# Zonification of Heavy Traffic in Mexico City

**Ruben F. Estrada-S.[1], Alejandro Molina[2], Adriana Perez-Espinosa[3]. Araceli L. Reyes-C.[1], Jose L. Quiroz-F.[3], and Emilio Bravo-G.[1]**
[1]Science and Technology College, Autonomous University of Mexico City, Mexico City
[2]The National Commission for Knowledge and Use of Biodiversity, Mexico City
[3]Electrical Engineering Department, Metropolitan Autonomous University, Mexico City

**Abstract—** *Mexico City has the most elevated traffic congestion in the world according to a recent report carried out by TomTom navigation products. A driver in Mexico City could spend up to 59% extra time trapped in traffic during a normal day. This paper applies a cluster analysis to find out the heavy traffic zones using massive data from the Social Navigation Network Waze. This zonification is fundamental for urban planning and transportation management, and may help develop accurate prediction models for heavy traffic.*

**Keywords:** Geo-spatial Data Analysis, Clustering, Vehicular Traffic, Data Mining

## 1. Introduction

The rapid urban growth of Mexico City has made it one of the most crowded in the world with a population close to twenty million people. As a result, the Metropolitan area is in constant expansion, forcing thousands of people to use their vehicles daily. Thus, this year the levels of air pollutants have exceeded the maximum exposure limits established by the World Health Organization (WHO) [1]. The heavy traffic in Mexico City causes a daily loss of 3.3 million man hours in traffic, which costs 33 billion pesos according to a study of The Mexican Institute for Competitiveness (IMC)[2].

An important characteristic of developing cities is the heavy traffic zones, where main roads converge and generate a large amount of traffic. The detection of such traffic zones allows seriously analyzing causes and proposals for realistic solutions. In the past, this problem has largely been approached by specialists with knowledge of the city. Nowadays, information technologies enable the detection of heavy traffic zones using massive and real time data. We refer to sensor, video and GPS devices [3].

The increasing use of smart devices with a GPS has promoted the development of different social navigation networks such as: Waze [4], Google Maps [5] and Inrix Traffic [6], etc. In Mexico City, the most popular of these networks is Waze, which collects data from user devices and allows them to report events and traffic levels, generating a real-time snapshot of the traffic situation.

This paper analyzes heavy traffic zones in Mexico City using data from Waze. In the light of Data Mining algorithms we propose a zonification of the metropolitan area based on massive quantities of user-reports. Section 2, shows an overview of the subject. In Section 3 the Waze network data is described. Results are presented in Section 4. Finally, the conclusion and future work are highlighted in Section 5.

## 2. Related Work

The problem of traffic congestion is a serious one in many cities the developing countries. In [7], V. Jain, A. Sharma and L. Subramanian presented an automated image processing mechanism for detecting the congestion levels in road traffic by processing CCTV camera image feeds. The algorithm is designed for noisy traffic feeds with poor image quality, based on live CCTV camera feeds from multiple traffic signals in Kenya and Brazil. Their analysis shows evidence of congestion collapse which lasts for long time-periods across multiple locations, allowing the detection of critical congestion zones. Furthermore, they gave causes of poor traffic management e.g. unplanned cities, poor discipline, alternate traffic means, archaic management and tighter budgets.

As for traffic studies in Mexico City in [8], the analysis of the traffic flow on the Mexico-City Cuernavaca highway is presented. They developed a traffic simulation system based on cellular automata with variable anticipation for single-lane traffic flow. Simulation results show that the congestion observed on days-off on this highway is not only a consequence of its complex topology. The main problem lies in the fact that there is an inefficient system of toll.

F.A. Armah,D.O. Yawson and A. Pappoe [9] applied system dynamics to identify related drivers, causes and effects of traffic problems in the Accra city, Ghana. Based on the analysis, they proposed policies, mainly economic instruments, to reduce traffic congestion in that city.

On the other hand, in [10] develops a new density-based algorithm to identify congested traffic routes, named FlowScan; instead of clustering the moving objects, road segments are clustered based on the density of the traffic they share in common.

