From Data to Knowledge: City-wide Traffic Flows Analysis and Prediction Using Bing Maps

Anna Izabel J. Tostes
Federal University of Minas Gerais (UFMG)
Belo Horizonte, Brazil
annatostes@gmail.com

Fátima de L. P. Duarte-Figueiredo
Pontifical Catholic University of Minas Gerais (PUCMG)
Belo Horizonte, Brazil
fatimafig@pucminas.br

Renato Assunção
Federal University of Minas Gerais (UFMG)
Belo Horizonte, Brazil
assuncao@dcc.ufmg.br

Juliana Salles
Microsoft Research
Redmond, WA, USA
jsalles@microsoft.com

Antonio A. F. Loureiro
Federal University of Minas Gerais (UFMG)
Belo Horizonte, Brazil
loureiro@dcc.ufmg.br

ABSTRACT

Traffic jam is a common contemporary society issue in urban areas. City-wide traffic modeling, visualization, analysis, and prediction are still challenges in this context. Based on Bing Maps information, this work aims to acquire, aggregate, analyze, visualize, and predict traffic jam. Chicago area was evaluated as case study. The flow intensity (free or congested) was analyzed to allow the identification of phase transitions (shocks in the system). Also, a prediction model was developed based on logistic regression to correct discovery future flow intensities for a target street.

Keywords

Urban Computing, Human Mobility, Traffic Prediction, Analysis, Visualization

1. INTRODUCTION

Distincts companies have proposed technological improvements in vehicles concerning the experience of driver and passengers [1]. However, traffic jam continues to reflect a significant impact in the economy, productivity, and environment in urban areas [3]. In 2011, Chicago has shot up to number one in road congestion, according to the Urban Mobility Report, issued by the Texas Transportation Institute [13]. Beyond the time it normally takes to travel without delays, the national average for traffic delays was 34 hours while commuters in the Chicago area spent an additional 70 hours behind the wheel in 2009, which is just increasing (55 hours of wasted time in 1999, and 18 hours in 1982).

So as to deal with traffic jam, the Intelligent Transportation System (ITS) uses infrastructure sensors to monitor traffic conditions in a vehicle environment. The following are possible ITS services: (i) cooperative monitoring traffic, (ii) assistance to the unmarked crossroads, and (iii) collision prevention. In ITS, cooperative communication systems have the potential to improve traffic safety and efficiency through continuous exchange of information [3]. The cooperative communication between vehicles, which is well known as Vehicular Ad hoc Network (VANET), provides them with alternative routes for vehicles.

In this context, distincts Geographic Information System (GIS) have been designed to capture, store, manipulate, analyze, manage, and present all types of geographical data. Bing Maps[1] and Google Maps[2] are mapping applications on the web that give the public access to huge amounts of geographic data. They have an Application Programming Interface (API) enabling users to create GIS applications. Web mapping offers street map, aerial/satellite imagery, geocoding, searches, routing functionality, and traffic jam.

Furthermore, there are crowdsourcing geodata in projects such as OpenStreetMap[3], which is a collaborative project to create a free editable world map. Although without traffic jam information as well as Bing and Google Maps, there are several ways to download the data from OpenStreetMap to a file for simulators of Urban Mobility such as SUMO (Simulation of Urban MObility)[4]. Thus, junctions, traffic lights, and road network can be a VANET scenario, allowing a more realistic evaluation of VANET protocols. The only missing information is traffic jam, which is available at Bing Maps. As Horvitz and Mitchell said in [5], methods for learning automated driving competencies from data will be crucial in the development of autonomous vehicles that drive without human intervention. First, how to acquire and predict the traffic flow are the matters to build safer cars that employ collision warning and avoidance systems.

This paper aims to acquire, aggregate, analyze, visualize, and predict traffic jam based on Bing Maps information. A methodology has been developed in order to establish a city-

UrbComp'13, August 11-14, 2013, Chicago, Illinois, USA

[1]Available at `www.bing.com/maps`.
[2]Available at `www.google.com/maps`.
[3]Available at `www.openstreetmap.org`.
[4]Available at `sumo.sourceforge.net`.

wide traffic jam modeling, which allows its analysis, visualization, and prediction. After acquiring Bing Maps traffic jam information, map's images were processed so as to establish a Chicago traffic jam database. Chicago has been chosen as a case study. Phase transitions (shocks in the system) were identified, and the probability of traffic jam intensity withal. Then, to correct discovery future flow intensities for a target street, a prediction model has been developed.

This work is organized as follows. Section 2 summarizes the topic of vehicular network and related works. Section 3 explains the Bing Maps web crawler, the image-processing algorithm used to extract its information, and afterwards the collected database. Section 4 presents the analysis and visualization of the data, with the phase transitions. Section 5 describe and analyzes the prediction model and the inference results. Finally, section 7 concludes this paper, pointing finally remarks.

