Master Thesis

Identification of anchor zones for floating content in VaNETs based on centrality measures

Fábio Massalino
fm@ic.ufal.br

Advisor:

Dr. André Luiz Lins de Aquino

Maceió - August of 2019

Fábio Massalino

Identification of anchor zones for floating content in VaNETs based on centrality measures

Thesis presented as a partial requirement to obtain the Master degree by the Master Course in Computational Modeling of Knowledge of the Computing Institute of the Federal University of Alagoas.

Advisor: Dr. André Luiz Lins de Aquino

Maceió - August of 2019

Catalogação na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 – 1767

M414i Massalino, Fábio.
Identification of anchor zones for floating content in VaNETs based on centrality measures / Fábio Massalino. – 2019.
43 f. : il.

Orientador: André Luiz Lins de Aquino.
Dissertação (mestrado em Modelagem Computacional de Conhecimento) – Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2019.

Bibliografia: f. 42-43.

1. Ancoragem (Modelagem e simulação de sistemas). 2. Aritmética de ponto flutuante. 3. Rede veicular. 4. Medidas de centralidade. I. Título.

CDU: 004.414.23

Folha de Aprovação

Fábio Massalino

Identification of anchor zones for floating content in VaNETs based on centrality measures

Dissertação submetida ao corpo docente do Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas e aprovada em 30 de agosto de 2019.

Prof. Dr. André Luiz Lins de Aquino
Instituto de Computação - UFAL
Orientador

Banca Examinadora:

Prof. Dr. Heitor Soares Ramos Filho
Departamento de Ciência da Computação - UFMG
Examinador Externo

Prof. Dr. Danielo Gonçalves Gomes
Departamento de Engenharia de Teleinformática – UFC
Examinador Externo

Raquel da Silva Cabral

Profa. Dra. Raquel da Silva Cabral
Instituto de Computação - UFAL
Examinador Interno

ABSTRACT

Floating Content is the information with local and temporal validity and spatial relevance. These contents are produced and consumed by mobile nodes in a region called anchor zone. This floating content remains alive during its lifecycle without the need for fixed infrastructure, being maintained only by the mobile nodes that consume and forward it — because of this, determining the most viable anchor zones to use floating content is one of the challenges of the Vehicular Network. This work presents the study and scientific experiment where the use of measures of centrality to identify the most viable anchor zones to spread the floating content proved to be effective. The study consisted of two parts: the first consisted of analyzing two real vehicular bases: one the Mobile Century, which traffic of 2 900 vehicles on a section of 18km of "Nimitz Freeway", near Union City, California, in the United States, during February, 02 of 2008; other the T-Drive, which traffic of 8 388 vehicles from 02 to 08 of February of 2008 in a region of Beijing, China. In these two bases, two techniques were presented to identify the most viable anchor zones: GRID POSITION and BETWEENNESS CENTRALITY MEASURE. This study allowed us to observe the regions and times where there was the most significant flow of vehicles, through GRID POSITION technique and the regions where there was the most influential flow of vehicles using BETWEENNESS CENTRALITY. Due to not being able to verify which regions are more viable to use floating content, which motivated the second part of the work. The second part consisted in simulate two real bases over synthetic traffic: The first one shows the traffic of 18 000 vehicles in a section of 18km of "Nimitz Freeway" near Union City, California, in the United States. The second has the traffic of 100 000 vehicles in a region of Beijing, China. On these bases, we presented four model strategies, based on FLOW OF VEHICLES(State-of-The-Art), BETWEENNESS AND DEGREE CENTRALITY MEASURES and the HYBRID, mixed flow of vehicles and betweenness centrality to determine the anchor zones most viable in VANET scenarios. After identifying the most viable anchor zones in each strategy, we used a simulation where we verified in which of the strategies the floating content remained alive for longer and the number of vehicles affected. These simulation experiments showed that the strategies of centrality measures performed better than other strategies. The total simulation time was 18000s. By using centrality measures, the floating content remained alive for 17 008s on "Nimitz Freeway", USA and 13 579s in Beijing, China.

Keywords: Anchor Zone. Floating Content. Vehicular Network. Centrality Measure.

ACKNOWLEDGEMENTS

I would like to thank everyone who made the completion of my Master's degree possible! First, my professors of the Master Program "Computational Modeling of Knowledge": Alejandro César Frery, Heitor Soares Ramos Filho, Raquel da Silva Cabral, Ig Ibert Bittencourt Santana Pinto and especially to the professor who was also my advisor, Andre Luiz Lins de Aquino, who from the beginning believed in my work and and helped me in all stages of the Master's program! To the members of the LACCAN - Laboratory of Scientific Computing and Numerical Analysis - Ufal, where through the seminars I was able to learn and the contributions were essential for its accomplishment. I thank my family for their patience in the moments when they required dedication in the accomplishment of the master's degree, especially my wife Eleusa, who besides supporting me, is giving me the greatest gift that a man can get: a daughter, Laura, God willing, she will born in the end of this year! Finally, acknowledge support from the Brazilian research agency CAPES, CNPq, FAPEAL e FAPESP.

Fábio Massalino

List of Figures

Table of Contents

1

INTRODUCTION

THIS master's thesis reports on the evaluation of the use of measures of the centrality of networks to identify regions where it is feasible to use content with temporal and spatial relevance shared between vehicles without the need of fixed infrastructure. For the accomplishment of this study, we perform a literature review, experiments in simulators of vehicular networks and analysis of real and simulated vehicular networks scenarios.

A vehicular ad hoc network (VANET) has enjoyed a variety of applications, from safety and comfort to entertainment and commercial services (Lee et al., 2014, 2016; Sookhak et al., 2017). For instance, traffic safety applications can be used to reduce the number of accidents, especially the loss of life on the road, by providing warning systems such as intersection collision, lane change assistance, overtaking the vehicle, emergency vehicle, wrong-way driver, risk of collision, hazardous location and signal violation (Chaqfeh et al., 2014).

On the other hand, driver comfort applications can significantly improve our daily lives. These applications allow the delivery of traffic data, as well as advertisements such as the congestion state of a distant road, parking slots available in one parking place, estimated time of arrival of the bus at a bus stop, sales information at a department store and meeting times in a conference. In general, VANETs have some issues relevant to floating content distribution (Lee et al., 2014):

1. Vehicles produce and consume a great quantity of content with locally relevant property (time, space and consumer);

2. Vehicles naturally seek content regardless of their providers;

3. Vehicles collaborate using their resources to create value-added services with minimal help from the Internet infrastructure.

To vehicles communicate without any support infrastructure, it is necessary to ensure that, throughout the life cycle of disseminated content, the regions where they will be alive there are always vehicles that can produce, consume, store and forward the floating content. For this, it is critical to determine which regions are most feasible for the use of floating content.

Ott et al. (2011) say that all users are mobile nodes with wireless interfaces and that there is no supporting infrastructure for the system.

Nodes originate floating content items with a defined anchor zone and lifetime, and the message must reside in the anchor zone when posting. For this, they need to be location-aware,e.g., by using GPS, triangulation-based methods using WLAN access points or cellular base stations, or any other method offering reasonable accuracy. Other users are interested in these items and accept copies to store and (probabilistically) further replicate within the anchor zone. Some considerations (Ott et al., 2011):

- Explicitly allow information items to disappear from the system;
- provide no guarantees about their availability;
- If no (or too few) nodes are around to replicate a content item; it may disappear;
- Once the lifetime of an item expires, it will be deleted by all nodes.

The floating content has characteristics temporal and spatial. The vehicles consume it only during its life cycle and in a specific region (anchor zone). Therefore the floating content protocol must contain in its header at least this information: Its identifier, anchor zone ID and its lifetime. Ott et al. (2011) defines a 4-phase protocol for exchanging floating content messages:

1. Neighbor discovery: Nodes continuously send neighbor discovery beacons to discover peers within radio range;
2. Offering content: After discovering a peer, a node sends a summary message with the anchor zone ID and content items available for replication;
3. Requesting content: As soon as a vehicle is aware of what a neighbor has to offer, it requests all or a subset of the items for;
4. Content transfer: The vehicles exchange different messages containing the requested items until to complete the batch or nodes lose contact. In the former case, the receiver discards incomplete messages; in the latter, the protocol returns to phase (2).

The floating content used in simulations on this, the header of the messages contained the anchor zone identifier and TTL. The vehicles were aware of the anchor zone that was at a specific moment, and if on receiving a message it belonged to anchor zone in which it was, the floating content was stored and forwarded while it remained in anchor zone or the TTL expired. We set the propagation frequency between nodes to 3 seconds and the TTL at 3600s.

Based on these issues, the main contribution of this study is to present that the use of centrality measures is useful in identifying the most viable anchor zones for the use of floating content. For this, experiments were carried out to identify these anchor zones in specific regions. To show that these techniques are the most efficient, we performed simulations in which traffic was generated in the anchor zones identified by each technique and analyzed in which the floating content remained alive for longer — the study consist of two parts. The first one presented real vehicular bases and the second, produced with simulated vehicular bases.

The first part (Massalino & Aquino, 2018) analyzes two real bases of vehicular networks: The Mobile Century and T-Drive. The first one presents the traffic occurred between 9h to 19h on February, 02 of 2008, into a stretch of 18km of "Nimitz Freeway", near Union City, California in the United States. They collected the data from GPS-equipped mobile phones of 2 900 vehicles. Each vehicle generated a record, on average every 4-8 seconds, where it recorded its location, speed, and the instant of time. The second one has the characteristic of very sparse, because collected data of 8 388 vehicles between 02 to 08 of February of 2008 in a region of 200km of height (latitude) x 264km of width(longitude). In these two regions, two model strategies to determine the most appropriate Anchor Zones in VANET scenarios. The first strategy, **Grid position**, the region where there was the traffic of vehicles was divided into grids, where each anchor zone has 200m of height and 200m of width and we verified that there was more traffic than other ones. The second strategy presents the measures of centrality were through the **Betweenness centrality**. We verified that the vehicles have a greater influence on the communication between the other vehicles. After that, it was possible to determine which regions these vehicles traveled. For this, we represent the Mobile Century, and T-Drive bases into a graph, where vehicles were a vertex, allowing that we verified which ones had more influence as forwarders, through the betweenness centrality measure.

The second part[1] we analyze two simulated bases of vehicular networks. The first simulation presents the traffic of 18 thousand vehicles during 5h, into a stretch of 18km of "Nimitz Freeway", near Union City, California in the United States. The second one corresponds the Beijing metropolitan area, with traffic of 100 thousand vehicles during 5h in a region of 200km of height (latitude) x 264km of width(longitude). In these two regions, four model strategies to determine most relevant anchor zones in VANET scenarios: The first strategy, **Flow of Vehicles**, identify the regions where there was a greater flow of vehicles during the interval of time. The second and fourth strategy presents the **Measures of Centrality** where through the centrality betweenness and degree where we verified the regions had vehicles that have greater influence in the communication between the other vehicles and vehicles with more number of connections, respectively. The third strategy, **Hybrid**, mixed the strategies of the flow of vehicles and betweenness centrality.

The remainder of this Master's Thesis is as follows. Chapter 2 shows the related works.

[1] It will be submitted to Ad Hoc Network Journal from Elsevier, JCR 3.151

Chapter 3 presents the techniques used in this paper. Chapter 4 comments about the evaluations and results. Chapter 5 concludes the study.

The chapter presented the approach to the problem-focused in this dissertation. The next chapter will talk about the related works about anchor zones and floating content in VaNETs.

