

# D4D Challenge

## Commuting Dynamics 4 Change

R. Lario <sup>\*</sup>M. Muñoz <sup>†</sup>R. Abad <sup>‡</sup>J. Gonzalez <sup>§</sup>A. Martín <sup>¶</sup>R. Maestre <sup>||</sup>

Paradigma Labs Research Group  
Paradigma Tecnológico

E. Perez <sup>\*</sup>I. del Bosque <sup>††</sup>  
Geographic Information Systems Unit  
Spanish National Research Council

4-Jan-2013

### Abstract

Our idea is to use the geolocation data from the antennas processing the mobile phones calls in order to know which sub-prefectures the customers have been getting around. The main goal of our project is developing spatio-temporal models to detect commuting patterns for the different sub-prefectures, including some other factors related to the region and/or time: wealth, development, infrastructure, investment, grants...

By means of GIS technology, we will be able to apply our generated models to the gathered data and to analyze their correlations over the Côte d'Ivoire surface, working with geographical layers: landcover, roads map, railways lines, water sources... Consequently, the reached conclusions from our study will be properly visualized, allowing a better explanation of the findings. With a bigger amount of data gathered for a longer period, more interesting and accurate trends could be discovered, allowing us to calculate associated coefficients.

Our analysis models will provide coherent data to support a correct urban design and will mean a monitoring tool for development, specially related to population dynamics. In the near future, some other measures could be included. For instance, hospitals and police stations locations, their calls rate... Thus, we could know its real use, being able to improve their service to the citizens: dangerous areas, crowded hospitals...

---

<sup>\*</sup>rlario@paradigmatecnologico.com

<sup>†</sup>mmunoz@paradigmatecnologico.com

<sup>‡</sup>rabad@paradigmatecnologico.com

<sup>§</sup>jgonzalez@paradigmatecnologico.com

<sup>¶</sup>amartin@paradigmatecnologico.com

<sup>||</sup>rmaestre@paradigmatecnologico.com

<sup>\*\*</sup>eperez@cchs.csic.es

<sup>††</sup>idelbosque@cchs.csic.es

## List of Figures

1	Theoretical Commuting Model . . . . .	7
2	Dynamic and static user patterns . . . . .	8
3	Total amount of calls grouped by Week days . . . . .	11
4	Dynamic users displacements . . . . .	13

## About us and why face this challenge

Paradigma Labs Research Group<sup>1</sup> (PLRG) core values are conducted by one motivation: "To figure out the fuzzy dynamics between Humanity and Technology", providing tools and methods to study, display and understand these dynamics. Therefore, an international challenge whose research subject can be chosen freely as long as it relates to an objective of development improving quality of life for people, quickly held our attention.

Geographical Information Systems unit<sup>2</sup> at Spanish National Research Council<sup>3</sup> (GIS-CSIC) is a multidisciplinary group with a huge experience in Remote Sensing and Geoprocessing, providing a quality support for plenty of researches carried out at CSIC.

Our final aim has been detecting geospatio-temporal patterns in order to obtain an useful knowledge to better manage the country resources. For example, if we could predict the traffic intensity segmented by road, week day and hour, then another secondary roads could be suggested or the budget for the most used ones could be increased. Through mobile communications, a specific user can be tracked along the day, not only by means of 'call communications' but also thanks to applications running on their handsets: IMS, RSS... The dataset provided by Orange is a sample, and it only uses 'call communications', however, with the whole set of data (i.e.: app and call communications), we strongly believe more accurate and complete models could be discovered helping to identify new kinds of dynamics.

From our own experience studying and modeling several kinds of Human Dynamics like during the ESF project DynCoopNet(Solana and Alonso, 2012) and while developing a Business Intelligence Tracking Tool on Twitter (Marin et al., 2012), we can claim there are two main exploring perspectives: the Geographical one and Temporal one. We believe a mathematical model related to Human Dynamics must be managed with these two viewpoints. The Temporal component is useful by providing a tool to go backward and forward in order to get a more detailed understanding of the dynamic, not only moving across the timeline, but also creating temporal windows to group events. The Geographical component provides a more high-level understanding related to the human mobility across the space in different levels and relating it to some other spatial features. Mixing both components in a final and single visualization has led our study during the project.