2. VEHICULAR NETWORK AND RELATED WORKS

Vehicular networks (VANETs) are different from Mobile Ad hoc Networks (MANET) in some aspects. VANETs do not have the problem of energy consumption and processing power of MANETs. But the vehicular topology is highly dynamic, which raise some challenges. The main characteristics are the high mobility of nodes, intermittent links and stringent latency requirements. We can define the organization and communication between nodes by three models, as illustrates figure 1: (i) pure ad hoc, (ii) infrastructure and (iii) hybrid network. Model (i) is a V2V (Vehicle-to-Vehicle) or VANET (Vehicle Ad Hoc Network) and it consists in the communication between vehicles. Model (ii) is the V2I (Vehicle-to-Infrastructure) that deals with the communication between vehicles and nearby fixed equipment among the road. This model can increase the network connectivity if it has a suitable amount of static nodes, although the network cost can be increased. Model (iii) combines both models (i) and (ii) [4]. It is clear that vehicular networks have several challenges to its widespread adoption, including: (1) connectivity loss for data transmission and (2) the reduced time for the communication between vehicles. Dealing with feature (2) is a challenge owing to the high mobility of vehicles, the dynamism of the network topology and the scalability.

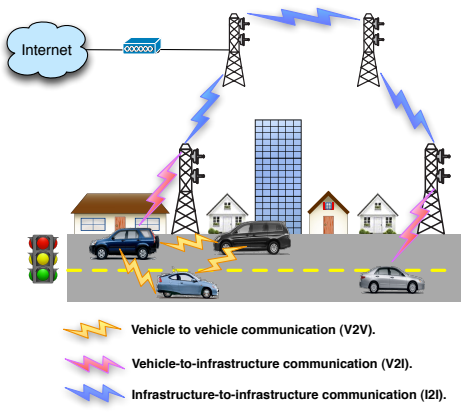


Figure 1: The architecture of vehicular networks.

Distincts routing protocols have been proposed and evaluated for VANETs [8] concerning whether its performance can satisfy the throughput and the delay requirements of safety and emergency applications. Traditional routing protocols, such as the protocols ad hoc on-demand distance vector (AODV) and the dynamic source routing (DSR) [9], can not be used due the route instability that causes packet loss and high overhead. Notwithstanding, geographical routing protocols, e.g. the greedy perimeter stateless routing (GPSR) protocol, have better path stability, but the matter is how to forward the packet through the next hop. To solve this, the main routing protocols use road information to choose the crossroad to forward the packet, but they do not use information of crossroads behavior or traffic flow inference. In this context, we believe that the traffic flow inference can be used to improve its performance.

From Microsoft Research to Bing Maps, refence [6] describe the JamBayes project, started in 2002, which provides estimates of flows inferences about current and future traffic flows. The challenge was to predict the future of traffic flow in Seattle area. When does the highway system would become clogged? The authors developed a probabilistic traffic forecasting system based on Bayes Theory, and predict future surprises about traffic congestion and flow.

Afterwards, following on JamBayes effort, the Microsoft Research project Clearflow focused on applying machine learning to learn how to predict the flows on all street segments of a greater city area [10]. It was based on GPS data collected from volunteers, buses, and vehicles for over five years. Clearflow considers all flows on all roads via predictive models in addition to real-time sensing while directions are provided based on best guesses about flows over all roads. Its main contribution is high coverage of traffic flow for 72 major cities in North America, inferring on over 60 million road segments in North America [10].

Bing Maps is a web mapping platform that can provide business intelligence and data visualization solutions [11]. Bing Maps services can be used to accurately pinpoint locations from geo-coding address (latitude and longitude), base maps and imagery, overlay customer locations, and data analysis. It is similar to GIS systems but without its complexity. This is done using Bing Maps APIs[5], including JavaScript/AJAX or Silverlight Controls. Data from SQL Server or other BI data sources can be easily visualized without the complexity of traditional GIS systems. Also, base maps and imagery can be manipulated. Bing map cloud platform infrastructure is divided in consumer offering and AJAX, Silverlight Control and web services APIs [11].

Urban computing deals with a massive amount of data, gathered by ubiquitous mobile sensors from personal GPS devices to mobile phone. In [12], the authors measure spatiotemporal changes in the population, identifying clusters of locations with similar zoned uses and mobile phone activity patterns. Beyond characterizing human mobility patterns and measuring traffic congestion, reference [2] shows how mobile sensing can reveal details such as intersection performance statistics. None of them use Bing maps information, which is the difference for this study. We have used AJAX/JavaScript APIs so as to visualize traffic layer over Chicago's map. Next section explains our methodology of traffic flow acquisition.

[5]A full overview of Bing Maps API can be found at `www.bingmapsportal.com`.

