D4D Challenge
Commuting Dynamics 4 Change

R. Lario[*] M. Muñoz[†] R. Abad [‡] A. Martín[§] J. Gonzalez [¶] R. Maestre [‖]
Paradigma Labs Research Group
Paradigma Tecnológico

E. Perez [**] I. del Bosque [††]
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 facts. 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...

[*] rlario@paradigmatecnologico.com
[†] mmunoz@paradigmatecnologico.com
[‡] rabad@paradigmatecnologico.com
[§] amartin@paradigmatecnologico.com
[¶] jgonzalez@paradigmatecnologico.com
[‖] rmaestre@paradigmatecnologico.com
[**] eperez@cchs.csic.es
[††] idelbosque@cchs.csic.es

1 Introduction and motivation

The Paradigma Labs Research Group (PLRG) core values are conducted by one motivation: "Figure out the dynamic and fuzzy intersection between Humanity and Thechnology", providing tools and methods to study, display and undeestand this dynamics. Therefore a Challenge which main goal is the human improvement and development by means of the study of mobile communications quickly captured our attention. Geographical Unformation Systems Unit at Spanish National Research Council (CSIC) is a multidisciplinary group with a huge experiencie in Remote Sensing and Geoprocessing providing support to the high and long scale researchs at CSIC. To reach the goal, our researchers and engineers are mixed efforts to achive the goal proposed in this paper.

Our final aim is detecting spatio-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. Trough mobile communitacions, a specific user can be tracked along the day, not only by means of "Call comunications" also by applications running on the user terminal like chats, RSS consumers, etc ... The dataset provides by Orange is a sampled one, and only apply "Call communications", however, we belive that with thw whole set of data i.e.: app and call comunications future models could be several kind of dynamics more accurate.

Our experience studing and modeling several kinds of Human Dynamics like ESF project DynCoopNet(Solana and Alonso, 2012) or Business Intelligence Tracking Tool on Twitter (Marin et al., 2012), has teach us to explore through two main point of view: Space and Temporal perspectves. We belive that a mathematical model related with Human Dynamics must be present this two components. The Temporal component is useful to provides to the final researcerh a tool to go forward and backward in order to get more deeply understand of the dynamic, not only move across the timeline, also creating temporal windows to group events. The Geographical component provides a more high-level understand related with the human mobility across the space, mixing the time component along the space in a final visualization. Therefore our model and the sudy will carry out by this two main components.

In this paper, we propose a Space-Temporal Model to achieve this goal. A Space Model, because user interactions with geolocated antennas are analyzed and treated, and a Temporal Model because several time windows are used to group this user dynamics. Later, the mix of this two variables (space and temporal) are consumed and visualized by a Geographic Information System. At the begining several results will be showed, however, the tools with the data will be available to the research comunity to study more deeply the studied 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.

Our aim is related to the Commuting concept and could be defined as follow: *Commuting* is regular travel between one's place of residence and place of work or full-time study (Wikipedia, 2012), but sometimes its refers to *any regular or often repeated traveling between locations when not work related.* Our first commuting approach is defined like: "Mobility patterns through inferring dynamic users movements grouped by temporal windows".
A dynamic user is defined like an user changing his antenna location within the studied temporal window (i.e.: each temporal window groups the whole user communication into one specific hour). Among these temporal windows, static users have been removed, i.e.: users that 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.