Consequently, in this paper, we propose a Geospatio-Temporal Model. A Geospatial Model, because user interactions with geolocated antennas are analyzed and treated, and a Temporal Model since several time windows are used to group these user dynamics. The combination of these two variables is used and displayed by a GIS. Initially, several results are showed supporting the project main conclusions. However, what's really important is the whole process for handling the data, that is, the code, tools and methodology, which will be available to the researcher community, allowing to study more deeply the dynamics. For instance, a Standard Kernel Density estimation (KDE) aims to produce a smooth density surface of spatial point events over a 2-D geographic space(Bithell, 1990; Alegria et al., 2011), final dynamics visualization across the several days of the week will show by means of KDE, in order to understand and proof which an where are the maximum commuting peaks.

We have focused on the Commuting concept, which could be defined as follows: *Commuting* is regular travel between one's place of residence and place of work or full-time study (Wikipedia, 2012), but sometimes it refers to *any regular or often repeated traveling between locations when not work related*.

---

<sup>1</sup><http://labs.paradigmatecnologico.com/>

<sup>2</sup><http://humanidades.cchs.csic.es/cchs/sig/>

<sup>3</sup><http://www.csic.es/>

Our first commuting approach is defined like: "Mobility patterns through inferring dynamic users movements grouped by temporal windows".

A *commuter* or *dynamic user* is defined as an user changing his antenna location within the studied temporal window (i.e.: each temporal window groups the whole user communication during a specific hour). Among these temporal windows, *non-commuters* or *static users* have been removed, i.e.: users who do not change their antennas locations within the temporal range. The justification to remove these users comes to focus our study on users that are moving into this temporal windows and perform micro-displacements. It is common that a same user performs these two kind of dynamics within the same temporal window. Note that we are not quantifying the distance, but only the fact of changing from a particular antenna to another one.

## State-of-the-art

Nowadays, the world has nearly as many cell phone subscriptions as inhabitants<sup>4</sup>. For the first time, the majority of humanity is linked and has a voice. Consequently, plenty of phone communications are being generated continuously everywhere, and, what is more relevant, they are being tracked: geolocation, start/end times... This is the key, mobile phone companies record data which are very closely associated with behaviour of people.

Analyzing these data in a proper way discloses a great deal of social knowledge (behaviour modeling, mobility patterns and trends, outliers discovery) which can be applied in countless and different areas<sup>5</sup>: transportation, urban planning, commuting, tourism, traffic congestion, demography, economy, advertising and commerce, public health... Even without Internet connections (e-mail, IMS and so on), that is, focusing only on speech-calls and text messages, there is a vast amount of information which can be 'read' to reach further conclusions. The ability to understand the patterns of human life by analyzing the digital traces that we leave behind will transform the world, specially poor nations. Reality mining of behavior data is just beginning.

Let's enumerate a few examples of current projects related to behavioural data

As can be seen, communication data are everywhere (*we are social animals!!*) and they can be used to obtain really interesting and high-value findings. Imagine, once we know the nature and meaning of these data, it is as if we had access to a lot of complete, reliable and immediate surveys. Honestly, we strongly believe that the future lies in knowing how to process this kind of data to get unique results. MIT's Technology Review has recently identified **reality mining on mobile communications** as one of '10 Emerging Technologies That Will Change the World'.

Note how the D4D datasets (there are only 4 and contain really simple data) have caused many and varied studies from all teams. As far as we are concerned, we discussed about several ideas: antennas network optimization in traffic terms, geospatial-temporal detection of real use for public services (hospitals, schools, police stations...), commuting patterns detection and the like.

Precisely, it has been the human urban mobility approach the one we chose as the core of our project. Why ?? Because it is a reality very tied to ordinary people daily lives, so that its study can reveal clues to improve people quality of life.