3. TRAFFIC FLOW ACQUISITION METHODOLOGY

Aiming a city-wide traffic flow information to establish inferences and big data analysis for patterns discovery, a methodology for acquiring traffic flow data from distincts sources was developed. Any GIS map service can be input of this methodology. Figure 2 presents the process of traffic flow acquisition. Through the map service API, a city flow web crawler have been developed. Then, a bash script was designed in order to collect the traffic flow image from the selected city. We used virtualization to print the screen and save it into the database. A image processing software was developed for extracting each road traffic intensity, saving the percentage of green pixels, yellow pixels, and red pixels, which correspond to the flow intensity (green is free while red is congested). Each image from the image database was processed and its flow intensity was saved to a specific date and time into the Traffic Flow Acquisition.

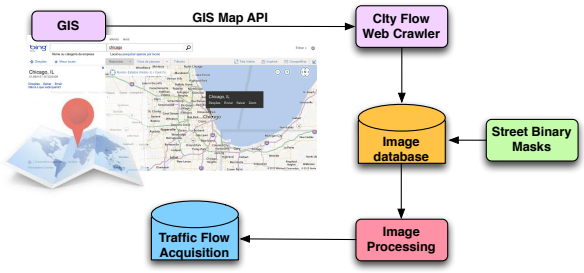


Figure 2: Traffic flow acquisition methodology through a GIS map web crawler.

In this work, Bing Map has been used as input. Algorithm 1 presents the procedure for loading Bing Map in a webpage, with the traffic layer on, and without the illustrations over the map. First the map is set to the specific city center location (geo-code of Chicago is 41.866768, -88.067741), with the specific zoom level (10). Then, the traffic layer is enabled and its opacity is turned off (1). To get a clear traffic image, some HTML tags were removed through Document Object Model (DOM) after 5 seconds so as to ensure the map is loaded with the traffic layer.

Algorithm 1: Bing Map Traffic Module Load

```
setMapView();
showTrafficLayer();
opacityTrafficLayer();
setTimeout(function() {domManager();}, 5);
```

The web crawler algorithm is presented by Algorithm 2. For each HTML scenario established as Algorithm 1, we get current hour, open the web browser with the created HTML file, print the screen, saving as the image name the current hour. This process is repeated every delay seconds.

After collecting image data, the next step is the processing. As input, the algorithm needs the image data and the masks for each road. The mask is a binary image with white background and black street line. Figure 3 illustrates a street mask for one street segment and a translucent image

Algorithm 2: Traffic Flow Web Crawler

```
DELAY = 10 seconds foreach Scenario s do
    get current hour;
    open web browser with the code from Algorithm 1;
    print the screen;
    kill web browser;
    sleep DELAY;
end
```

of Chicago map overlapped, which were used in this work. For such scenario, 100 street masks were manually drawn.

Figure 3: Street mask example and a Translucent Chicago map overlapped.

Algorithm 3: Traffic Flow Image Processing

```
Input: Image file i, Set of Road Masks k_r
GreenPixels = 0;
YellowPixels = 0;
RedPixels = 0;
NoCategoryPixels = 0;
foreach Road Mask k_r do
    foreach Pixel p in the k_r image do
        if p is black then
            // Increase the counter of its respective color
            if hue(p) < 30 or hue(p) ≥ 330 then
                RedPixels++;
            end
            else if hue(p) < 70 then
                YellowPixels++;
            end
            else if hue(p) < 150 then
                GreenPixels++;
            end
            else
                NoCategoryPixels++;
            end
        end
    end
end
```

Algorithm 3 presents the steps followed to process one map image. For each black pixel in the street mask, the counter for each flow category (green, yellow, red, or no category – error) was increased according to its color. To estab-

lish a band for each flow intensity, HSL (Hue, Saturation, Lightness) was used. As HSV (Hue, Saturation, Value), HSL is one of the most common cylindrical-coordinate representations of points in an RGB color model. The variation of the hue corresponds to the values 0–360, in which 0 is a red band, followed by a yellow band, and other band colors, and finally a red band again. So it is possible to identify color bands to Bing's traffic flow intensity.

4. CHICAGO'S TRAFFIC FLOW ANALYSIS AND VISUALIZATION

Chicago area was defined as a study case. The methodology of Flow Acquisition was applied. We collected data from April 10th 2013 to April 24th 2013. Bing map acquisition occurred every 7 minutes. Next, the database has the following information: (i) date; (ii) hour; (iii) street number; (iv) number of green pixels; (v) number of yellow pixels; (vi) number of red pixels; and (vii) number of no category pixels. Its map was divided into geo-code sectors, as Figure 4 illustrates. Each street has its influence area, considering the direction of the street. It is important to notice each street number.