2 Mathematical model

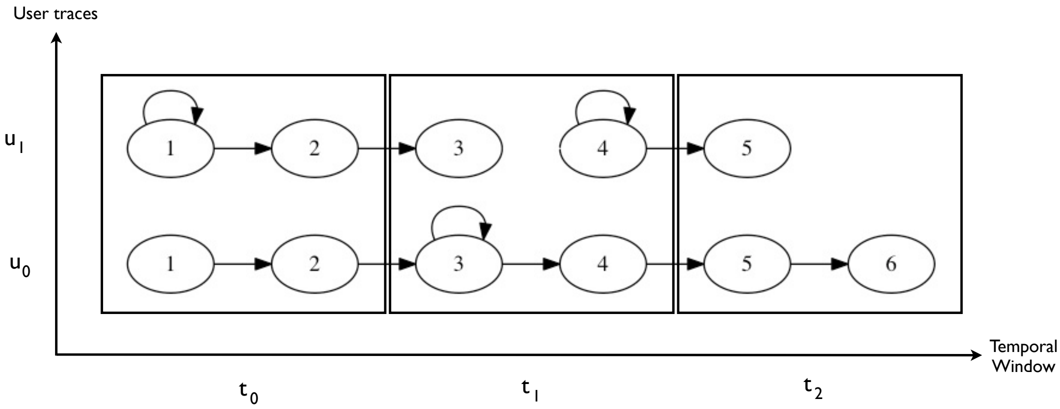


Figure 1: Dynamic and static user patterns

The above figure shows two users u_0, u_1 repesented on the vertical axe, grouped by three time windows t_o, t_1, t_2. A time window is defined as t_n where $n \in \{0, .., 24\}$. Each t_n groups the comunnications traces of the whole set of users in a 60 minutes range.
More formally, an user trace is defined as follows:

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

where, $p_n \in \mathbb{R}^2$, $n \geq 2$, t is a temporal window range and , u the unique user 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.: 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 with distance as 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 with the number of antenna connections into a trace, the key point is to count only the dinamic transitions, i.e.: remove the self edges over a given trace as follows:

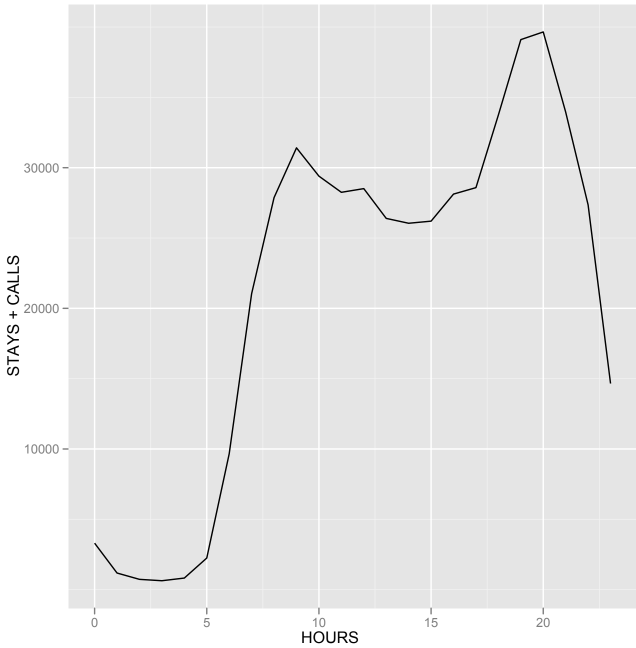
$$\text{S}(p_0, p_1) = \begin{cases} 0 & \text{if } D(p_0, p_1) = 0 \\ 1 & \text{if } D(p_0, p_1) > 0 \end{cases}$$

therefore, the function is defined as follows:

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

Temporal and geographical components

Using the proposed model, the first step is to understand the main component of the commuting dynamic. The result of our approach will emerge from this main component as we can see in the next picture. This main component model the first logic result and shows the time segments that people have more activite, in call numbers terms. This picture shows static and dinamic users calling i.e.: nodes with self-edges are counted and transition edges together.



(a) Total number of calls grouped by Week days

As the above picture shows, we can see two peaks p_1 from 7 Am to 9 Am and another one p_2 from 6 Pm to 8 Pm. As we can see $p_1 < p_2$ in a number of call terms and p_2 start to grow at 6 PM and get its maximum value at 8 PM, after this hour, the number of calls dicrease linearly. Another main thing showed by the picture is that exists a central valley between p_1 and p_2, i.e.: from 9 Am to 6 Pm, that people stop or "relax??" the number of calls. This main component is common to the seven days of the week, therefore, we can assume that people have the same behavior on this sample dataset.