Chapter 2

RELATED WORK

OVER the last few years, the use of applications and services in vehicular networks has been increasing in order to provide safety and comfort for drivers and passengers (Fernando et al., 2013). Despite its increasing use, it is difficult to exploit its full potential due to vehicular networks presenting some characteristics such as scarcity of resources, frequent disconnections, and mobility. Some authors have presented some concepts related to this environment: Gerla (2012) says that vehicles represent an increasingly important source of computing and sensing resources for drivers as well as for urban communities. They use these resources to define a mobile vehicular platform were several utilities with locally relevant can be created and shared among all vehicles in a specific region. The authors called this mobile vehicular platform of *Mobile Cloud Computing*, which aims to study mobile agents (people, vehicles, robots) as they interact and collaborate to feel the environment, process data, propagate the results and share resources. Already Fernando et al. (2013) highlights the specific concerns in mobile cloud computing. They discuss the different approaches taken to address these issues and challenges that have not yet addressed. Lee et al. (2014) presents the *Vehicular Cloud Computing* where it seeks to present the mobile cloud model for vehicle networks where vehicles can produce, consume, and share content between them, thus forming Vehicle Cloud Networking.

In this way, vehicles exchange content with local relevance and space-time constraints. Hyytiä et al. (2011) calls this content of Floating Content because it does not need to use a fixed infrastructure, will remain alive while it is "floating" between the vehicles within range, ceasing to exist when there is no vehicle to share this content.

In order to make the concept of floating content feasible in practice, Ott et al. (2011) propose a system, algorithms and protocol specification in detail of the use of floating content on mobile devices without the need of fixed infrastructure, using principles of an opportunistic network. The performance of floating content is evaluated in general and especially of different replication and exclusion strategies through extensive simulations using a map-based mobility model in the downtown of Helsinki, Finland.

Rizzo et al. (2018) present studies where they report that the viability of floating content varies according to mobility patterns of mobile nodes, where they consider two different mobility models that describe pedestrian behavior and information transfers occur in periods when users pause between the movements. The results reveal the critical relevance to the performance of the floating content in the group dynamics in the user movements.

Because vehicles act as forwarders of floating content, this can generate much traffic, causing congestion. Hagihara et al. (2017) proposes a floating content delivery control method called *PFCS*(Proportional control for Floating Content Sharing), which controls the message ownership rate, that is, the fraction of mobile nodes carrying the message in an anchor zone. In Ciobanu et al. (2017) is presented interest-based content dissemination for opportunistic networks entitled *Interest Spaces*, which allows nodes to only share contents that are of interest, allowing the proposed framework to take care of caching, routing, routing and dissemination. In Duarte et al. (2018), they present approaches that combine *Content Centric Networking*, Floating Content and *Software Defined Networking*(SDN) to present an innovative and adaptive VANET architecture for intermittent connectivity, node density fluctuation, and one-sided mobility patterns and performance of applications and network resources on the other, achieving high QoS.

In order for the floating content to remain "alive" in the time, it is useful, defining the regions where it is "floating" is fundamental, since they must be viable to use it. Some authors (Hyytiä et al., 2011; Virtamo et al., 2013) called this region of anchor zone, where users exchange the information item on condition that they are within each others' transmission ranges.

Hyytiä et al. (2011) analyze the expected lifetime of the information in these systems, considering situations in which anchor zone is significantly larger than the node's transmission range. In this case, was aimed to characterize the viability of the operating region, especially the so-called critical condition, related to the average number of encounters that a randomly chosen node experiences per visit, provides the threshold above which the expected information lifetime increases very rapidly, and content can actually "flow" without any supporting infrastructure.

Di Maio et al. (2017) proposes a method to improve the performance of the floating content by optimizing the size of the anchor zone with the support of SDN controller, that collects information of mobility, such as speed and position, of the vehicles in the coverage area. Manzo et al. (2017) presents how to size an anchor zone in a realistic vehicle scenario, through a simple set of sizing strategies, based on the estimation of some critical parameters of mobility and floating content performance. These strategies were evaluated in measurement-based mobility patterns, providing the first indication of their relative performance and the feasibility of floating content in practical scenarios. Virtamo et al. (2013) seeks to study under what conditions the floating content can stay alive. This analysis, where the authors consider anchor zone as a circular region and used a mobility model of the random type, use the transport equation, adopted from the nuclear reactor theory, that derive the condition of criticality that defines a lower limit for the product of the node density, communication distance and the radius of the anchor zone required for the information to fluctuate. Manzo et al. (2018) proposes a solution to scale anchor zone through deep learning to minimize the communication capabilities to ensure the availability of the messages. This proposed method is validated its effectiveness and efficiency through a numerical study.

Some authors use centrality measures in mobile networks: Magaia et al. (2015) presents the use of betweenness centrality measure to identify the most influential nodes in a delay tolerant network, which employ a store-carry-and-forward approach, keeping the messages until to found an appropriate node to forward them. The identification of the best routing nodes is made using this approach. Through the use of centrality measures, Kim & Yang (2017) proposes a new routing scheme, where nodes with higher probabilities of contacts with other nodes tend to gather while nodes with higher centrality of distance compensate for intermittent connection interruptions between nodes in the network. To avoid a network overhead if all nodes act as a router, Vázquez-Rodas & de la Cruz Llopis (2015) presents a protocol based on centrality metrics where only the most influential nodes, identified by the three most common centrality measures: degree, betweenness, and closeness; act as routers, contributing to better performance from the network. Khan & Ghamri-Doudane (2015) goes further and identifies the most influential nodes to be used as routers, using the algorithm called *CarRank*, which takes into account not only the use of centrality measures but the importance of information, its spatio-temporal availability and neighborly topology to find its relative importance analytically in the network.

The previous papers presented the concept of floating content and the dissemination area. We do not find, in the current literature, works to determine, based on complex networks concepts, the regions with higher viability for the use of floating content, being the proposal of this present one. Before using complex networks to identify these most viable anchor zones, we divided the regions into grids and checked the regions where there was the most significant flow of vehicles, and the results were very satisfactory.

> The present chapter presented the works related to the identification of *anchor zones* and the use of *floating content*. The next chapter will present the techniques used to identify feasible anchor zones used in this work.

Chapter 3

IDENTIFICATION OF ANCHOR ZONES

In this chapter, we present the techniques used in this work to identify the most viable anchor zones to disseminate floating content in vehicular networks.

In VANETs there is a specific region named "Anchor Zone", where users can exchange information once they are each in transmission range. Once the user leaves this region, it deletes the information because it is considered obsolete outside the anchor zone (Hyytiä et al., 2011). Mobile nodes assume that there is no supporting infrastructure for the system. These nodes are interested in items of information posted by other users, cooperating by replicating the content to the other interested nodes.

Anchor zone (AZ) is a delimited area in which items must be available. Figure 3.1 is an example of an Anchor Zone, with the nodes and their range of communication: The information dissemination is within the scope of the anchor zone. The node in the anchor zone receives the information, can consume and forward it to other nodes during the period that is inside the anchor zone. We illustrate the probability of replication and deletion on the right diagram.

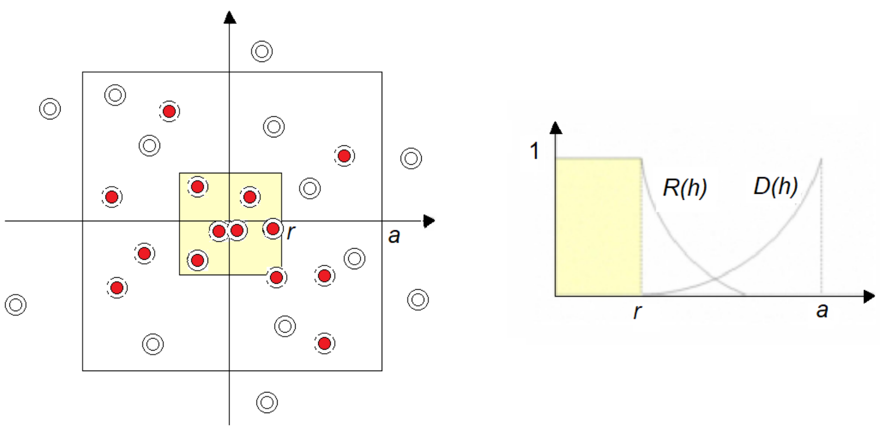

Figure 3.1: A sample of Anchor Zone

The probability of replicating a content item is equal to 1 when the node is within the region r. When the node leave this region, the probability will range from 0 to 1 while remaining in region a: this value will be close to 1 while this node is close to region r and will decrease as it move away from it, being 0 when the node leaves the region a. This can be observed in the equation,

$$P_r(h) = \begin{cases} 1, & h \leq r \\ R(h), & r < h \leq a \\ 0, & \text{otherwise} \end{cases} \tag{3.1}$$

where h corresponds to the distance of the node from the center point.

Consider the node A having an item J, with an anchor zone called r. When a node A finds another node B, the node A will replicate the item J to B with the probability of $R(h)$, where the replication probability $R(h) \in [0,1]$ is 1 within the Anchor Zone, and decreasing to 0 outside this anchor zone.

The deleting probability this content item is inverse: being equal to 1 when within the region r, being between 0 and 1 in the region a (the closer to the region r is close to 0), 1 being outside this region a. This can be seen in the equation

$$P_d(h) = \begin{cases} 0, & h \leq r \\ D(h), & r < h \leq a \\ 1, & \text{otherwise} \end{cases} \tag{3.2}$$

For the deletion of the items, can be defined a similar function $D(h)$ which is 0 if the node A is within the region r and the function is increasing when A is outside the region r but within the region a. The item j remaining in this region allows protection so that the items do not disappear when the nodes move out of it for a brief moment and then return. When the node finally exits the region a, the items are deleted.

In order to identify the most feasible anchor zones for the dissemination of floating content, we verify the regions that had the higher *flow of vehicles*: this technique is the State-of-The-Art; and we compared with three other techniques: *betweenness centrality measure*, *degree centrality measure*, and *hybrid*, was created by us and merges the techniques of flow of vehicles and betweenness centrality measure. In each technique, ten regions with the best values were used, since selecting a larger number of regions did not show a significant change in the choice of the most effective technique. We detail each technique below:

Flow of vehicles The technique is the State-of-The-Art. We divide the regions into grids, and we check the number of vehicles in each region every 10 seconds. We use the average number of vehicles that each anchor zone had in the time interval of the simulations.

Network centrality measure Opsahl et al. (2010) say in the study of social networks that the nodes can represent individuals, organizations, or even countries, and the links of these refer to communication, cooperation, friendship, or commerce. One of the goals of social networking is to capture the global or local relationship between nodes. According to Opsahl et al. (2010), rather than identifying whether two nodes are interconnected, it is important to find out the weight of these links.

The measure centrality has fundamental importance in social network analysis (Kim & Yang, 2017). The centrality of nodes, or the identification of which nodes are more "central" than others, has been a critical issue in network analysis (Freeman, 1978; Borgatti, 2005).

Freeman (1978) argued that central nodes were those in the "thick of things" or focal points. The Figure 3.2 show the network consisting of 5 nodes and can clarify this idea (Opsahl et al., 2010):

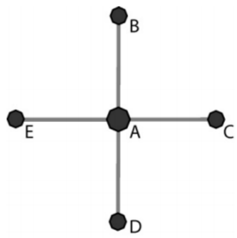

Figure 3.2: A star network with 5 nodes and 4 edges. (Opsahl et al., 2010)

The middle node has three advantages over the other nodes, and this respect with three different measure of node centrality, formalized by (Freeman, 1978): *Degree*: it has more neighbors than other nodes; *Closeness*: it can reach all the others more quickly and *Betweenness*: it controls the flow between the other ones. About these measures of centrality, we tried to use the *Closeness* but the results were unsatisfactory compared to the State-of-the-Art and we decided to use only the *Betweenness* and *Degree*.

1. **Betweenness centrality**: It indicates the number of shortest paths from all nodes to all others that pass through a node.

   A node with a high betweenness centrality can enhance interactions between the nodes that it links. Betweenness centrality for N_i is

$$Betweenness_i = \sum_{j=1}^{m} \sum_{k=1}^{j-1} \frac{g_{j,k}^{(i)}}{g_{jk}}, \tag{3.3}$$

   where g_{jk} is the total number of geodesic paths between N_j and N_k, and $g_{jk}(i)$ is the number of those geodesic paths that include N_i.