Here below a few current researches showing how identified commuting patterns are really useful to understand human motion dynamics better and to perform accurate plans and actions:

- a) Exploring spatio-temporal commuting patterns in a university environment allows making more appropriate decisions to decrease the automobile dependence of students, promoting the non-motorized and public transportation. It is a green initiative looking for sustainability: reducing pollution and noise, avoiding congestion, improving public health and urban planning...
- b) Classifying different urban areas based on their mobility patterns from mobile phone data. The results can be used to better understand this dynamic allowing more efficient environmental and transportation policies for the time being and for the future (since due to the regularity of the individual trajectories, it can be claimed that human mobility is highly predictive).

---

<sup>4</sup>[http://www.huffingtonpost.com/2012/10/11/cell-phones-world-subscribers-six-billion\\_n1957173.html](http://www.huffingtonpost.com/2012/10/11/cell-phones-world-subscribers-six-billion_n1957173.html)

<sup>5</sup><http://www.insead.edu/v1/gitr/wef/main/fullreport/files/Chap1/1.6.pdf>

- c) Time patterns and geospatial clustering based on mobile phone network data provide accurate statistics about mobility of people, population density and economic activity with detailed regional and time resolution.
- d) Visual analytics system to study people's mobility patterns from mobile phone data.

## Problem description & Hypothesis

We believe that we can better manage the country resources through the knowledge extracted from data communications, in this paper the solution is built on the hypothesis of mobility patterns to predict common and well-known, geographical and time based patterns to manage roads and infrastructures in a correlated way with the results figured out.

The figure below shows a theoretical commuting model proposed like a main pattern.

Two peaks are modeled,  $p_1$  in the range  $[07 : 00, 8 : 00]$  and  $p_2$  in  $[17 : 00, 18 : 00]$ . This first approach to modelize this dynamic set up the two peaks  $p_1, p_2$  with the same weight, however, numerical results will show that the weight of each peak depends of the day of the week. A central valley is defined between  $[09 : 00, 16 : 00]$ , with a uniform displacement distribution.

An ad-hoc mathematical model is defined in the next section in order to confirm this hypothesis; focusing, filtering and processing the main data to contrast the hypothesis.

The main idea behind the two main peaks in the model,  $p_1, p_2$  and the central valley, is that people perform great displacement distances early in the morning i.e.:  $p_1$  related with the common business activity. After the first peak  $p_1$ , people resides in this target destinations, working, eating, etc ..., but in a more static point of view and always performing displacements. The last point in the hypothesis approach showed by Figure 1 is in the second peak  $p_2$ , when people return to his destination or the last business activities are realized.

The geographical behaviour of the dynamic, always mixed with the temporal component, will be contrasted by means of GIS tools in order to visualize the expansion and contraction in the main points that hypothesis shows:  $p_1, p_2$  and the central valley. Expansion when maximum displacement is reached on the first peak  $p_1$  and contraction, but not quite, when the central valley is reached. Another displacement expansion when peak  $p_2$  is reached and its corresponding contraction when among  $p_2$  is declining.