Several assumptions could be suggested at this point, for instance, in the p_1 range, people start to work on a regular daily activite, in the central valley, some people stop (but not entirely), perhaps to eat, and in the next peak p_2, people return to regular daily activite and perhaps start to return home. The question here is to figure out the commuting dynamic through the average displacement per person and where are located the time ranges or windows that people realize this action. This last assumption, will be tested with the proposed model and the final GIS visualization as main core of our work. Geographical displacements are studied by means of GIS tool on the map in order to detect and understand the micro and long movements across the country.

Start point here is show our theoreticall model of commuting Poner algunas referencia or something del supuesto modelo teórico!!. As we can see two peaks are defined corresponding with maximum displacement hours and a central valley corresponding to "eat hours" queda poco científico, hay que poner algo mejor

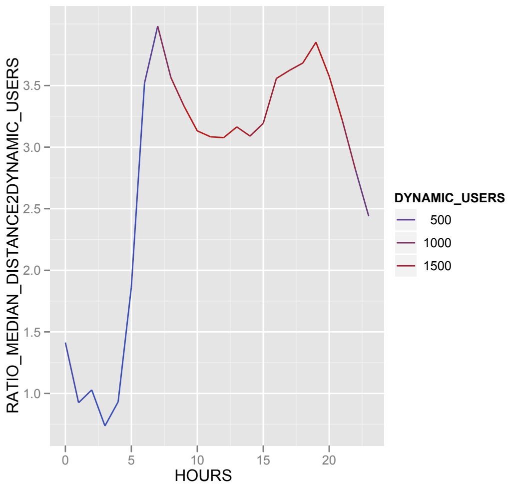
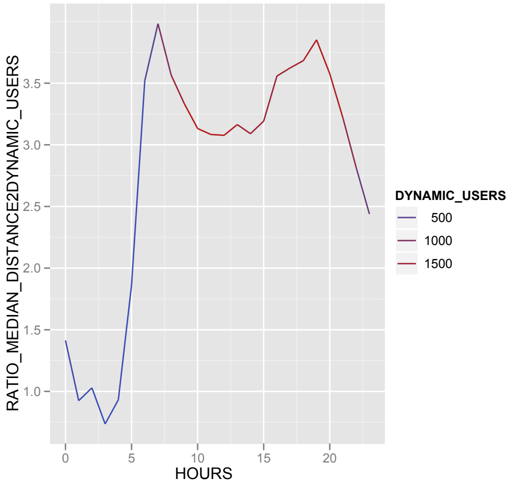


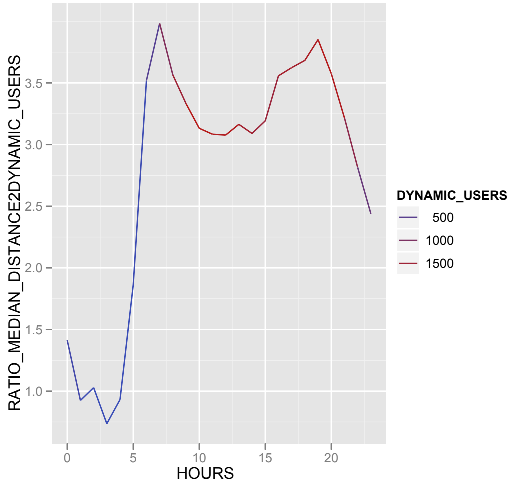
Figure 2: Theoretical Commuting Model

Datasets has been processed carried out by the mathematical model proposed, therefor, static users has been removed and dinamyc users, better its tracks, has been grouped into a temporal windows and grouped by the day of the week. Seven pictures are shown below in and their respective data table. The normalized median with the number of dynamic users is showed in each temporal window of each picture corresponding with each day of the week.

