Retrieved from http://www.jstor.org/stable/10.2307/117093

Gabaix, X., & Ioannides, Y. (2004). The evolution of city size distributions. *Handbook of Regional and Urban Economics*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1574008004800105

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 7821–6.

Jedwab, R. (2010). *Urbanization With Growth in Africa : On Cocoa and Cities in Ivory Coast and Ghana* (pp. 1–35). London, UK.

Jensen, O. B. (2009). Flows of Meaning, Cultures of Movements – Urban Mobility as Meaningful Everyday Life Practice. *Mobilities*, *4*(1), 139–158.

Jost, N., & Michon, P. (2009). *Abidjan en mouvement. Etude de la mobilité dans une métropole de l'Ouest Africain. Abidjan (Côte d'Ivoire)*. Ecole Polytechnique Fédérale de Lausanne. Retrieved from http://infoscience.epfl.ch/record/174154

Liu, X., Murata, T., & Wakita, K. (2013). Exploring the Multilevel Community Structure in D4D Datase. In V. D. Blondel, N. de Cordes, A. Decuyper, P. Deville, J. Raguenez, & Z. Smoreda (Eds.), *Mobile Phone Data for Development* (p. 849). NetMob Conference 2013.

Losada, J. C., Benito, R. M., Morales, A. J., & Creixell, W. (2013). Understanding ethnical interactions on Ivory Coast. In V. D. Blondel, N. de Cordes, A. Decuyper, P. Deville, J. Raguenez, & Z. Smoreda (Eds.), *Mobile Phone Data for Development* (pp. 115–120). NetMob Conference 2013.

Manfredini, F., Pucci, P., Secchi, P., Tagliolato, P., Vantini, S., & Vitelli, V. (2012). *Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the Milan urban region*. *MOX–Report*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract

Mao, H., Shuai, X., Ahn, Y., & Bollen, J. (2013). Mobile Communications Reveal the Regional Economy in Cote d'Ivoire. In V. D. Blondel, N. de Cordes, A. Decuyper, P. Deville, J. Raguenez, & Z. Smoreda (Eds.), *Mobile Phone Data for Development* (pp. 17–35). NetMob Conference 2013.

McCann, P. (2008). Globalization and economic geography: the world is curved, not flat. *Cambridge Journal of Regions, Economy and Society*, *1*(3), 351–370.

Morgan, K. (2004). The exaggerated death of geography: learning, proximity and territorial innovation systems. *Journal of Economic Geography*, *4*(1), 3–21.

Nelson, M. (2013). The big picture on big data. *Inter Media*, *41*(1), 2013.

Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323–351.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(23), 8577–82.

Nitsch, V. (2005). Zipf zipped. *Journal of Urban Economics*, *57*(1), 86–100.

O'Connor, B. (2009). *Zipf's law and world city populations*. Retrieved May 06, 2013, from http://brenocon.com/

Reggiani, A., & Nijkamp, P. (2012). *Did Zipf Anticipate Socio-Economic Spatial Networks?* (No. 816) (p. 30). Bologna, ITA.

Rodriguez-Pose, a., & Crescenzi, R. (2008). Mountains in a flat world: why proximity still matters for the location of economic activity. *Cambridge Journal of Regions, Economy and Society*, *1*(3), 371–388.

Sassen, S. (2010). Urban Economies and Fading Distances. In S. Buijs, W. Tan, & D. Tunas (Eds.), *Megacities: Exploring a Sustainable Future* (I., p. 379). Rotterdam, NL: OIO Publishers.

Sutherland, E. (2011). Mobile telecom in Africa : unforeseen success ? *Inter Media*, *39*(2), 26–31.

Tranos, E., & Nijkamp, P. (2012). The Death of Distance Revisited: Cyberplace, Physical and Relational Proximities. *Tinbergen Institute Discussion Paper*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2103024

United Nations. (2012). *Big Data for Development: Challenges and Opportunities*. (Global Pulse, Ed.)*Naciones Unidas, Nueva York, mayo* (p. 47). New York, USA: United Nation Organization - Global Pulse Initiative.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (p. 573). Cambridge, Mass. USA: Addison-Wesley Press.