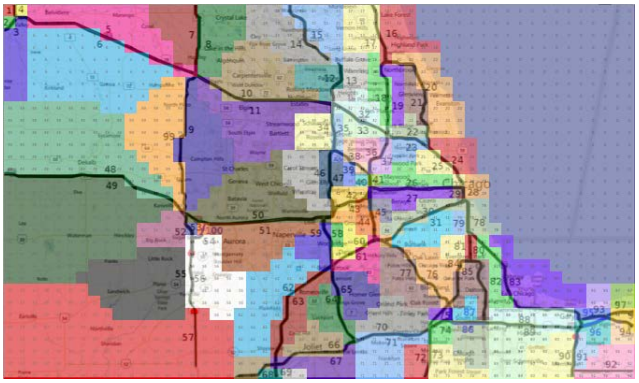


Figure 4: Street sectors over the Bing map of Chicago area.

To analyze the street performance, we mapped as a graph in which a road is a vertex and the edges are the direction that can be taken by a vehicle in that street. Thus, a directed graph is established. In order to present the most important streets, the betweenness centrality has been calculated. It indicates the number of shortest paths from all vertices to all others that pass through that vertex. Figure 5 summarizes the most important roads, which are the downtown streets (red squares in the heat map). The top is street 79 (759), followed by street 28 (731), street 78 (719), street 31 (713) and street 26 (702).

Another graph metric that has been evaluated was the edge betweenness, i.e. the number of shortest paths between pairs of nodes that run along the edge. Such metric corresponds to the importance of the road convergence. Table 1 presents the top-10 roads importance and the convergence importance (edge betweenness). One can notice that the 5 most important roads involve the top-5 vertex betweenness that is vertices 79, 28, 78, 31, and 26. A traffic jam in such roads has more impact on the network availability.

Next analysis will be presented according to the period of the day (dawn, morning, lunch, afternoon, and night), which is indicated by Table 2 with their corresponding times.

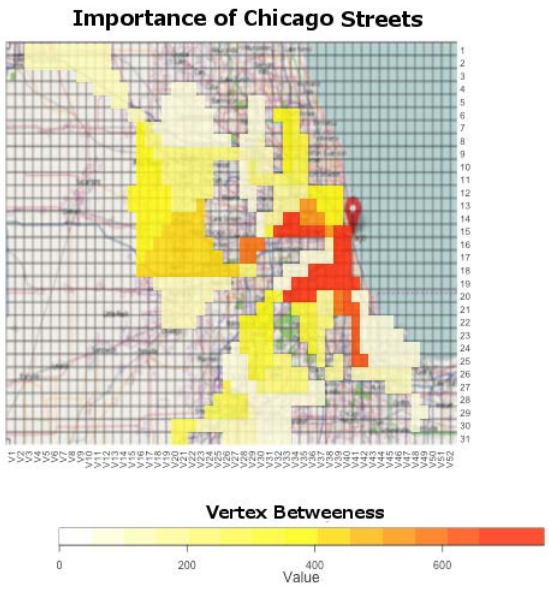


Figure 5: Street importance according to the vertex betweenness.

Table 1: The most important roads and convergences according to the city infrastructure.

Road	Importance	Edge	Convergence Importance	
79	759	31 → 79	775	100.00%
28	731	78 → 28	754	97.29%
78	719	80 → 78	633	81.68%
31	713	79 → 82	628	81.03%
26	702	82 → 80	615	79.35%
82	611	26 → 42	520.5	67.16%
80	597	50 → 9	427	55.10%
42	582	25 → 26	383	49.42%
25	531	76 → 31	380	49.03%
50	454	28 → 26	372	48.00%

Table 2: Parameters of the period of the day.

	Description	Time
1	Dawn	0:00am–5:00am
2	Morning	5:00am–10:00am
3	Lunch	10:00am–15:00am
4	Afternoon	15:00am–20:00am
5	Night	20:00am–0:00am

Figure 6 presents the average road flow intensity (metric 1) of working days in the left graphic while the two last graphics illustrate the multiplication of the average road flow intensity and the edge betweenness (metric 2) in dawn and afternoon. One can notice that metric 1 did not characterize the most important roads with a higher value, which is the advantage of metric 2. Therewithal, distincts behavior patterns stand out depending on the period of the day. The weekend graphics varies as well as the work days'.