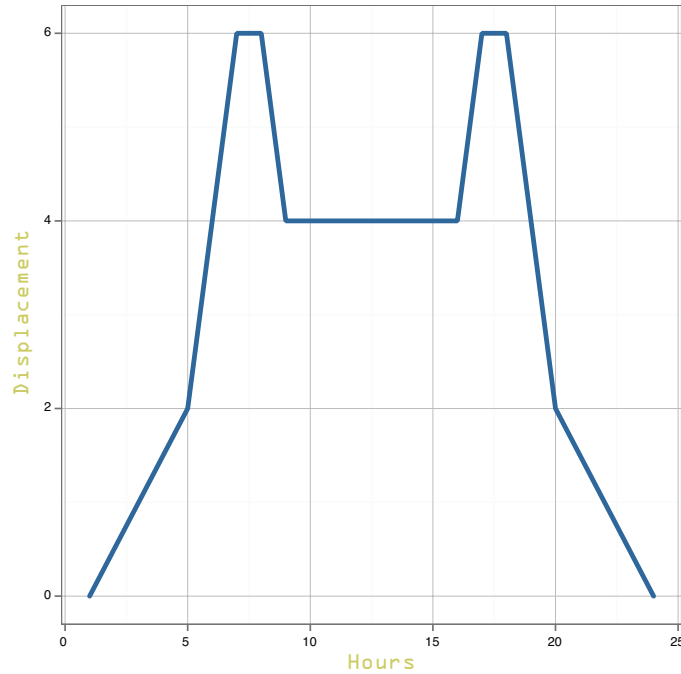


Figure 1: Theoretical Commuting Model

## Mathematical model

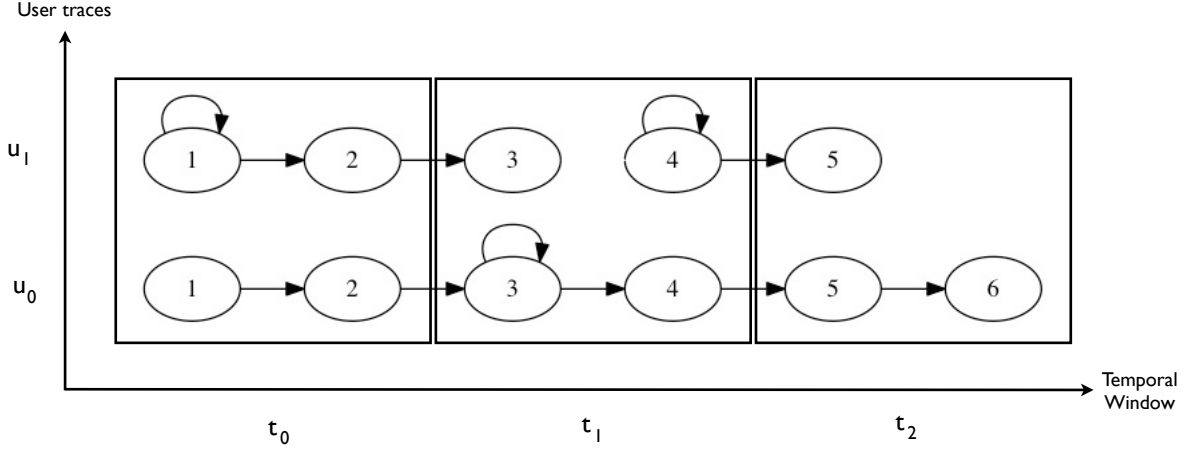


Figure 2: Dynamic and static user patterns

The above figure shows two commuters  $u_0, u_1$  represented on the vertical axe, grouped by three time windows  $t_0, t_1, t_2$ . A time window is defined as  $t_n$  where  $n \in \{0, \dots, 24\}$ . Each  $t_n$  groups the communications traces of the whole set of commuters in a 60 minutes range.

Formally, a commuter trace during a particular time window is defined as follows:

$$\vec{T}_{ut} = (p_0, p_1, \dots, p_n)$$

where,  $p_n \in \mathbb{R}^2$ ,  $n > 1$ ,  $t$  is a temporal window range and,  $u$  the unique commuter id. For instance,  $\vec{T}_{00} = (p_1, p_2)$ ,  $\vec{T}_{10} = (p_1, p_1, p_2)$  and so on.

Also, two functions are defined in order to measure the distance (Cook, 2012), given a set of points in spherical coordinates (i.e.:  $\text{user}_{ut}$ ):

$$D(p_0, p_1) = \text{acos}(\sin(\phi(p_0^0)) * \sin(\phi(p_1^0)) * \cos(\theta(p_0^1) - \theta(p_1^1)) + \cos(\phi(p_0^0)) * \cos(\phi(p_1^0)))$$

where:

$$\phi(x) = (90 - x) * \frac{\pi}{180}$$

$$\theta(x) = x * \frac{\pi}{180}$$

therefore the function related to the distance is defined as follow [result in Km]:

$$U(u, t) = 6373 * \sum_{i=0}^{n-1} D(\vec{T}_{ut}^i, \vec{T}_{ut}^{i+1})$$

The second function is related to the number of antenna connections into a trace, the key point is to count only the dynamic transitions, i.e.: remove the self edges over a given trace as follows:

$$S(p_0, p_1) = \begin{cases} 0 & \text{if } D(p_0, p_1) = 0 \\ 1 & \text{if } D(p_0, p_1) > 0 \end{cases}$$



therefore, the function is defined as follows:

$$N(u, t) = \sum_{i=0}^{n-1} S(\vec{T}_{ut}^i, \vec{T}_{ut}^{i+1})$$

## Methodology

## Results

According to the proposed model, a very important feature has been deduced for the Commuting Dynamic. As we can see in the picture below, there is a couple of time zones when people perform more phone calls from their handsets than usual. Static and dynamic users have not been distinguished, that is, both self-edges and transition edges are counted together.

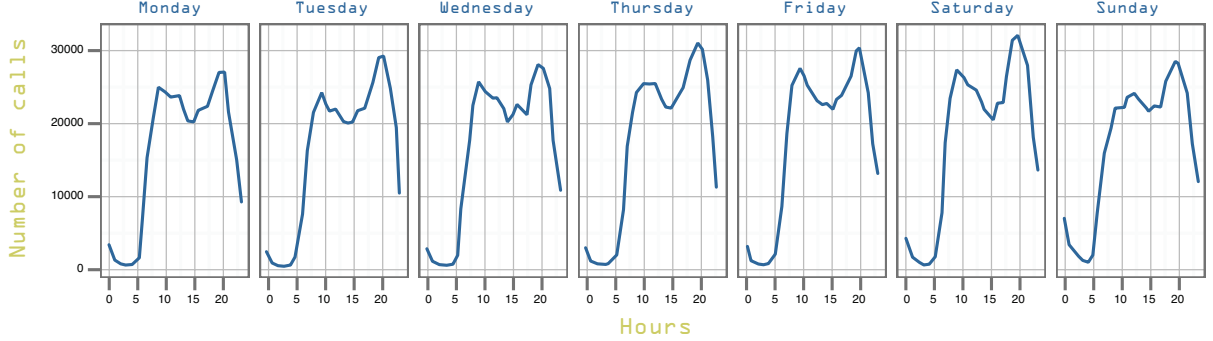


Figure 3: Total amount of calls grouped by Week days

There are two peaks,  $p_1$ , from 7:00 am to 9:00 am, and another one,  $p_2$ , from 6:00 pm to 8:00 pm. As can be seen, the 'morning-peak' is lower than the 'evening-peak'. Moreover, both peaks reach their maximum value at the end of their corresponding time zone, i.e.,  $p_1$  at 9:30 am and  $p_2$  at 7:30 pm. Note how the amount of calls increases linearly since 5 am and how it decreases linearly too since 8:00 pm. Another remarked feature is the existence of a central valley between  $p_1$  and  $p_2$ , that is, from 9:00 am to 6:00 pm. This 'peak-valley-peak' pattern is shown for all days in the week, so that we can assume people have the same behaviour (at least, according to our target datasets).

Let's extract some ideas from the previous chart, focusing on what commuters could be probably doing (as we will see shortly, most of the phone calls belong to 'dynamic users'): they get up early in the morning and start to perform more and more phone calls from their handsets until  $p_1$ ; next, the amount of phone calls decreases a bit, keeping itself more or less balanced (workhours, lunch) until the beginning of  $p_2$ ; later, they people leave the office and plan the rest of the day (errands, leisure...), what is reflected on a marked rise in the amount of phone calls.

Since it is really complicated to measure the displacements of the people (antenna locations instead of users locations, missing antenna identifiers...), what will be analyzed is the amount of callers who are really commuters, that is, dynamic users. This one will be the essential magnitude of our research, leading us to figure out the Commuting Dynamics key-features for different regions and times. Eventually, all conclusions deduced from the G/T Model (charts, formulae...) will be ultimately tested with the final GIS visualization as the main core of the work. In this last phase, geographical displacements are estimated so that they can be plotted in a dynamic and interactive map which makes easier to detect peaks, trends...