## 3. Waze: A Social Navigation Network

Waze [4] is a noticeable social navigation network used by over 50 million users worldwide, which provide to its

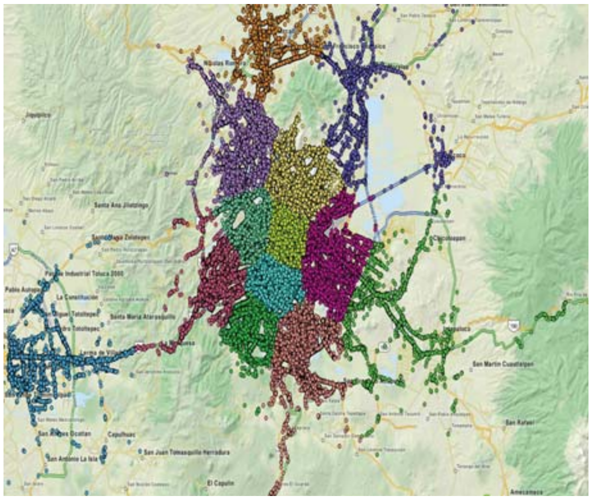

Fig. 1: A million datasets of heavy traffic reports were classified and projected using Lloyd's algorithm.

users real-time traffic information collected through a Waze application installed in smart devices with built-in GPS such as tablets or phones. Moreover, Waze application allows users to report traffic jams, accidents, speed and police traps. This information is processed to estimate the level of traffic on the roads, the time that a user will spend in traffic congestion and suggest alternative roads and the time estimated to arrive at a destination point. Waze has five level users: Baby, Grown-Up, Warrior, Knight and Royalty. As a user advances to high levels its reports have greater influence on the routing. In this work, we are interested in the traffic alerts, which can be of three type: moderate, heavy, and standstill.

# 4. Experiments and Results

In our study, we gathered the Waze public information using a web application which collects reports submitted by users. Our application focused on events in the Metropolitan Zone of Mexico City. The study area is bounded by the geographic coordinates North: 20.075, South=18.785, East=-98.582 and West=-99.659. More than a million reports were used to carry out the experiments. The reports have been collected every ten minutes since November 2015 but, as we want to model heavy traffic scenarios, we have selected the data when the number of traffic reports was greater than 12500 at the same time in the study area.

Figure 1 shows data points classified and projected, after a run of Lloyd's algorithm [11] to find a partition into well-shaped clusters. The initialization spans from $k = 2$ to $k = 20$ clusters after selecting several random sets of distinct points as the initial centers. Clustering results were used to define convex cells (hulls) that suggest a first approach for the zonification of the Metropolitan area of Mexico City in terms of real heavy traffic reports (see Figure 3).

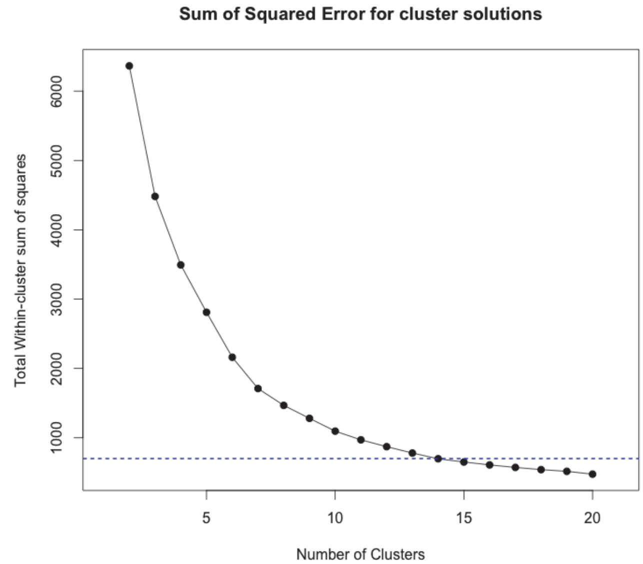


Fig. 2: Sum of squared error for different cluster solutions.