   Defining the betweenness centrality of a node is difficult due to the lack of global network topology. Hence, the equation used in the ego betweenness (Everett &

Borgatti, 2005) is $A^2[1-A]$, where A is the adjacency matrix of N_i. $A_{ij} = 1$ if there is contact between N_i and N_j, and 0 if there is not. Figure 3.3 shows an example of the calculation of the betweenness centrality $B(1)$ for N_1, assuming that N_1 has encountered N_2, while N_3 had encountered N_1, N_2, and N_4, and N_1 had earlier encountered N_4 and N_5. When $i \neq j$, since the walk is of length 2, it must be a path. The number of paths of length 2 is needed to count for non-adjacent pairs of nodes since these will be geodesics. It follows that $A^2[1-0]_{i,j}$, where 1 is a matrix of all 1's, gives the number of geodesics of length 2 joining i to j. The sum of the reciprocal of the entries gives the ego betweenness.

$A =$

	N_1	N_2	N_3	N_4	N_5
N_1	0	1	1	1	1
N_2	1	0	1	0	0
N_3	1	1	0	1	0
N_4	1	0	1	0	0
N_5	1	0	0	0	0

$A^2[1-A] =$

	N_1	N_2	N_3	N_4	N_5
N_1	*	*	*	*	*
N_2	*	*	*	2	1
N_3	*	*	*	*	1
N_4	*	*	*	*	1
N_5	*	*	*	*	*

Figure 3.3: Calculation of the betweenness centrality of N_1. (Kim & Yang, 2017)

2. **Degree centrality**: It indicates the number of nodes that a focal node is connected to, and measures the involvement of the node in the network. Its simplicity is an advantage: we must know only the local structure around a node to calculate it (McPherson et al., 2001).

   The degree centrality of node N_i is

$$Degree_i \quad = \quad \sum_{j=1}^{m} l(i,j), \tag{3.4}$$

   where $l(i,j) = 1$ if a link exists between N_i and N_j and m is the number of nodes. The degree centrality of N_i is the number of nodes connected to it.

   However, there are limitations. The measure does not take into consideration the global structure of the network. For example, although a node has many connections, it might not be in a position to reach others quickly to access resources (Borgatti, 2005).

Hybrid We created this technique where mixed the techniques of the *flow of vehicles* and *betweenness centrality measure*. For this, we checked the maximum value of betweenness each anchor zone had and multiplied by the number of vehicles it had. It calculated the average that each anchor zone had in the interval of the simulation.

In the first phase of the study, we used real data to identify these regions. In this way, it was possible to identify the anchor zones, but to prove in fact that these are the most viable anchor zones, it became necessary to use a simulator. By using a simulated data, in the second phase, once we identify the anchor zones, floating content was disseminated in these regions to verify in which technique the identified anchor zones kept the floating content alive longer.

3.1 Real data

Using real data, in order to determine which anchor zones have the higher viability of the floating content, we use two databases of Vehicular Network: Mobile Century[1]; and T-Drive[2]. The first one was collected on I-880, the "Nimitz Freeway", between Winton Avenue to the north (Latitude: 37,656 19, Longitude: −122,101 5) and Stevenson Boulevard to the South (Latitude: 37,529 6, Longitude: −121,987), near Union City, California, in the United States. There are logs files of 2 900 vehicles, being that 1 388 from South to North and 1 512 from North to South, on 02 February 2008. Each vehicle was equipped with Nokia model N95 with a GPS-enabled mobile phone, traveling about 18km on the I-880 motorway. The vehicles generated a `.csv` file (`veh_number.csv`), where the number corresponds to a numeral identifier of the vehicle containing the fields listed in the Table 3.1:

Table 3.1: Columns of `veh_number.csv` files.

Column	**Description**
unix.time	Datetime in Unixtime format (seconds, UTC)
latitude	Latitude coordinate, in degrees
longitude	Longigute coordinate, in degrees
postmile	Miles from the begin of I-880
speed	Speed(MPH)

For the accomplishment of the work, all 2 900 files were joined, generating the file `vehicles_mc.csv`. We add two fields to this file: the "ID", corresponding to the vehicle that generated these records and the "SN", which with the "S" value for vehicles which North-South direction and the "N" for vehicles in the South-North direction. The Figure 3.4 shows a scatter plot of locations these 2 900 vehicles obtained from `vehicles_mc.csv` file.

The second database (T-Drive) was collected in Beijing, China, within limits: P1:(Latitude: 40,645 132, Longitude: 115,004 561), P2: (Latitude: 41,098 065, Longitude: 117,989 349), P3: (Latitude: 39,637 456, Longitude: 117,999 905) and P4: (Latitude: 39,299 329, Longitude:

[1] http://traffic.berkeley.edu/project/downloads/mobilecenturydata
[2] https://www.microsoft.com/en-us/research/project/t-drive-driving-directions-based-on-taxi-traces/

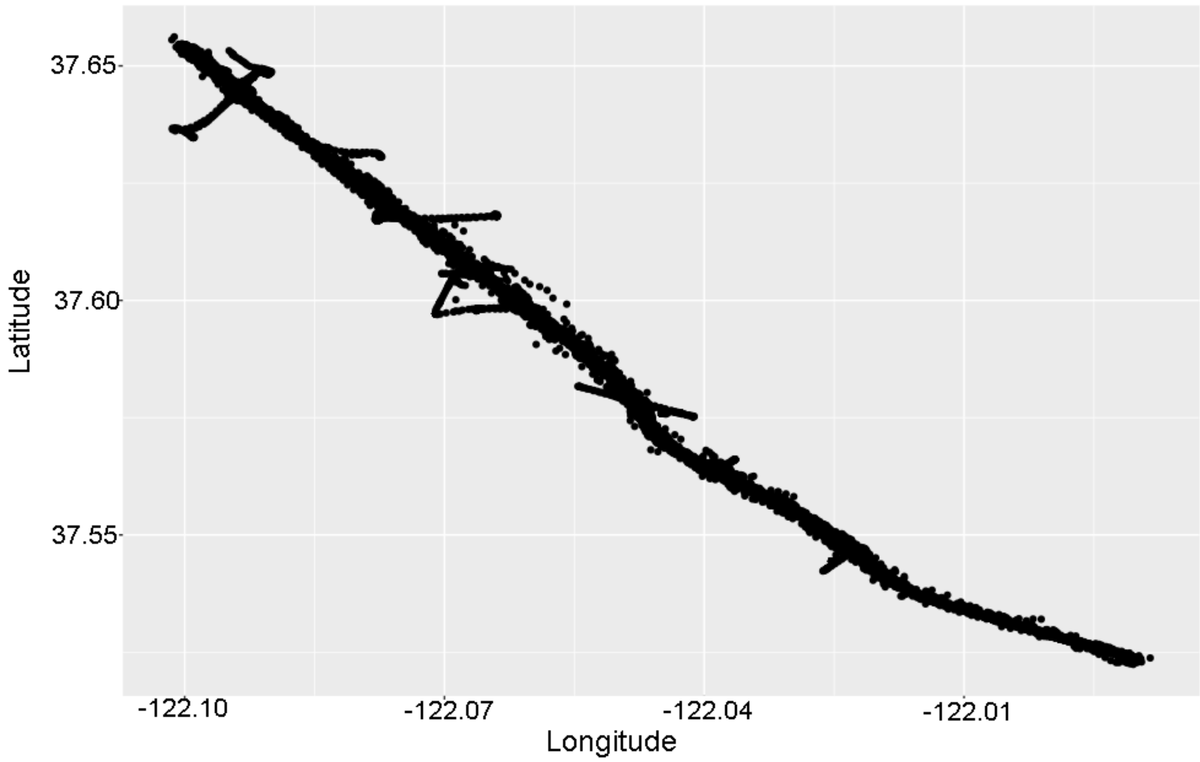

Figure 3.4: Mobile Century Scatter Plot.

115,917 965). There are logs files of 8 838 vehicles collected between February 02-08 of 2008. Each vehicle generates a `.txt` file (`number.txt`), where the number corresponds to a numeral identifier of the vehicle containing the fields listed in the Table 3.2:

Table 3.2: Columns of `number.txt` files

Column	**Description**
id	Vehicle identification
datetime	Datetime in YYYY-MM-DD HH:MM:SS format
longitude	Longitude coordinate, in degrees
latitude	Latitude coordinate, in degrees

All 8 388 files were joined, generating the file `vehicles_T.txt`. The Figure 3.5 shows a scatter plot of locations these 8 388 vehicles obtained from `vehicles_T.csv` file.

To perform the identification of anchor zones we use the first two techniques presents in section 3: **Flow of Vehicles** that divides the region of each base into grids, where each square has 200m of width by 200m of height. We verify that this square size presents the higher flow of vehicles; and **Betweenness Centrality Measure** that modeling the VANET communication as a graph, where each vehicle is a vertex in the graph. Thus, it is possible to identify the vehicles

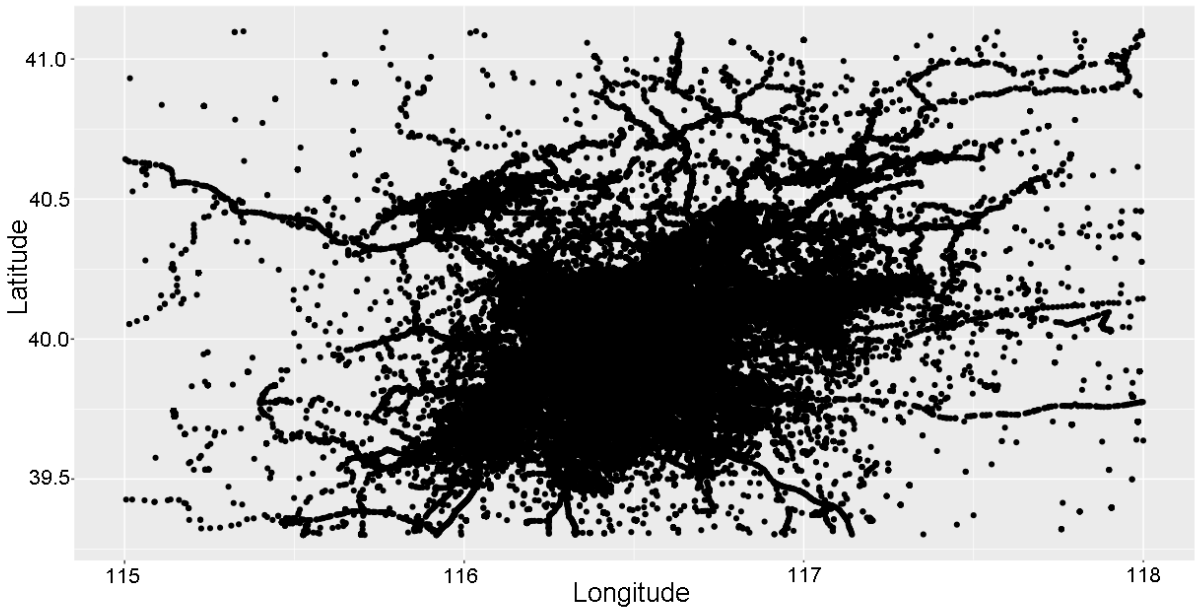

Figure 3.5: T-Drive Scatter Plot.

that had more importance as routers between the other vehicles, based on the measure of centrality.

In the **Flow of Vehicles** technique, through log files collected, the anchor zones were determined by dividing the region into squares of 200m of height and width, creating a grid. Each row is equivalent to latitude and column to longitude. The Mobile Century grid, therefore, has 3 636 regions, identified by **YYXX**, where the first two digits (**YY**) correspond to the line and (**XX**) to the column. Of these, 263 are regions where there was a vehicle log. The T-Drive created grid has 1 324 323 regions, identified by **YYYYXXXX**, where the first four digits (**YYYY**) match the rows, and the other (**XXXX**) rows correspond to the columns. Of these, 84 447 are regions where there was a vehicle log. After that, it was created a dataset (`tot_ah_veh.txt`) with each base, where each line was registered the number of vehicles that were present in each region per hour. The Table 3.3 shows the structure of this dataset.