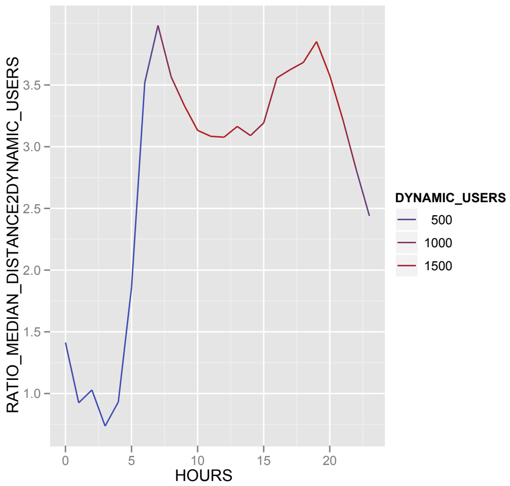
Pictures shows that emerge a new p_1' and p_2' where $p_1' > p_2'$, in a displacements terms, we can see that people start to perform moves about 4 AM until to reach p_1' at 7 Am. At this point, displacements are reduced



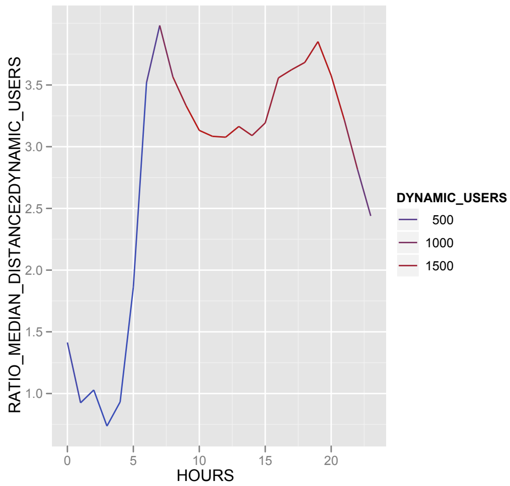
(a) Monday



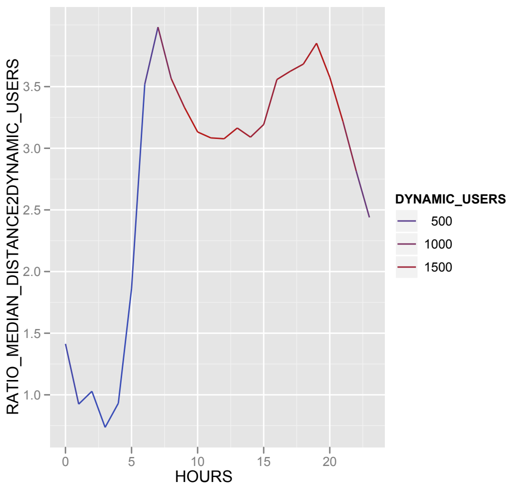
(b) Thuesday



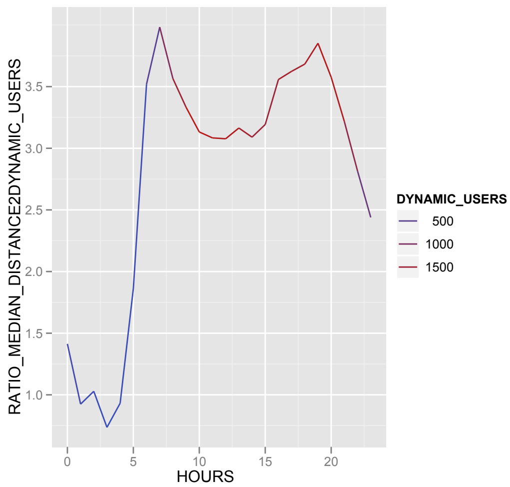
(c) Wednesday



(d) Thursday



(e) Friday



(f) Saturday

Figure 3: Dynamic users displacment

Representation and visualization

Pequeño estado del arte de los gis ... y porque usarlos.
Poner los enfoques por los que hemos pasado: los puntos de colores por intensaida de paso, la grilla y finalmente, acabar hablando del kernel density

Final Conclusions

Poner las conclusiones finales

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. [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].