Besides, phase transitions are difficult identify in a daytime analysis, but not in each period of the day. Figure 7 presents the flow intensity in April 19th 2013 in the morning

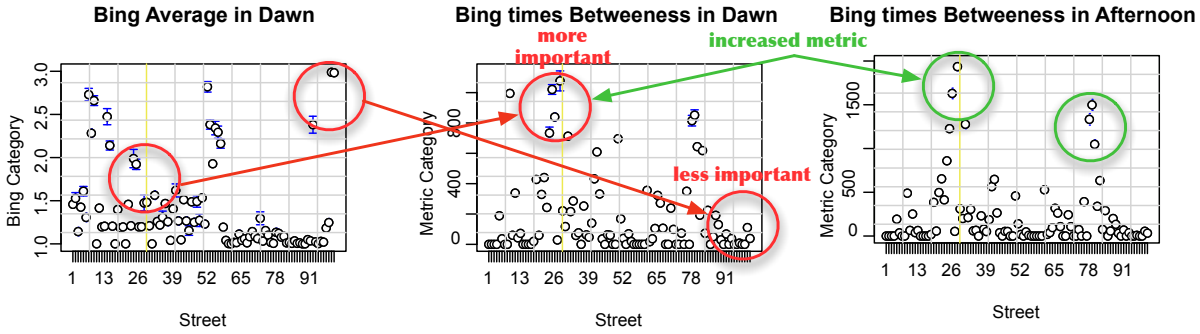
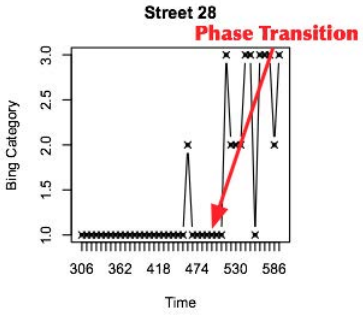
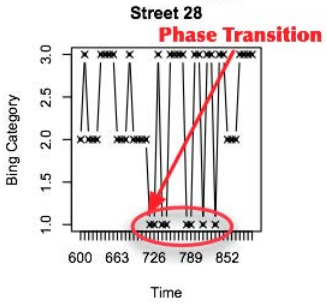


Figure 6: Boxplot of the average road flow intensity and such value updated by the edge betweenness.

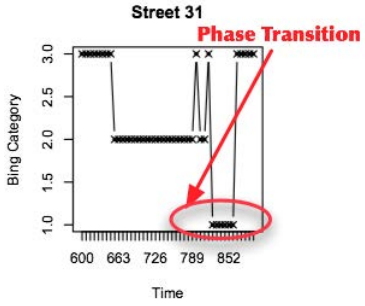


(a) Morning.

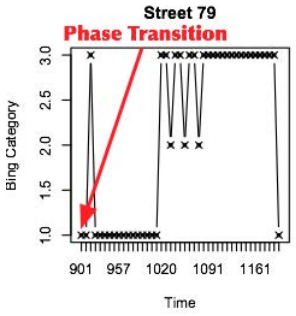


(b) Lunch.

Figure 7: Flow intensity analysis for street 28.



(a) Lunch.



(b) Afternoon.

Figure 8: Flow intensity analysis for streets 31 and 79.

and during lunch. One can notice that are several hours of the day that if we leave 7 minutes earlier, it will make a difference.

Figure 8 presents the flow intensity for streets 31 and 79, which is the most important road. About street 31, we can identify 852 minutes (14:12 hours) as a phase transition due the street flow intensity remains highly congested (level 3).

Through daytime analysis, we can discovery flow patterns during hours of the day, as well as phase transitions. Figure 9 presents the flow intensity result for the top-4 important streets in April 19th, 2013. One can notice that the street 79 has more traffic jam at night, as well as the street 28. Notwithstanding, the street 28 has more flow fluctuation during lunch, and in the afternoon and in the night it is highly congested for hours.

Finally, the average probability a street be free (green), medium (yellow), or congested (red) is determined by the frequency in the acquired Chicago area database. As the probability of each road is not equally likely, grouping the data is required. Thus, four clusters have been made, according to the average flow intensity so as to preserve the street main characteristics: (i) streets with values between 1 and 1.5; (ii) streets with values between 1.5 and 2; (iii) streets with values between 2 and 2.5; and (iv) streets with values higher than 2.5. Such division has been made with data of working days and of weekends, creating two distincts clustering of behavior.

So as to define the street categories, Figure 10 presents the probabilities of flow intensity for each. One can notice that category 1 has a higher probability of being green during almost all the day, except at dawn that presents a higher probability of being yellow. The same happens for category two, highlighting the red higher probability at 8pm. Streets of category 3 presents a variation of green and red probabili-

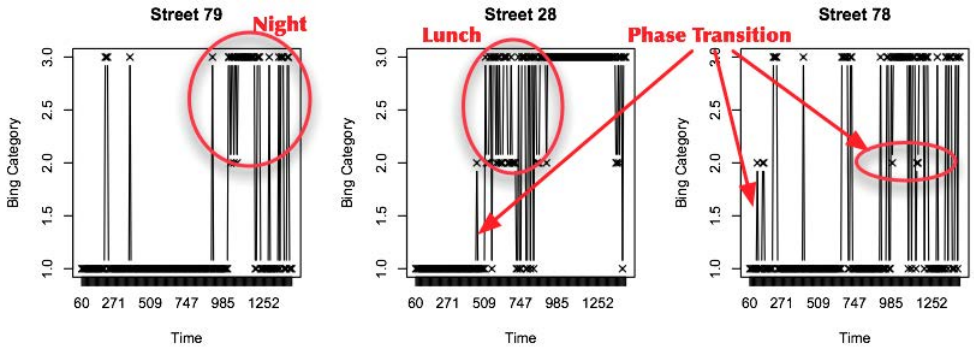


Figure 9: Day time analysis of the top-4 streets' flow intensity.