With this new chart, it is the ratio between commuters and all users what it is being emphasized. During a particular day (24h), there is no need to know if a concrete displacement is or not longer than other, no, what it is being highlighted here is the movement detection itself, for 1h windows which groups all users calling within it. The colour of the chart shows how many users are really commuters.

Datasets have been processed according to the proposed G/T Model, filtering non-commuters when commuters tracks have been calculated. Here below, seven charts display Commuting Dynamics for each week-day. As the results shows, there is not correlation between the number of dynamic users and the maximum displacement peaks, actually, at the same time the number of dynamic users are growing, the central valley and the two maximum commuting peaks are always presented in the sample.

These seven charts below shows a fitter correlation with the theoretical commuting model proposed in this paper (Figure 1), showing two high peaks and a lower central valley. However more qualitative data is necessary to figure out the performance of the first high peak, because is not related with the number of dynamic user, therefore a first approach can be applied to say that few dynamic users (in comparison with the mean) travel long distances in this first peak, specially in the first uphill to the first peak. This is a common pattern figure out for the proposed model. However, because there are people that start his travel from a far distances early in the morning since the main people lives near the work places and travel less distances than the first one.

falta cruzar lo que se acaba de decir con las visualizaciones del kernel density, y listo more or less

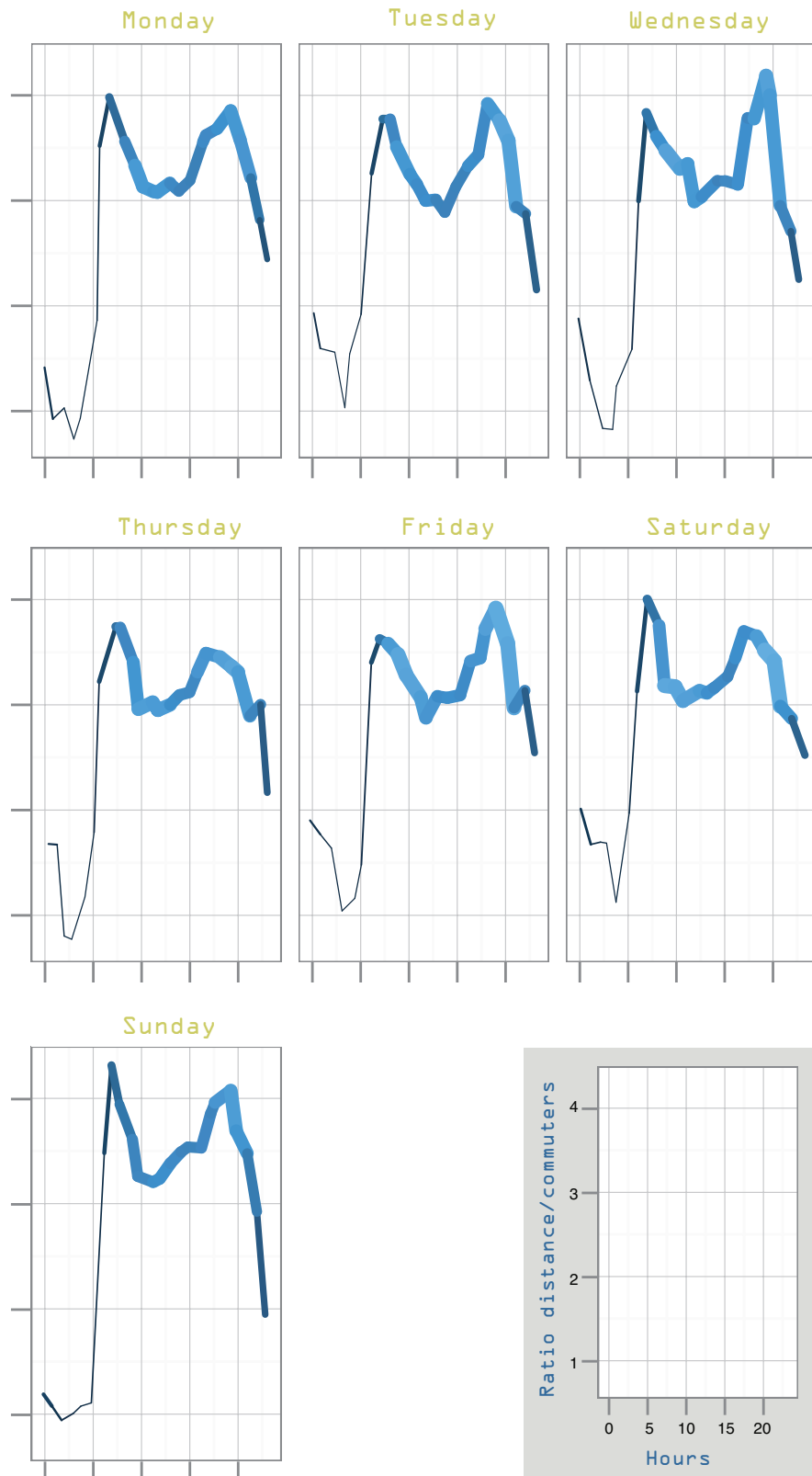


Figure 4: Dynamic users displacements

## Final Conclusions

Summing up the methods and tools used in this novel study, a mathematical model to shape Space and Temporal user-antenna communications has been proposed, two kind of user dynamics have been identified in order to carry out and focus the study on the commuting dynamics. A commuting pattern model is showed to test with the Orange sample dataset of user communications.

A temporal set of windows defined in 60 minutes ranges and grouped by hours and day of the week has been proposed in the model, however, this assumption could be changed in order to explore the data with another point of view.

Dynamic users has been splitted from statis users to detect commuting patterns, an the mathematical model proposed and applied has shown two maximum displacement peaks and a central valley following the main component of total call numbers, normalizing the maximum displacement using the median with the total number of dynamic users.