Table 3.3: Columns of `tot_ah_veh.txt` files

Column	**Description**
az_id	Anchor zone identification
line	Line number of anchor zone
column	Column number of anchor zone
day	Day
hour	Hour
sum	Number of vehicles encountred at time in anchor zone

By using the **Betweenness Centrality measures**, the objective was to use these two databases as a graph and from this, to verify which nodes are most important. For this, we create a dataset containing registers where each line contains pairs of vehicles that were within the radius of 100m at the same instant of time. It was possible using the *geosphere* library[3] in R environment. This procedure generated a dataset with 50 811 records about Mobile Century and 20 292 records about T-Drive. After that, the dataset about T-Drive was divided into seven datasets, corresponding to each day: 02 to 08, February. From these datasets created, we verify the pairs of vertices (vehicles) contacted during the course. The dataset `graph_mc` and `graph_t_drive` was created with the structure of Table 3.4:

Table 3.4: Graph table structure

Column	**Description**
vert1	ID of Vehicle 1
vert2	ID of Vehicle 2
edge	Weighted edge

When the vehicle contact happens, we increase the "edge" field. This contact generated a weighted graph. We perform the dataset conversion to a graph by using the *igraph* library of the R environment. After created the undirected and weighted graphs, the centrality was used to verify which vertex, per hour, have more importance. In the Mobile Century base, graphs were created per hour (from 10:00 am to 6:00 p.m.) and based on T-Drive, two days were used: on 03 and 04 February 2008 (Sunday and Monday, respectively) and created time graphs of these days. From this analysis, we verified which instants there were vertices with higher values of betweenness. It was verified then, which schedule of each day there was greater betweenness, and these schedules were used to verify the most critical vertices. After that, we verified the regions where these vehicles traveled on this day and time to then determine the regions with higher viability for the use of floating content.

3.2 Simulated data

In order to identify the most feasible anchor zones for the dissemination of the floating content using simulated data, we used the four techniques presents in section 3. We use these techniques in two regions: the I-880 highway, the "Nimitz Freeway" near Union City, California, USA; and the Beijing metropolitan area, China. We simulate the traffic of vehicles in these regions. The software used to create routes and traffic was **SUMO** [4]. Simulation of

[3] Available in https://cran.r-project.org/web/packages/geosphere/index.html
[4] Available in http://sumo.dlr.de/

Urban MObility (SUMO) is an open-source, highly portable, microscopic, and continuous road traffic simulation package designed to handle large road networks. It has tools for importing maps from various platforms, such as OpenStreetMap, for example, and APIs to remotely control the simulation. The version used was 0.30.0. The first step was to obtain the corresponding maps of these two regions. For this, we use the software OpenStreetMap [5].

In the region corresponding to the I-880 freeway, the region between latitudes of 37.52239 to 37.65714 and longitude of -122.1015 to -121.9882 was delimited, obtaining the region according to Figure 3.6.

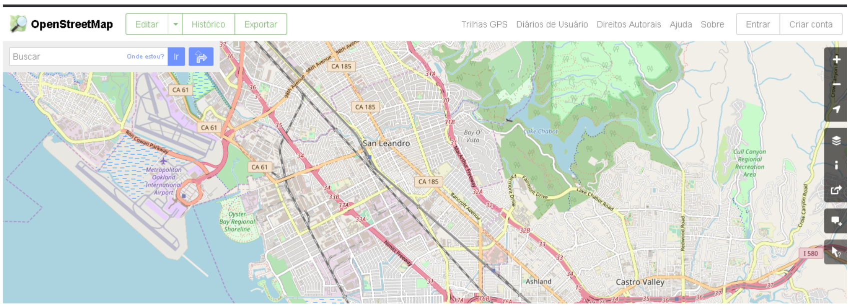

Figure 3.6: Region corresponding to the I-880 freeway

As the region of interest was only the I-880 highway, the JOSM software [6] (Java OpenStreetMap Editor), version 12921 (Java 1.8.0_144) was used to remove from the map all regions that did not match this freeway. Figure 3.7 shows the results.

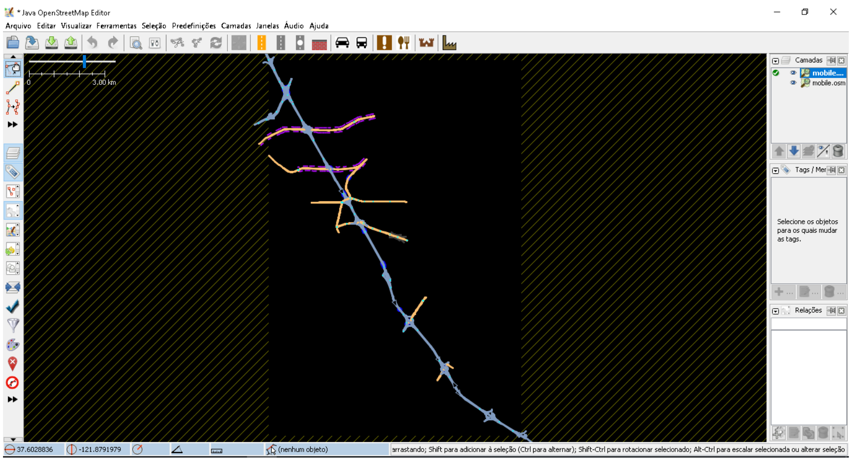

Figure 3.7: I-880 after the use of JOSM

Concerning the metropolitan area of Beijing (Figure 3.8), we use the region delimited between latitudes of 39.3 to 41.0589 and longitudes of 115.0093 to 118.0019.

[5]Available in https://www.openstreetmap.org/
[6]Available in https://josm.openstreetmap.de/

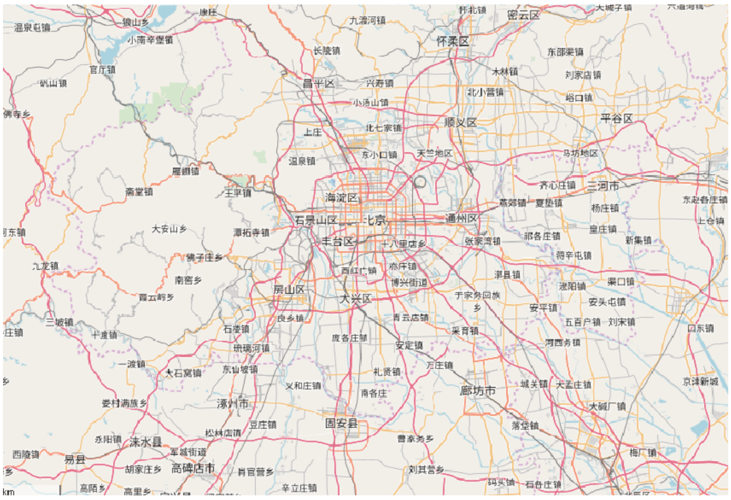

Figure 3.8: Beijing Metropolitan Area

After the regions were delimited, the maps were exported and converted to the format compatible with the simulator SUMO. Then, using the script *randomTrips.py* (available in Sumo) written in the *phyton* language, simulated traffic was generated for each of these regions in five hours, wherein the region corresponding to the I-880 freeway, traffic with 18 thousand vehicles, while in the Beijing region, we generate traffic of 100 thousand vehicles.

We store the results of the simulation of these two regions in files, and we convert them to text format (using the R environment through the libs "XML" and "data.table" for conversion). We depicted the files content in Table 3.5. Each row contains the instant and the position (latitude and longitude) of the vehicle and the speed. Each vehicle generated a log in intervals of 3s.

Table 3.5: Columns of `Vehicles file.txt` files

Column	Description
vehicle_id	Vehicle identification
instant	Instant of simulation, in second
longitude	Longitude coordinate, in degrees
latitude	Latitude coordinate, in degrees
speed	Vehicle's speed, in m/s

To verify if the size of the anchor zone interferes with the choice of the best technique to identify the most viable anchor zones for the dissemination of floating content, in this paper, we define that the anchor zones would be squares of five sizes: 200m, 400m, 600m, 800m and 1000m.

Anchor zones smallers or larger than those used did not present satisfactory results, which led to discarding them. The variation of 200m in each size was the one that presented the best benefit-cost ratio so that could verify if the size of the anchor zone interferes with the choice of the best technique.

In this way, we divided the I-880 and Beijing regions into grids, with the height and width of each of these five defined sizes.

3.2.1 Application of Techniques

With the anchor zones defined, for each of the sizes defined in each of these bases, the four techniques mentioned above were applied: Flow of Vehicle; Betweenness and Degree centrality; and Hybrid.

The first technique, *Flow of Vehicle*, calculated the average number of vehicles that each anchor zone had at each moment and identified the ten regions where there was a higher flow of vehicles.

In order to use the Betweenness and Degree centrality measures, it was necessary to convert the bases of the I-880 freeway and the Beijing metropolitan area into graphs. For this, we created an R script that verified the encounters of vehicles at each moment occurred in each anchor zone. Every hour, we verify the amount of each pair of vehicles, and we increase the edge field according to each encounter. This proceeding generates a weighted graph. The table 3.6 shows the structure of these generated graphs.

Table 3.6: Graph table structure

Column	**Description**
vert1	ID of Vehicle 1
vert2	ID of Vehicle 2
edge	Weighted edge
az_id	ID anchor zone

In the *Betweenness centrality measure*, with the database converted into a graph, we verify the betweenness value of the nodes belonging to the anchor zone. We use the regions were found the vehicles with the highest betweenness values. As the interval of simulation corresponds to five hours, we calculate the average of these betweenness values to identify the ten anchor zones with the highest values.

Similar to the second one, in the *Degree centrality measure*, with the database converted into a graph, we calculate the degree centrality measure. In each anchor zone, we verify the degree value of the nodes belonging to this anchor zone. As the simulation interval corresponds to five hours, we calculate the averages of the degree value of each anchor zone to identify the ten anchor zones with the highest degree values.

The fourth technique, *Hybrid*, multiplied the betweenness centrality value by the number of vehicles in each anchor zone and identified the ten regions that presented the highest values.

After identifying the anchor zones of each technique, the simulations were done using the *Veins framework* [7], wherein these selected regions, a message was generated at each hour, containing the anchor zone identification and routed to vehicles in its range. The Veins is an open-source framework for running vehicular network simulations. It runs over two well-established simulators: *OMNeT++* [8], an event-based network simulator, and *SUMO* [9], a road traffic simulator. It extends these to offer a comprehensive suite of models for Inter-Vehicle Communication (IVC) simulation.

The vehicles, when received messages, if they were in the same anchor zone, consume, store and forward the message. The floating content remained active in that region until vehicles are receiving and forwarding the content. At the end of each simulation, the average time that each Floating Content remained alive and the number of vehicles that consumed the floating content. In this way, it is possible to verify which technique is the most effective to detect viable regions for the dissemination of floating content.

For this, we modified the sources of the Veins framework [10], adding *ID_AZ* fields in messages and vehicles. With this, through parameters in the *omnetpp.ini*, we generate the regions of traffic, set the time of this floating content, the latitude and longitude values, and the size that each anchor zone. During the simulation, at each moment, it is possible to identify the vehicle' anchor zone. If one vehicle is the first to enter a selected anchor zone, it creates the floating content with the anchor zone ID. This floating content is then transmitted, and the vehicles within reach receive the same. If at the instant it receives, they are in the same anchor zone, it consumes, stores and forwards the same.

In this way, the floating content remains alive until the moment it has vehicles. We repeat this cycle every 3600s, i.e., we generate five floating content in each simulation. The generation of more than one per simulation was necessary because we observe that some regions became more active during the simulation, that is, some of them, by disseminating the floating content at the beginning of the simulation, it disappears in a few seconds. When we disseminate floating content during the simulation, we observe that it remained alive for a longer time, proving in fact that it was a viable region to the dissemination of floating content.