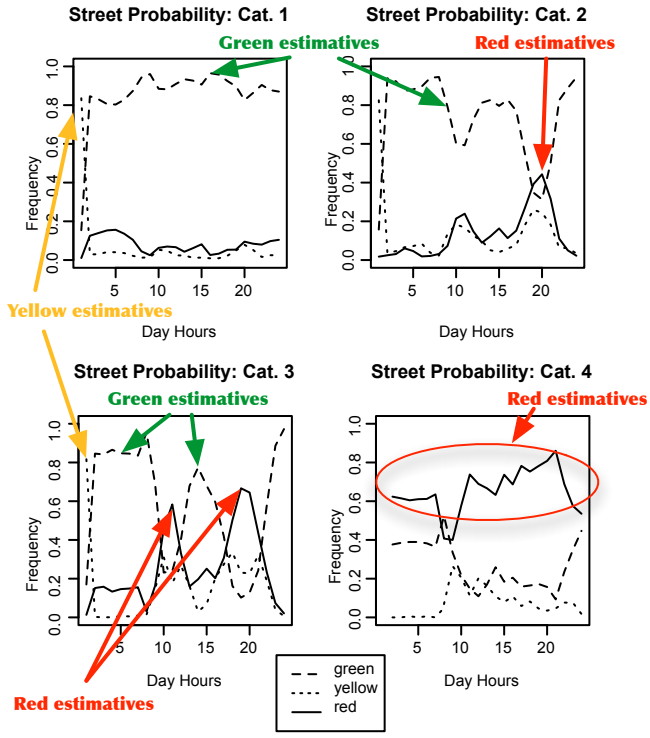


Figure 10: Probabilities of flow intensity per street category.

ties, being more susceptible of being red at 11am and 20pm. Finally, category 4 presents a red probability during most of the day, except for 9am that has a higher probability of being green.

Table 3 shows the classification of streets 79, 78, 28, and 31. One can notice that almost every period of the day the street category is 1, being sometimes 2. Categories 3 and 4 are rarer, standing out street 28 that change its category during the periods of the day. Even in the weekend, one can notice that street 28 has a distinct behavior in comparison with streets 79, 78, and 31.

Although analyzing such probabilities, shall the probabilities be valid for the next few weeks? Next section discusses the topic of whether this behavior is maintained during the days and if a prediction model can be developed.

Table 3: Classification of streets in periods of the day.

	Working Days				Weekends			
Streets	79	28	78	31	79	28	78	31
Dawn	1	1	1	1	1	1	1	1
Morning	1	2	1	1	1	1	1	1
Lunch	1	3	1	2	1	2	1	1
Afternoon	2	4	2	2	1	4	1	1
Night	1	2	1	1	1	2	1	1

5. TRAFFIC FLOW PREDICTION MODEL

5.1 Overview

In statistics, logistic regression is a type of regression analysis used for predicting the outcome of a categorical dependent variable based on one or more predictor variables. Its main characteristic is that the probabilities describing the possible outcomes of a single trial are modeled, as a function of the predictor variables, using a logistic function. Logistic regression measures the relationship between a categorical dependent variable and one or more independent variables. The probability function follows a logistic function, presented by equation 1.

$$P(t) = \frac{1}{1 + e^{-t}} \quad (1)$$

In this work, the categorical dependent variable is the flow intensity, being: (i) green; (ii) yellow; or (iii) red. The independent variables are the day, the hour of the day, the street number, and the number of pixels green, yellow, and red in the week.

Algorithm 4 synthesizes the prediction model algorithm, which is composed by two logistic regressions. The logistic regressions were made with only the first week work days. After gathering data, we considered each independent variable as categorical or numeric. The green flow intensity was considered as class 0, while class 1 was yellow and red intensities. The logistic regression 1 classified the flow intensity as green (class 0) or not (class 1 – yellow or red). Removing the green data, the logistic regression 2 classified the flow intensity as yellow (class 0) and as red (class 1). The regressors was made through the `glm` library in R–Project[6].

5.2 Results and Analysis

[6]See more information about R–Project in `www.r-project.org`.