A visualization of the displacement with a Kernel Density estimation by means of GIS using the antennas positions and users displacements like a direct graph has shown the contraction and the expansion of the commuting dynamics across the 24 hours of the day. A tool to visualize parallel this dynamic across the seven days of the week is release on line<sup>6</sup>.

Poner

## Future work

We strongly believe that working with more detailed data i.e.: data from applications and other kind of communications, researchers will be able to work on more accurate mathematical models and therefore, more accurate and useful visualizations. GIS tools are itself a method to research on geolocated data. Several algorithms and methods could be applied to figure out new dynamics and to return the profit of these results into a new develop policies in order to improve the human condition.

---

<sup>6</sup>"Poner la dirección de las herramientas"

## References

- A.C. Alegria, H. Sahli, and E. Zimanyi. Application of density analysis for landmine risk mapping. In *Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2011 IEEE International Conference on*, pages 223–228, 29 2011-july 1 2011. doi: 10.1109/ICSDM.2011.5969036.
- J. F. Bithell. An application of density estimation to geographical epidemiology. *Statistics in Medicine*, 9(6):691–701, 1990. ISSN 1097-0258. doi: 10.1002/sim.4780090616. URL <http://dx.doi.org/10.1002/sim.4780090616>.
- John D. Cook. Computing the distance between two locations on earth from coordinates, 2012. URL [http://www.johndcook.com/python\\_longitude\\_latitude.html](http://www.johndcook.com/python_longitude_latitude.html). [Online; accessed 4-January-2013].
- Oscar Marin, Alejandro Gonzalez, Roberto Maestre, Julio Gonzalez, Marco Martinez, Ruben Abad, and Leonardo Menezes. 15th october on twitter global revolution mapped, 2012. URL <http://labs.paradigmatecnologico.com/2011/12/19/15th-october-on-twitter-global-revolution-mapped/>. [Online; accessed 23-January-2013].
- Ana Solana and David Alonso. Self-organizing networks and gis tools cases of use for the study of trading cooperation (1400-1800). *Journal of Knowledge Management, Economics and Information Technology*, page pp. 402, 2012. ISSN 2069-5934. URL <http://www.scientificpapers.org/special-issue-june-2012/>.
- Wikipedia. Commuting — wikipedia, the free encyclopedia, 2012. URL <http://en.wikipedia.org/w/index.php?title=Commuting&oldid=525136196>. [Online; accessed 4-January-2013].