At the end of each simulation, the results were analyzed by *R* to obtain the following data: the time which each Floating Content was alive in each region, and the number of vehicles reached. These results were analyzed by the R, through the *omnetpp* library, these results were analyzed to identify which technique presented the floating content for a longer time

[7] Available in https://veins.car2x.org/
[8] Available in https://omnetpp.org/
[9] Available in https://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883
[10] Available in https://github.com/SensorNet-UFAL/anchorzoneInVaNETs

and reached more vehicles.

In this chapter, we present the techniques used to identify the most viable anchor zones for the dissemination of floating content used in this work. In the next chapter, we will present the implementation of these techniques.

Chapter 4

EVALUTIONS AND RESULTS

IN this chapter we will present the results of the techniques used in the chapter 3, both with real and simulated data. The results of the data obtained from real bases using *Grid Position* and *Betweennes centrality* techniques are presented in section 4.1 and the results obtained from the simulated bases using the techniques *Flow of vehicles, Betweenness and Degree centrality measure* and the *hybrid*, are presented in section 4.2.

To perform the analysis, we use the R platform version 3.4.1 [1]. The regions of the I-880 highway, the "Nimitz Freeway" near Union City, California, USA; and the Beijing metropolitan area, China; were analyzed using the techniques presented in Section 3. The results using real data will be presented in the section 4.1 and utilizing simulated data in the section 4.1.

4.1 Real data

Using **Flow of Vehicles** technique, we verified the number of vehicles that were in each region per hour. For this, we define the number of vehicles that traveled per hour in each anchor zone. This generated the dataset `tot_ah_veh_mc` to Mobile Century and `tot_ah_veh_T` to T-Drive database. The Listings 4.1 and 4.2 show the minimum, maximum, medium, and median the quantities of vehicles in each anchor zone per hour.

Algoritmo 4.1: "Summary of Mobile Century"

```

> summary(tot_ah_veh_mc$sum)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0     5.0   104.0   258.9   412.0  2498.0
```

[1]https://www.r-project.org/

Algoritmo 4.2: "Summary of T-Drive"

```

> summary(tot_ah_veh_T$sum)
   Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
  1.000    1.000    3.000    6.383    8.000 2166.000
```

The suggested zones, illustrated in Figures 4.1 and 4.2, show the regions with the higher flow of vehicles in the Mobile Century and T-Drive respectively.

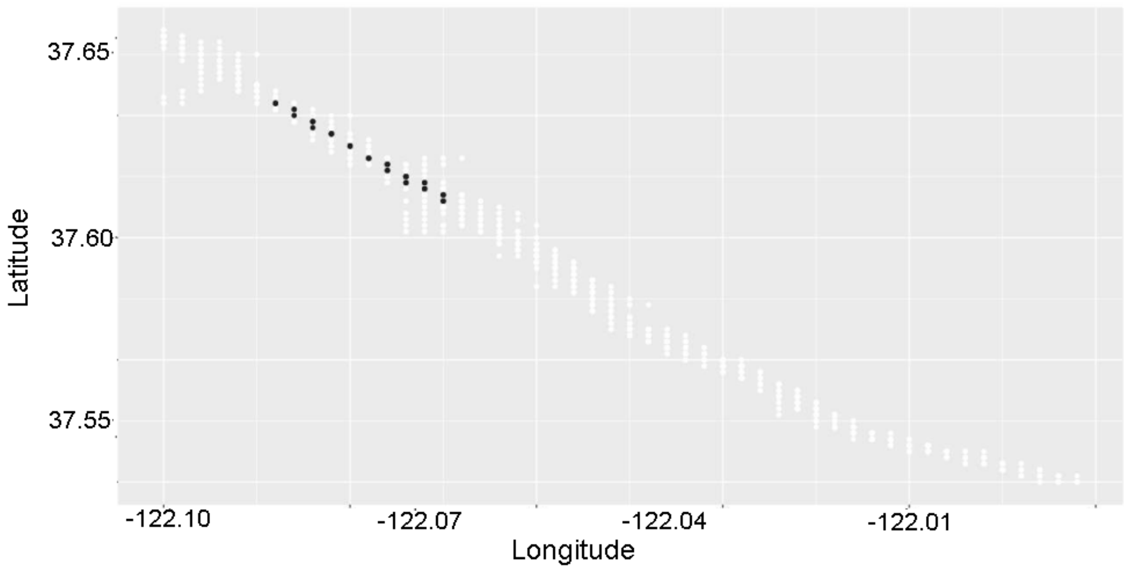

Figure 4.1: Mobile Century - Regions with the greater flow of vehicles.

We select as anchor zones the ten regions with the higher flow of vehicles in each base — the flow of vehicles of these regions where filtered per hour. Using the *lattice* package created a plot showing the number of vehicles per hour in each selected anchor zone in the Mobile Century and per day and hour in T-Drive. The Figures 4.3 and 4.4 show the results.

Figure 4.3 presents the number of vehicles in the ten selected anchor zones in the interval of 9h to 18h. We have in X-axis the time interval of 9h to 18h and in Y-axis the number of vehicles. The amount of vehicles is between 1 to 2498.

Figure 4.4 shows the number of vehicles in the ten selected anchor zones in the interval of 9h to 18h. We have in X-axis the interval of 0h to 23h, each color represents the day, and in Y-axis the number of vehicles. The amount of vehicles is between 1 to 2166.

Using the **Betweenness centrality measures** was verified the maximum values per hour in each base. To Mobile Century we generate nine graphs, which corresponds from 10h to 19h. To T-Drive was used just files of days 3 and 4, and each day generated 24 graphs,

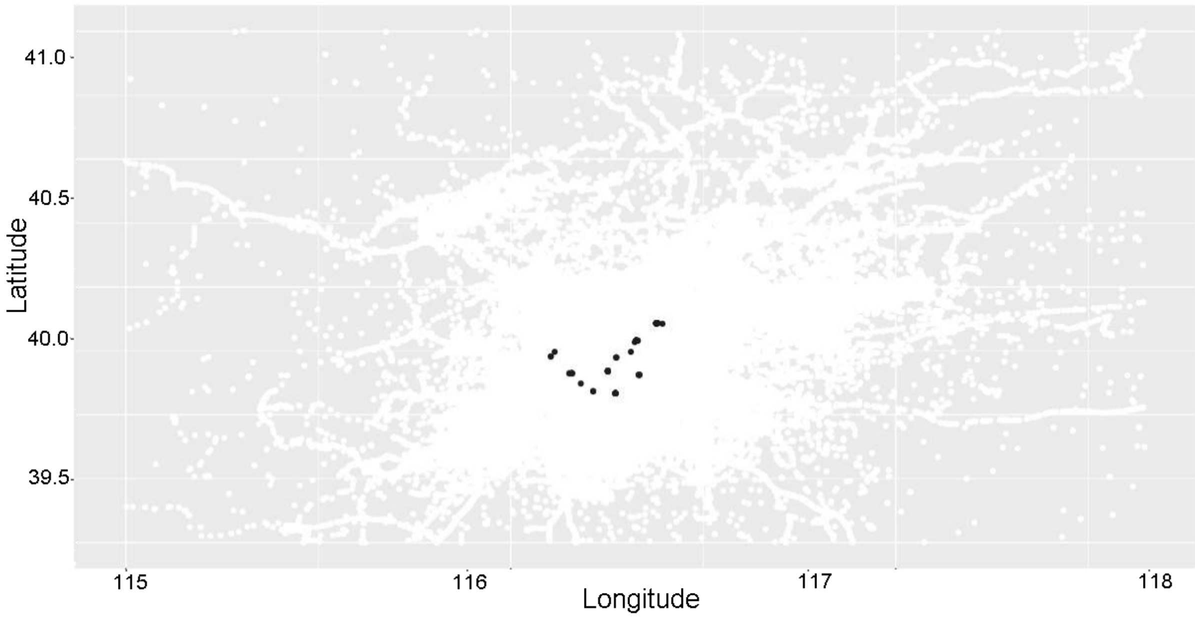

Figure 4.2: T-Drive - The regions which the greater flow of vehicles.

corresponding to each hour of these days. The listings 4.3, 4.4 and 4.5 show the minimum, maximum, medium and median the value of betweenness, calculated by *igraph* library.

Algoritmo 4.3: "Summary of Mobile Century"

```

> summary(mobCentury$max)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   2485    5647    6667    9255   10155   25744
```

Algoritmo 4.4: "Summary of T-Drive - day 3"

```

> summary(t_drive3$max)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0     1.0    11.0   164.6    77.5  1751.7
```

Algoritmo 4.5: "Summary of T-Drive - day 4"

```

> summary(t_drive4$max)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    2.75   17.75  107.33   68.00 1247.00
```

From these graphs, we calculate the value of betweenness centrality measure of each vertex. After that, we create the datasets *mobcentury*, *t_drive3* and *t_drive4*, where each record corresponds to one-hour betweenness summary information of minimum, median, average

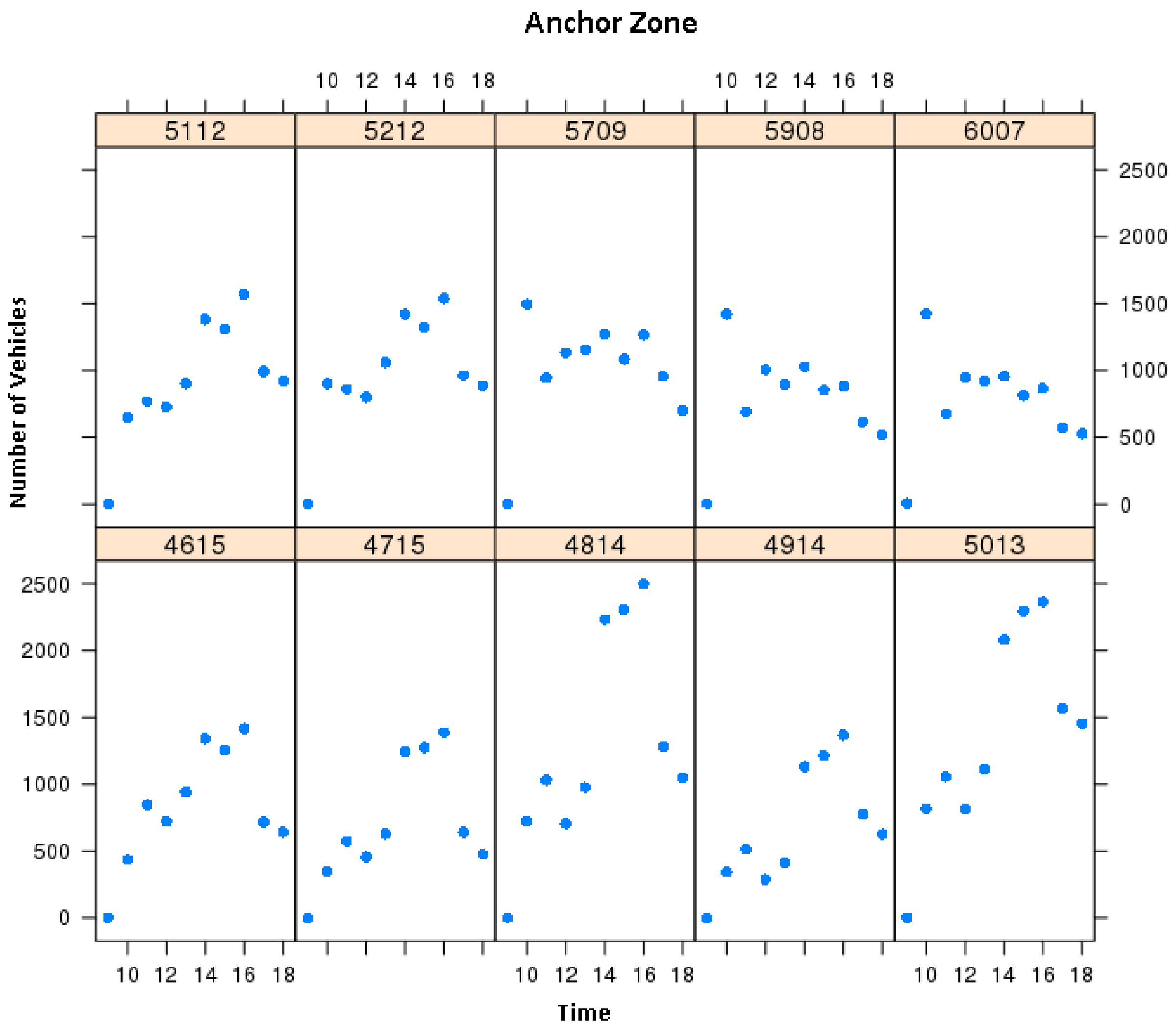

Figure 4.3: Mobile Century Top 10.

and maximum values. With datasets, we generate Figure 4.5. We verify that the time in each case, there was a higher value, and from this hour, verified which vertices have more importance. This relationship is presented in the Table 4.1.