Algorithm 4: Prediction Algorithm

```
Input: Database data
data1 <- previous weeks data;
// The categorical dependent variable
y <- vector of bing flow intensity (1, 2 or 3) of data1;
x <- matrix of the six independent variables of data1;
// Consider bing = 1 as class 0 and others as class 1
y = y-1;
y[y==2] = 1;
class0 <- y == 0;
class1 <- !class0;
// Run the logistic regression 1
flrm1 <- glm(y (factor(x[,1]) + factor(x[,2]) +
as.numeric(x[,3]) +
as.numeric(x[,4]) + as.numeric(x[,5]) +
as.numeric(x[,6])
), family=binomial("logit"));
// Prepare data for regressor 2;
y <- vector of bing flow intensity (only 2 or 3) ;
x <- matrix of the six independent variables data;
// Consider bing = 2 as class 0 and bing = 3 as class 1
y = y-2;
class0 <- y == 0;
class1 <- !class0;
// Run the logistic regression 2
flrm2 <- glm(y (factor(x[,1]) + factor(x[,2]) +
as.numeric(x[,3]) +
as.numeric(x[,4]) + as.numeric(x[,5]) +
as.numeric(x[,6])
), family=binomial("logit"));
```

Table 4 and 5 present the confusion matrix for each logistic regression (more details about Machine Learning concepts in [7]). As one can notice, the 109 of all data were false positive and 100 were false negative. In this context, falsely predicting an event (green considered as yellow) is better then missing an incoming event (yellow considered as green). Also, the percentage of False Positive (FP) is higher than the percentage of false negative in both regressors.

Table 4: Confusion matrix for each logistic regression using the week data.

Logistic Regression 1A			Logistic Regression 2A		
Green	FALSE	TRUE	Red	FALSE	TRUE
FALSE	11267	109	FALSE	7105	0
TRUE	100	70093	TRUE	0	4271

Table 5: Confusion matrix for each logistic regression using the previous week data.

Logistic Regression 1B			Logistic Regression 2B		
Green	FALSE	TRUE	Red	FALSE	TRUE
FALSE	1486	12272	FALSE	4737	1796
TRUE	728	79615	TRUE	2244	2599

After generating the confusion matrix, table 6 summarizes both regressors results. The accuracy was high (approximately 99%) indicating the proportion of correct predictions. The precision was also 99% indicating the probability of the predicted positive cases that were correct. For the second group of regressors, using the previous week data, the precision of free (green category) is 87% while the precision of the congested (yellow or red) is 60%, which can be explained by the less amount of yellow/red data in comparison with the green label.

Table 6: Other analysis of the regressors.

#	Accuracy	Recall	FP Rate	Specificity	Precision
1A	99.74%	99.85%	0.01%	99.04%	99.84%
2A	100%	100%	0%	100%	100%
#	Accuracy	Recall	FP Rate	Specificity	Precision
1B	86%	99%	90%	10%	87%
2B	65%	54%	27%	72%	60%

The recall and the specificity were also high, which show that one can have a high confidence in the model concerning correct predictions of positive and negative classes, respectively. The FP rate demonstrates that the flow intensity was predicted as congested while in fact it was free. Thus, the obtained value was very low due the proportion of negatives cases that were incorrectly classified as positive, as it was expected. Also, the Chi-squared test was made and the probability value (p-value) (i.e., the area under the chi-square distribution from the chi-square value to positive infinity), given the chi-square value and the degrees of freedom, was zero. The p–value shows that the observed result was highly unlikely under a null hypothesis.

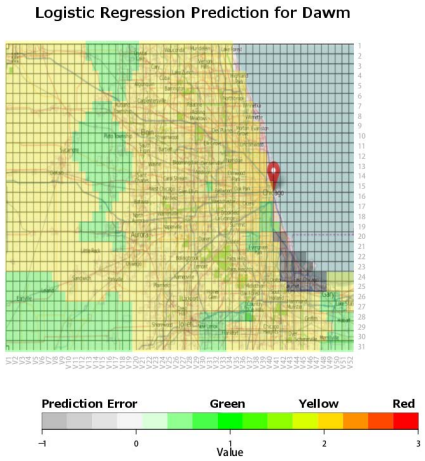


Figure 11: Heat map of prediction model results at dawn.

For a better visualization of the prediction model, the second week data was used as validation. Figures 11, 12, and 13 presents the prediction model and errors, in relation to the real flow intensity data, over Chicago area map for the following period of the day: dawn, morning, and lunch. Prediction model errors are presented with gray color. One can notice that at dawn we have a higher mobility in all area, and the prediction model not missed much. The maximum error was obtained for the lunch period. This visualization great advantage is the perception of movement in the morning to downtown, and from downtown at lunchtime. In general,

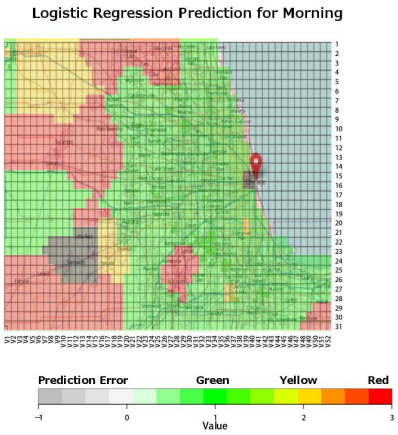


Figure 12: Heat map of prediction model results in the morning.