However, two crucial questions immediately arise. First, what is the appropriate number of polygons? Second, how could we adapt our zonification to the existent political frontiers in the area? In the following sections these questions are discussed.

## 4.1 Clustering Evaluation

One common method of choosing the appropriate cluster solution is to compare the sum of squared error (SSE) for a number of cluster solutions. This is the sum of the squared distance between each member of a cluster and its cluster centroid. We have used SSE as a global measure of error. Figure 2 shows the SSE against a series of sequential cluster solutions. This useful graphical information allows us to choose an appropriate number of clusters. An appropriate cluster solution could be defined as the way at which the reduction in SSE slows dramatically. This produces an "elbow" in the plot of SSE against cluster solutions. In Figure 2, such "elbow" is found at the $k = 14$ cluster solution suggesting that $k > 14$ does not have a substantial impact on the total SSE.

## 4.2 Zonification of Vehicular Traffic in the Metropolitan area of Mexico City

Using official maps information from the National Institute of Statistics and Geography (INEGI[1]) we have assigned each municipality to its corresponding traffic zone (Figure 4). When more than one zone was assigned to the same

[1] *Áreas Geoestadísticas Municipales*, 2012, scale: 1:250000. INEGI. Online: http://www.conabio.gob.mx/informacion/metadata/gis/muni_2012gw.xml?&_xsl=/db/metadata/xsl/fgdc_html.xsl

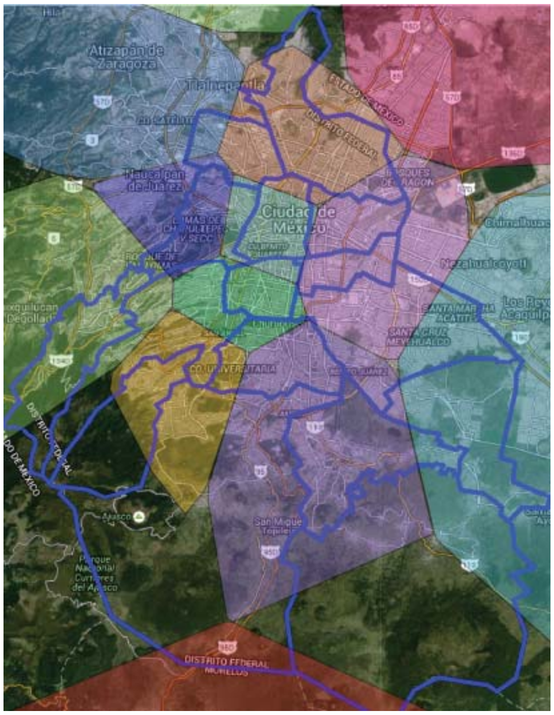


Fig. 3: The convex hull for each heavy traffic zone in the Metropolitan area of Mexico City.

| k | $\rho_{k'}$ | $\rho_k$ | $\rho_{kk'}$ |
|---|---|---|---|
| 1 | 45.138 | 194.372 | 4.306 |
| 2 | 13.361 | 12.159 | 0.910 |
| 3 | 14.919 | 49.063 | 3.289 |
| 4 | 1.837 | 14.263 | 7.762 |
| 5 | 17.473 | 284.688 | 16.292 |
| 6 | 132.023 | 895.862 | 6.786 |
| 7 | 0.325 | 9.283 | 28.498 |
| 8 | 167.602 | 390.356 | 2.329 |
| 9 | 12.336 | 55.256 | 4.479 |
| 10 | 1.778 | 1.406 | 1.194 |
| 11 | 0.480 | 1.896 | 3.946 |
| 12 | 1.715 | 4.885 | 2.848 |
| 13 | 106.632 | 39.344 | 0.369 |
| 14 | 16.574 | 375.366 | 22.647 |

Table 1: Density of each heavy traffic zone in the Metropolitan area of Mexico City. Column 1 considers data only during the special contingency period and column 2 considers all the data.

municipality we have considered the zone density defined in equation 1.

$$\rho_k = \frac{\text{reports in k}}{\text{area of k } (km^2)} \quad (1)$$

Thus, given a municipality $m$ and its assigned zones $Z_m : \{i \in 1 \dots k\}$, the final assigned zone is the one whose density is greater in the territory of $m$; i.e. $\arg\max_i \rho_i$. Note that the final proposed zonification is a product of official information and crowdsourcing data.