Then, it generated the graph corresponding to the route that each of these vehicles made, in each corresponding base. Figure 4.6 shows the route of the ten most influential vehicles at 14h of Mobile Century. Figure 4.7 shows the route of the ten most influential vehicles from day 3 to 18h and from day 4 to 17h.

Concerning Mobile Century, we observe that using this strategy of **Centrality measures**, the region of vehicles with the most significant influence traveled is more significant than using the **Grid position**'s strategy. This behavior occurs because this dataset is a Freeway, where vehicles travel more distances than in the city since in 1 hour most vehicles tend to

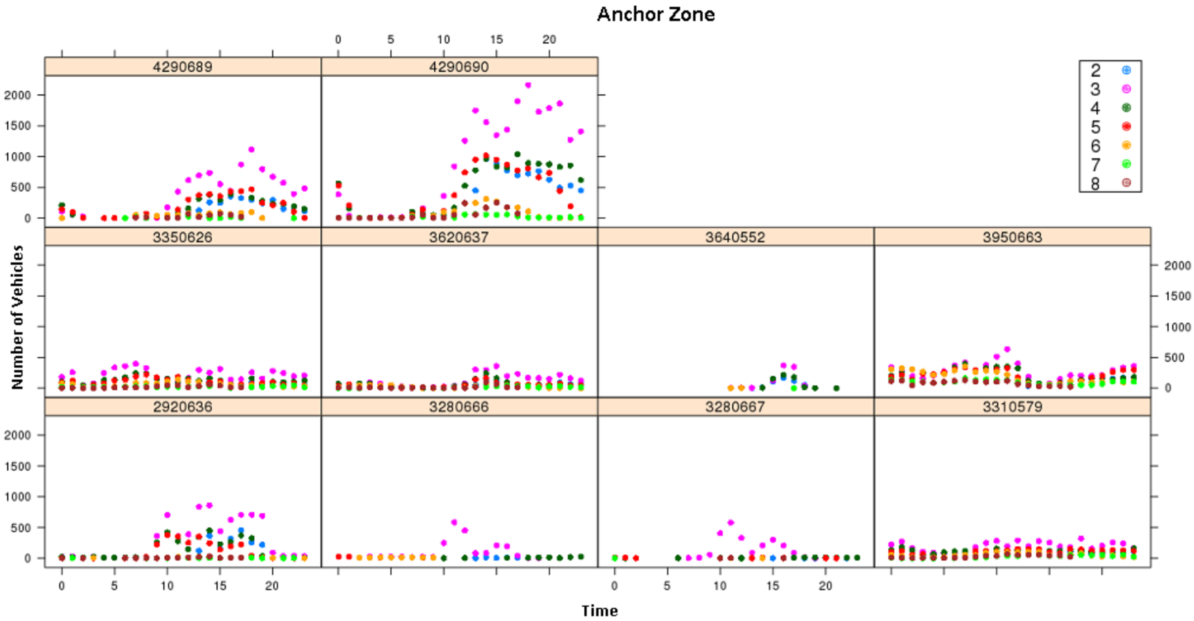

Figure 4.4: T-Drive Top 10.

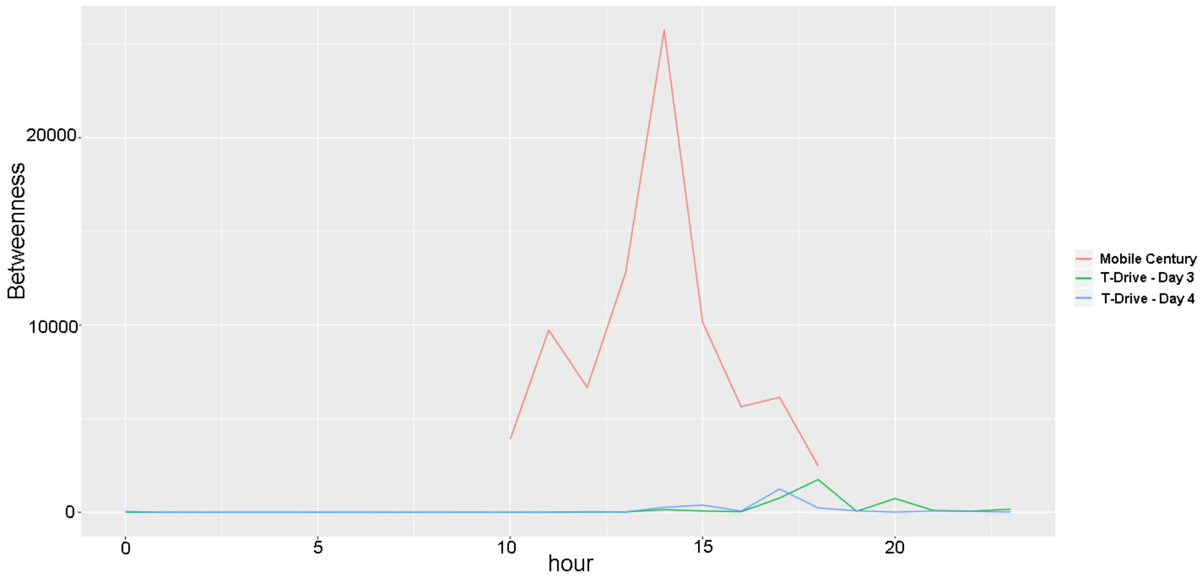

Figure 4.5: Betweenness

travel extensively. As these vehicles had a more significant influence in the communication with the others, they must have stayed longer in this region and at a slower speed than the others. We use the schedule of 14h in the second strategy because it was the schedule that had vehicles with higher values of influence, through the betweenness centrality measure. What can be observed through figure 4.3 that the schedules that the anchor zones with higher traffic, had more traffic in the schedules of 10h and 16h.

In contrast to the T-Drive base, the opposite behavior is observed, as the regions that the vehicles with the most influence traveled are in this time concentrated in two points of Beijing,

Table 4.1: Top 10 Vehicles

Position	MobCent-14h	TDrive-03 18h	TDrive-04 17h
1	747N	7090	5140
2	1319S	8688	5585
3	983S	2399	4481
4	792N	5977	3636
5	353N	4028	3077
6	567S	666	7984
7	1361S	86	4210
8	795S	7950	1605
9	1437S	3420	8205
10	1370N	8083	468

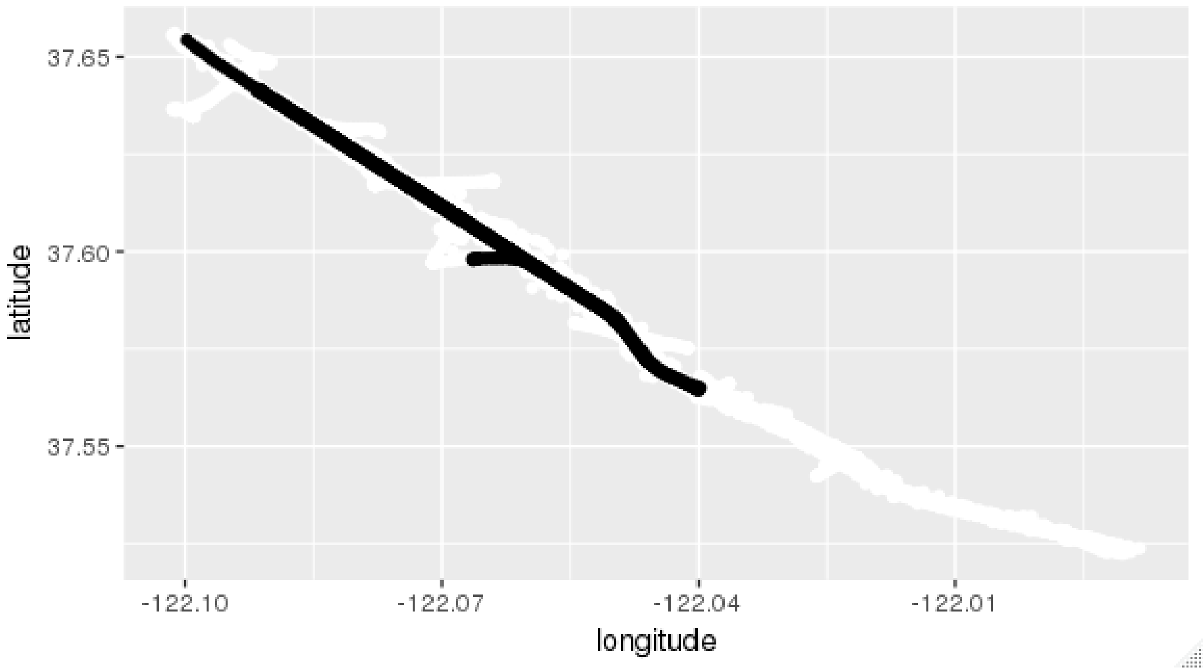

Figure 4.6: Mobile Century - regions where the most influential vehicles travelled at 14h.

and using the **grid position** strategy there was a greater spread of these regions. Because this base is the central region of the city, these vehicles during the time interval used went through small stretches at low speeds. The schedule of 18h on day 03 and 17h on day 04 were used in the second strategy because it was the schedule that had vehicles with higher values of influence, through the betweenness centrality measure. What can be observed through Figure 4.4 that the schedules that the Anchor Zones with higher traffic, had more traffic in the schedules of 10h and 16h. Using the strategy of **grid position**, 3 of the most populated AZ had more significant traffic on day 03 at 18h.

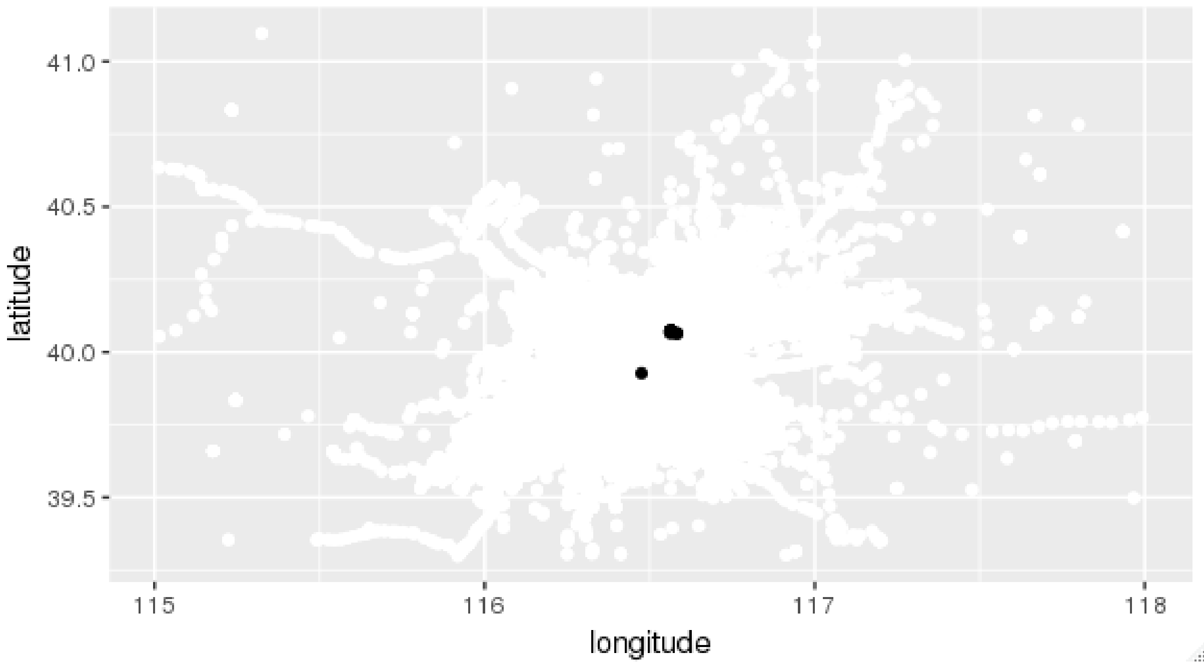

Figure 4.7: T-Drive - regions where the most influential vehicles travelled on day 3 at 18h and day 4 at 17h.