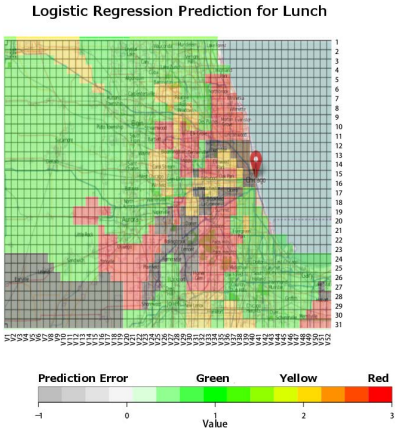


Figure 13: Heat map of prediction model results at lunch.

the obtained error of flow intensity was 10.98% for green, 3.36% for yellow, and 8.79% for red.

6. CONCLUSIONS

This work presented a methodology to acquire flow intensities from map services such as Bing maps and Google maps. Chicago area was evaluated as study case. A metric was applied to analyze the flow intensity according to importance of the streets (betweenness), in which a top-5 streets importance was established. Then, streets were aggregated in four categories in which phase transitions (shocks in the system) were identified. The main contribution was the proposal of a traffic flow acquisition metodology based on Geographical Information System (GIS), such as Bing Maps, which can be used for developing realistic mobility models for VANETs for a better evaluation of protocols.

Finally, a prediction model was designed to discovery future flow intensities for a target street. Based on logistic regression, the prediction model obtained an accuracy of 98%, a recall of 98%, a FP rate of 4%, a specificity of 96%, and a precision of 98%, in average for both regressors. In the prediction model validation, we presented visualization through a heat map over Chicago area. The obtained errors were 10.98% for green intensity, 3.36% for yellow intensity, and 8.79% for red intensity.

Future works include analyzing other cities, improving the prediction model using other information such as social networks. Also, a comparison between distincts map services will be developed.

7. ACKNOWLEDGMENTS

The authors thank the support of CAPES, CNPQ, and FAPEMIG.

8. REFERENCES

[1] R. d. S. Alves, I. do V. Campbell, R. de S. Couto, M. E. M. Campista, I. M. Moraes, M. G. Rubinstein, L. H. M. K. Costa, O. C. M. B. Duarte, and M. Abdalla. *Redes Veiculares: Princípios, Aplicações e Desafios*, chapter 5, pages 199–254. Minicurso do Simpósio Brasileiro de Redes de Computadores – SBRC'2009, 2009.

[2] X. J. Ban and M. Gruteser. Towards fine-grained urban traffic knowledge extraction using mobile sensing. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, UrbComp '12, pages 111–117, New York, NY, USA, 2012. ACM.

[3] R. Bauza, J. Gozálvez, and J. Sánchez-Soriano. Road traffic congestion detection through cooperative vehicle-to-vehicle communications. In *LCN*, pages 606–612, 2010.

[4] H. Hartenstein and K. Laberteaux. A tutorial survey on vehicular ad hoc networks. *Communications Magazine, IEEE*, 46(6):164 –171, june 2008.

[5] E. Horvitz and T. Mitchell. From data to knowledge to action: A global enabler for the 21st century. *Data Analytic Series, Computing Community Consortium, Computing Research Association (CRA)*, June 2010.

[6] R. S. J. Apacible and L. Liao. Prediction, expectation, and surprise: Methods, designs, and study of a deployed traffic forecasting service. *Twenty-First Conference on Uncertainty in Artificial Intelligence, UAI-2005*, July 2005.

[7] R. Kohavi and F. Provost. Glossary of terms. volume 30, February/March 1998.

[8] F. Li and Y. Wang. Routing in vehicular ad hoc networks: A survey. *Vehicular Technology Magazine, IEEE*, 2(2):12 –22, june 2007.

[9] J. Nzouonta, N. Rajgure, G. Wang, and C. Borcea. Vanet routing on city roads using real-time vehicular traffic information.

[10] M. Research. Predictive analysis for traffic.

[11] M. Research. Developing business intelligence and data visualization applications with web maps, 2013. [Online; accessed May 31, 2013].

[12] J. L. Toole, M. Ulm, M. C. González, and D. Bauer. Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, UrbComp '12, pages 1–8, New York, NY, USA, 2012. ACM.

[13] C. Tribune. Chicago no. 1 in road congestion, 2011. [Online; accessed May 31, 2013].