The definition of density (1) allows also to compare automatically the behavior of the same traffic zones in different situations. For example, we have assessed the impact of the public measures taken by the local government due to the poor air quality in recent days. Table 1 reports the density for each zone using all the data $\rho_k$ as well as the density using only the data for the contingency period of time $\rho_{k'}$. The last column in Table 1 corresponds to the relative density between the normal traffic and the special period. We observe that the local government measures have had a great impact on traffic for some zones of the city, while for some others, the heavy traffic situation remains ($\rho_{kk'} \simeq 0$).

## 5. Conclusions and Future Work

Figure 4 presents the final proposed zonification map of vehicular traffic in the Metropolitan area of Mexico city. We have identified fourteen traffic zones in the analysed area.

Zonification is needed because clearly, the alerts of heavy traffic occur practically in all municipalities of the city. Moreover, thousands of heavy traffic reports are generated in almost any place in Mexico City at the same moment during normal days. As a consequence, a traffic model for all the city is too complex. For future work, we plan to use the proposed zonification to elaborate a finer analysis over each zone.

As part of a quality analysis we identify that Gustavo A. Madero, Miguel Hidalgo, Cuauhtemoc, Iztacalco and Benito Juarez municipalities are the most affected by traffic, because their access implies crossing many traffic zones. Data is consistent with the fact that the municipalities with less traffic are M. Contreras, Tlalpan and Milpa Alta, mainly in the south-western or south-eastern areas, since they are rural areas.

We conclude that the availability of traffic reports and official information could be processed with data mining algorithms in order to get insights to deal with such a complex problem as vehicular traffic. Our ambition is to go further and use data to create models of traffic prediction.

## References

[1] Air-pollution-in-Mexico-City http://www.theguardian.com/world/2016/
[2] Viviendas para desarrollar Ciudades http://imco.org.mx/
[3] Honghui Dong, Mingchao Wu, Xiaoqing Ding, Lianyu Chu, Limin Jia, Yong Qin, Xuesong Zhou, Traffic zone division based on big data from mobile phone base stations, Transportation Research Part C: Emerging Technologies, Volume 58, Part B, September 2015, Pages 278-291, ISSN 0968-090X, http://dx.doi.org/10.1016/j.trc.2015.06.007.
[4] Waze https://www.waze.com
[5] Google Maps https://maps.google.com/
[6] Inrix http://www.inrix.com
[7] J. Vipin, A. Sharma, L. Subramanian. "Road traffic congestion in the developing world," ACM DEV 2012: 11:1-11:10.
[8] J.A. del Rio and M.E. Lárraga, . ‘Transient Situations in Traffic Flow: Modelling the Mexico City Cuernavaca Highway,” AIP Conference Proceedings;2005, Vol. 757 Issue 1, p190, April 2005.

Fig. 4: Proposed zonification of Vehicular Traffic in the Metropolitan Area of Mexico city.

[9] Frederick A. Armah, David O. Yawson, and Alex A N M Pappoe. "A systems dynamics approach to explore traffic congestion and air pollution link in the city of Accra, Ghana." Sustainability, 2 (1):252–265, 2010. ISSN 20711050. doi: 10.3390/su2010252.

[10] X. Li, J. Han, J. G. Lee, and H. Gonzalez. ‘Traffic density-based discovery of hot routes in road networks,” In Proceedings of the 10th International conference on Advances in spatial and temporal databases, SSTD’07, pages 441–459, Berlin, Heidelberg, 2007. Springer-Verlag.

[11] Lloyd, S. P. Least squares quantization in PCM. Technical Note, Bell Laboratories. Published in 1982 in IEEE Transactions on Information Theory 28, 128–137.