4.2 Simulated data

All simulations were performed under a computer model SGI Rackable Standard-Depth Servers, with 12 cores in 2 sockets of 2.53 GHz Intel Xeon Six-Core 5649, 1,333 MHz 144 GB DDR3 Memory (18×8 GB), 2× NVIDIA Tesla C2075 controllers, and Linux operating system. We use two regions with simulated data: The first is the stretch of 18km of "Nimitz Freeway - I-880", near Union City. We simulated the traffic of 18 000 vehicles, which corresponds to 8.37 % of the daily average of vehicles of this section, which is 215 000[2]. In the second region, it corresponds to the metropolitan region of Beijing. This region has 22 026km total length of roads and 57 million cars [3]. In real data, it used the real traffic of 8 388 vehicles. As this number in the simulations proved to be inferior to achieving consistent results, it simulated the traffic of 100 000 of vehicles, which corresponds to 1.75 % of the total of vehicles.

4.2.1 Identification of anchor zones viable per 4 techniques

The regions corresponding to the stretch of I-880 and the Beijing metropolitan area were divided into grids, varying the sizes in 200m, 400m, 600m, 800m and 1000m. These five sizes were used to verify if the anchor zone size interferes with the viability of the floating content in these regions. The four techniques were applied using the vehicles that transported through these regions during the time interval.

[2]Available in https://www.interstate-guide.com/i-880-ca/
[3]Available in http://www.ebeijing.gov.cn/feature_2/BeijingbyNumbers/

In the Flow of Vehicles technique, we verified the regions where there was a higher flow of vehicles in them, during the interval of time. In the Betweenness and Degree centrality techniques, in each of the anchor zones, we obtain the respective centrality values and the ten regions with the highest ones. We convert these bases into graphs, where each vehicle became a node and the meeting of them in an edge. At each encounter of a pair of vehicles, the weight of the edge was increased, making a weighted graph. In the Hybrid technique, the flow of vehicles in each anchor zone was multiplied by the Betweenness centrality value of this region, obtaining the regions with higher values.

4.2.2 Analysis of the results of the simulations

Using the Veins framework, modified by the authors so that the messages transmitted in the simulations had the identifier of the anchor zone original and the vehicles had "conscience" of which anchor zone they are at each instant of time. This floating content was shared by vehicles that were in the same anchor zone and remained alive while there was a flow of vehicles. In each simulation, the ten selected anchor zones had floating content generated every 3600s, totaling five floating contents per anchor zone.

In each technique, we run ten simulations, wherein each seed, of the movement of the vehicles, was modified. We obtain the confidence interval for the number of vehicles and the lifetime of the floating content in each anchor zone. For each technique, we use five different sizes of anchor zones, i.e., two hundred in each base, totaling four hundred simulations. We verified in which techniques the floating content remained alive longer, verifying if changing the size of the anchor zone will alter the most efficient technique.

Figures 4.8 and 4.9 show the results, where they present the lifetime of the floating content at the bases of I-880 freeway and Beijing, respectively.

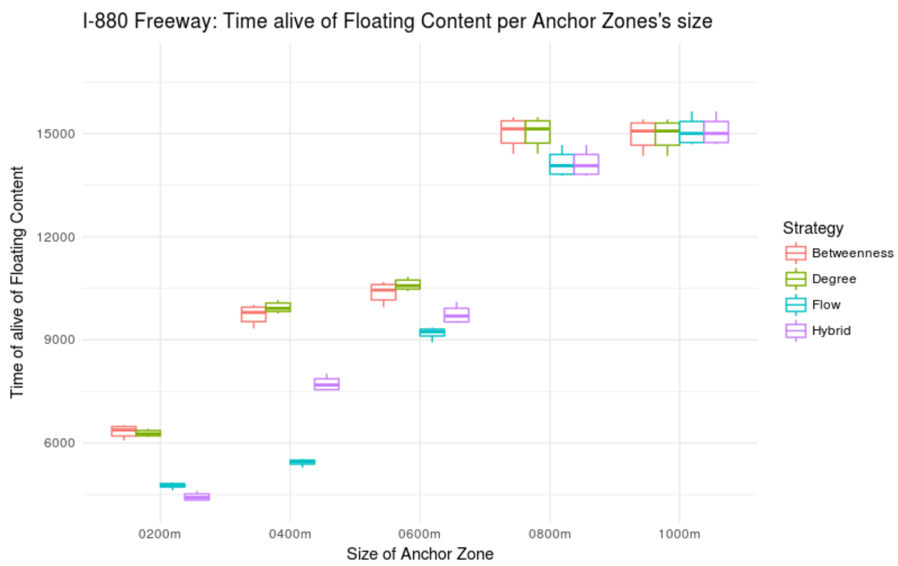

Figure 4.8: I-880 - Lifetime of Floating Content

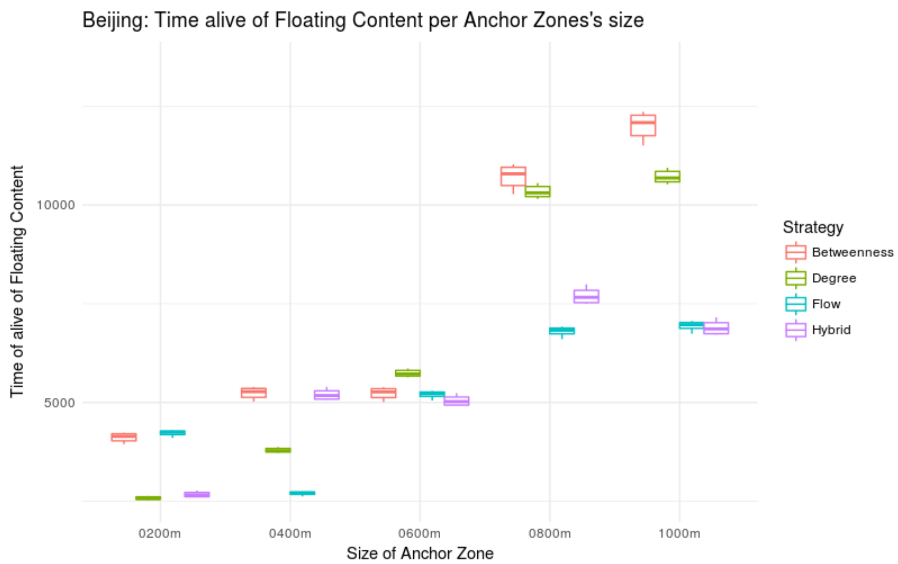

Figure 4.9: Beijing - Lifetime of Floating Content

In addition to obtaining the lifetime of the Floating Content, in each simulation, the number of vehicles that consumed the floating content was calculated, and the sum of the vehicles reached in each anchor zone, by technique, presented in the figures 4.10 e 4.11.

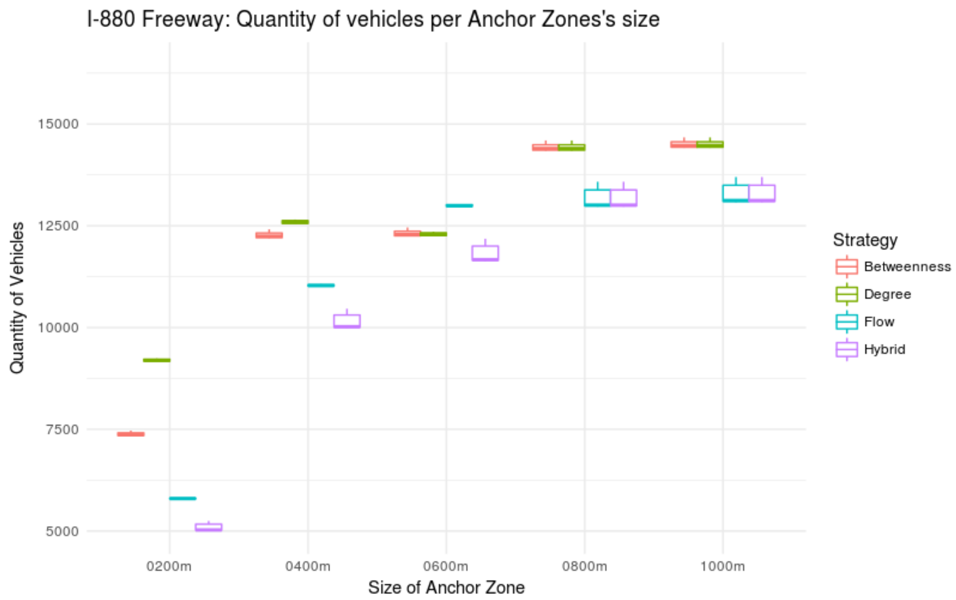

Figure 4.10: I-880 - Quantity of vehicles

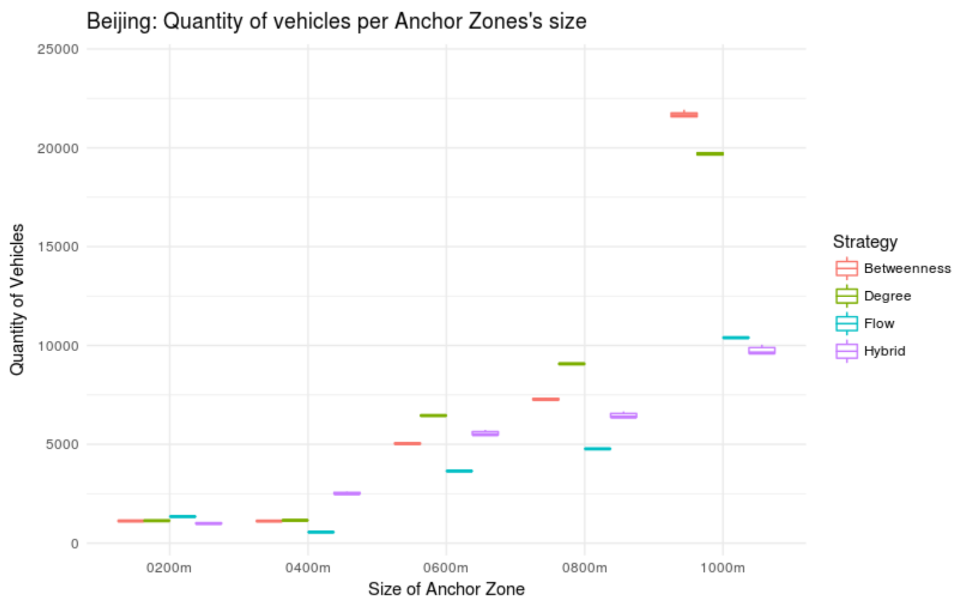

Figure 4.11: Beijing - Quantity of vehicles

This chapter presented the results of the identification of the most viable anchor zones for the use of floating content through the presented techniques. We conclude with the next chapter presenting the main impacts and contributions of this work, and the future steps.

Chapter 5

CONCLUSIONS

IN the first step, when worked in real bases, based on the results of these two databases, we observe that the T-Drive is much more sparse than the Mobile Century one. This behavior occurs because this region is larger than the latter.

Based on **Flow of Vehicle**, we observe that in the hours of 14h to 16h there was a higher flow of vehicles and **Betweenness Centrality Measure**, the time of 14h was the schedule where there were vehicles with the higher value of the betweenness. We deduce that these are the best time for using Floating Content. We observe that the regions located near the north of this section of the highway there were the anchor zones with the higher flow of vehicles. These regions, although of the peaks between 14h and 16h, during the time interval, a significant flow of vehicles was constant, being the regions most appropriate for the use of the Floating Content. The T-Drive base, although the data is 07 days, the highest concentration of records occurred on 03 (Sunday) and in the regions near the center of Beijing. Using **Flow of Vehicles** on day 03 from 4:00 p.m. to 9:00 p.m. and **Betweenness Centrality Measure**, day 03 at 6 p.m. and day 04 at 5 p.m. were the most viable times for the use of floating content. Through the strategy of **Betweenness Centrality measure**, we can identify the vehicles with the most considerable influence. Through them, it was possible to identify the most viable regions for the use of floating content. This behavior is possible because vehicles with this feature allow floating content to remain alive during its life cycle.

By using the real bases, we apply the techniques and identify only the regions that each technique showed the viable regions for the dissemination of the floating content. So that it could be verified, in fact, if these identified regions can keep the floating content alive for longer, in this way the technique is the most efficient, it was used simulators, where besides these two techniques, the measure of degree centrality and a hybrid technique, merging the techniques of flow of vehicles and the betweenness centrality. In simulated scenarios, in the regions identified by each technique, we disseminate the floating content and, in each technique, we verify the lifetime of the floating content, and the number of vehicles reached.

Analyzing the results of simulator data, we conclude that the Betweenness and Degree centrality techniques, both on the I-880 freeway and in Beijing bases, were those that identified the anchor zone in which the Floating Content remained alive for more time. Only when the size of the anchor zone was 200m in Beijing, that the flow of vehicles technique managed to be slightly better than the others.

Something to be reported, that on the I-880 freeway, when the size of the anchor zones was 800m or more, the selected regions of the Betweenness and Degree centrality techniques were the same. This behavior is because, since it was a stretch of about 18km, with anchor zones divided into these sizes, the two techniques ended up selecting the same regions. We observe the same behavior for the flow of vehicles and hybrid techniques, whose anchor zones were the same for both techniques in these sizes.

In Beijing the Betweenness centrality measure was the best in most scenarios, being overcome by Degree centrality only when the anchor zone size was 600m and by the flow of vehicles when it was 200m.

On the other hand, the Hybrid technique merged the flow of vehicle technique with the betweenness centrality technique, in both bases, was inferior concerning the two techniques of measures of centrality. It managed to overcome degree's centrality only at the base of Beijing when the size of the AZ was of 400m, being overcome by the technique betweenness in all the scenarios.

In the first technique the flow of vehicle was the one that presented the worst result because although it identified the regions where there was a higher flow of vehicles, it was not enough to keep the floating content alive the time that the other techniques maintained.

In both cases, the lifetime of floating content increased concerning the increase in the size of the anchor zone, and the increase was significant until the size of the anchor zone was 800m. In the I-880 freeway base, when the size of the anchor zone was 1000m, it was observed a reduction of the lifetime of the Floating Content, in the techniques of centrality measurement betweenness and degree. In the other two techniques (Flow of vehicles, and Hybrid) there was an increase of 800m to 1000m. Thus the regions detected by each technique were almost the same in this size. These results indicate that the size of 800m was the ideal size for an anchor zone for the dissemination of floating content. When applying the centrality techniques on anchor zones with 800m size, the identified anchor zones kept the floating content alive at about 94 % of the simulation time, which was 5h.

At the Beijing base, the anchor zones identified by the betweenness centrality kept the Floating Content live longer, and when the size of the anchor zone was 1000m, it remained alive at about 75 % of the simulation time, which was 5h.

About the number of vehicles that consumed the Floating Content generated in anchor zones identified by each technique, in both bases, no technique could be the best in all sizes of anchor zones.

At the base of the I-880 freeway, centrality measurement techniques were best when AZ size was 200m and 400m. Hybrid was better in all other sizes.

Concerning the Beijing base, the centrality techniques were better in the scenarios in which the size of the anchor zones was 600m, 800m and 1000m, being overcome by the technique of Flow and Hybrid, in 200m and 400m, respectively. We observe that about the number of vehicles, the Degree Centrality was superior to betweenness in almost all scenarios, losing to it only when the size of the AZ was 1000m. The flow of vehicles technique, although it presented a more significant number of vehicles that consumed floating content when the size of the AZ was of 200m, in the other scenarios was worse than the others.

When we increase the size of anchor zones, the number of vehicles increased proportionally in both bases. At the base of the I-880, anchor zones identified by the Flow and Hybrid technique reached 16 436 vehicles when anchor zone was 1000m in size. At Beijing base, the number of vehicles reached about 24 000 vehicles when the size of the anchor zones was 1000m, in the technique of betweenness centrality.

The target of this work was to verify the best strategy to identify the most feasible anchor zones for the dissemination of floating content. Compared to the strategy that identified the regions with the highest flow of vehicles, the centralities betweenness and degree were promising. The use of other strategies is useful to verify if these strategies are, in fact, the most effective.

Although the betweenness and degree centralities have similar results, degree centrality is better for one reason: not having to know the whole network like betweenness centrality. Thus, although both were effective, the degree is the most efficient!

The use of floating content allows city managers, entrepreneurs and other sectors of society to create communication channels for specific needs, where it is desired to reach a specific public in a specific region and period of time. The proposed solution aims be an answer to this problem. Cities and regions that have mechanisms to check how many vehicles they have on their streets, avenues and roads may be applying these techniques used in this work to determine the better anchor zones to use floating content.

Other strategies are also welcome to eliminate the problem that the use of the centrality betweenness presents: the need for knowledge of the whole network. Although the study has been very promising, it is necessary to move forward to find strategies that present results in real-time, where analyzes made in past traffic can allow prediction with this to identify the regions in real-time. Another approach can be to verify the influence of the size of the anchor zones on the lifetime of the floating content and may even identify the ideal size that an anchorage zone should have for better dissemination of floating content.

REFERENCES

Borgatti, S. P. (2005), 'Centrality and network flow', *Social Networks* **27**(1), 55 – 71.

Chaqfeh, M., Lakas, A. & Jawhar, I. (2014), 'A survey on data dissemination in vehicular ad hoc networks', *Vehicular Communications* **1**(4), 214–225.

Ciobanu, R.-I., Marin, R.-C., Dobre, C. & Pop, F. (2017), 'Interest spaces: A unified interest-based dissemination framework for opportunistic networks', *Journal of Systems Architecture* **72**, 108 – 119.

Di Maio, A., Soua, R., Palattella, M. R., Engel, T. & Rizzo, G. A. (2017), A centralized approach for setting floating content parameters in vanets, *in* '2017 14th IEEE Annual Consumer Communications Networking Conference (CCNC)', pp. 712–715.

Duarte, J. M. G., Kalogeiton, E., Soua, R., Manzo, G., Palattella, M. R., Maio, A. D., Braun, T., Engel, T., Villas, L. A. & Rizzo, G. A. (2018), 'A multi-pronged approach to adaptive and context aware content dissemination in vanets', *Monet* **23**, 1247–1259.

Everett, M. & Borgatti, S. P. (2005), 'Ego network betweenness', *Social networks* **27**(1), 31–38.

Fernando, N., Loke, S. W. & Rahayu, W. (2013), 'Mobile cloud computing: A survey', *Future Generation Computer Systems* **29**(1), 84 – 106.

Freeman, L. C. (1978), 'Centrality in social networks conceptual clarification', *Social Networks* **1**(3), 215 – 239.

Gerla, M. (2012), Vehicular cloud computing, *in* '2012 The 11th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)', pp. 152–155.

Hagihara, R., Yamasaki, Y. & Ohsaki, H. (2017), On delivery control for floating contents sharing with epidemic broadcasting, *in* '2017 14th IEEE Annual Consumer Communications Networking Conference (CCNC)', pp. 353–356.

Hyytiä, E., Virtamo, J., Lassila, P., Kangasharju, J. & Ott, J. (2011), When does content float? characterizing availability of anchored information in opportunistic content sharing, *in* '2011 Proceedings IEEE Infocom', pp. 3137–3145.

Khan, J. & Ghamri-Doudane, Y. (2015), Carrank: An information-centric identification of important smart vehicles for urban sensing, *in* 'Proceedings of IEEE 14th International Symposium on Network Computing and Applications'.

Kim, S.-K. & Yang, S.-B. (2017), 'Effective forwarding scheme for opportunistic networks based on refined contact probability and betweenness centrality', *Journal of Information Science and Engineering* **33**(1), 217–232.

Lee, E.-K., Gerla, M., Pau, G., Lee, U. & Lim, J.-H. (2016), 'Internet of vehicles: From intelligent grid to autonomous cars and vehicular fogs', *International Journal of Distributed Sensor Networks* **12**(9), 1550147716665500.

Lee, E., Lee, E.-K., Gerla, M. & Oh, S. Y. (2014), 'Vehicular cloud networking: architecture and design principles', *IEEE Communications Magazine* **52**(2), 148–155.

Magaia, N., Francisco, A. P., Pereira, P. & Correia, M. (2015), 'Betweenness centrality in delay tolerant networks: A survey', *Ad Hoc Networks* **33**, 284 – 305.

Manzo, G., Montenegro, J. S. O. & Rizzo, G. (2018), 'A deep learning mechanism for efficient information dissemination in vehicular floating content', *CoRR*.

Manzo, G., Soua, R., Di Maio, A., Engel, T., Palattella, M. R. & Rizzo, G. (2017), Coordination mechanisms for floating content in realistic vehicular scenario, *in* '2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)', pp. 589–594.

Massalino, F. & Aquino, A. L. L. (2018), Identification of anchor zones for floating content in vanets based on centrality measures, *in* 'Proceedings of the 33rd Annual ACM Symposium on Applied Computing', SAC '18, ACM, New York, NY, USA, pp. 2117–2124.

McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001), 'Birds of a feather: Homophily in social networks', *Annual Review of Sociology* **27**(1), 415–444.

Opsahl, T., Agneessens, F. & Skvoretz, J. (2010), 'Node centrality in weighted networks: Generalizing degree and shortest paths', *Social Networks* **32**(3), 245–251.

Ott, J., Hyytia, E., Lassila, P., Kangasharju, J. & Santra, S. (2011), 'Floating content for probabilistic information sharing', *Pervasive and Mobile Computing* **7**(6, SI), 671–689.

Rizzo, G., Mancuso, V., Ali, S. & Ajmone Marsan, M. (2018), 'Stop and forward: Opportunistic local information sharing under walking mobility', *Ad Hoc Networks*.

Sookhak, M., Yu, F. R., He, Y., Talebian, H., Safa, N. S., Zhao, N., Khan, M. K. & Kumar, N. (2017), 'Fog vehicular computing: Augmentation of fog computing using vehicular cloud computing', *IEEE Vehicular Technology Magazine* **12**(3), 55–64.

Virtamo, J., Hyytia, E. & Lassila, P. (2013), 'Criticality condition for information floating with random walk of nodes', *Performance Evaluation* **70**(2), 114–123.

Vázquez-Rodas, A. & de la Cruz Llopis, L. J. (2015), 'A centrality-based topology control protocol for wireless mesh networks', *Ad Hoc Networks* **24**, 34 – 54.

Este trabalho foi redigido em LaTeX utilizando uma modificação do estilo IC-UFAL. As referências bibliográficas foram preparadas no JabRef e administradas pelo BibTeX com o estilo LaCCAN. O texto utiliza fonte Fourier-GUT*enberg* e os elementos matemáticos a família tipográfica Euler Virtual Math, ambas em corpo de 12 pontos.