Centro de Enseñanza Técnica y Superior
With recognition of official validity of studies of the Government of the State of Baja California according to Agreement dated October 10, 1983

The properties and their effects of Tijuana Metropolitan Network

Thesis
to partially meet the requirements for obtaining the degree of
Master in Engineering Sciences

by

Belgica Jazmín Valencia García

Director:

Prof. Dr. Adán Hirales Carbajal
Centro de Enseñanza Técnica y Superior (CETYS University)

Tijuana, Baja California, México
2020

TITLE OF THESIS

Thesis/Application Project that to obtain the degree of
Master in Engineering Sciences

by

Belgica Jazmín Valencia García

and approved by the next Committee

Prof. Dr. Adán Hirales Carbajal
Thesis' Director

Prof. Dr. Ulises Orozco Rosas

Prof. Dr. German Eduardo Serna Valenzuela

Prof. Dr. Ricardo Martínez Soto
Postgraduate Coordinator in Engineering

Abstract of the thesis presented **by Belgica Valencia** as a partial requirement to obtain the Master of Engineering Sciences degree

The properties and effects of Tijuana Metropolitan Network

Abstract approved by:

Prof. Dr. Adán Hirales Carbajal
Thesis' Director

Tijuana is the 6^{th} largest city in Mexico. Due to its social - geographic characteristics, major avenues and arteries are often congested. Since 2006 and early 2018 onward, Tijuana's public transport fleets have undergone major renovation, by replacing small capacity vehicles, vans, and taxis, with larger capacity ones (busses). However, the city transport network has been analyzed in less degree. In order to understand its state, we analyze the transport network structural properties, node importance and robustness level. Among the major findings, we find that for a commuter traveling along the network diameter or radius requires in average 16 and 10 route transfers correspondingly. Four, out of nine, delegations concentrate the most important nodes. 80% of the Tijuana Mass Transit network can be disconnected with less 1.5% of nodes under random failures or target attacks.

Keywords: transport network, structural properties, spatial analysis, robustness.

Dedication

To my family for their unconditional love and care…

Acknowledgments

First and foremost, I would like to thank my thesis supervisor, Prof. Dr. Adan Hirales, for his support, advice and constant optimism on this path, thanks to his patience and dedication in my academic and personal growth. It has been an honor to be their student.

Thanks to CONACYT for the financial support I received for my education, thanks to CETYS University for the opportunity to be part of the Master's Program and thanks to all my teachers for sharing their knowledge.

I would like to thank my parents, and family for their support and guidance, not only during these months, but throughout my entire education.

Last but far from least, immense gratitude goes to my better half Alejandro. This is a journey we have done together. You have been there for me, and I can only hope to help you as much in your future endeavors.

Table of Contents

List of figures

List of tables

Chapter 1. Introduction

Tijuana, being part of the San Diego – Tijuana megalopolis agglomeration, has a complex transport network that adapts to the immediate needs of its population. Its sociodemographic characteristics are very dynamic, it experiences high volume of domestic, transborder, and transcounty traffic; high international exchange results in long lines at US-Mexico truck ports of entry; high migratory flows, among other characteristics (Ley, 2006) Immigration phenomenon have accelerated the city growth, with more than 50% of its settlements of irregular origin, mostly in periphery (Alegría and Ordoñez, 2005). Thus, the topological configuration of Tijuana is of regular and irregular human settlements.

To our knowledge the Tijuana Mass Transit Network (MTN) has not been analyzed. Previous studies have focus on understanding socio-cultural aspects and public policy, i.e., sustainable urban mobility (Avelar, 2014), quality of service (Castro, 2012; Millan, 2018), network supply (IMPLAN, 2017), relationship between urban form and the environment (Ramos, 2014), and public policy and governance of the transit network (Castro, 2012; Millan, 2018). In this study, our focus is on analyzing the network structural and node properties.

The Tijuana's transit network is used by 32% of Tijuana's population (1.64 million) (Millan, 2018). It changes constantly as urban attracts are born, re-locate, or die. Historically, a route is added as transport demand occurs in new human settlements (Castro, 2012). Often the routes are allocated to preexisting hubs and pass thought streets with preexisting routes, thus increasing congestion and redundancy in shared route segments (Ramos, 2014). The previous or any other transport policy must be measurable in order to assess its spatio-temporal implications. Since we lack historical data route maintenance and demographics, we analyze the transit network corresponding to one-time instance, April 2019.

Three classes of routes operate atop Tijuana's transit network, namely: MTN, which operate vehicles with an average capacity of 30 passengers; taxis, with an average capacity of 12 passengers; and Bus Rapid Transit (BRT), with capacities of 50 and 80 passengers. Of each class, there are 129, 157, and 22 routes accordingly. In this study, we analyze the MTN that result from merging 104 independent routes corresponding to the mass transport class. Twenty-five routes were left out the network, since these were incomplete. Route data was provided by the offices of the Municipal Public Transport Directorate.

The rest of this thesis is organized as follows. First, a brief review of the existing literature of network science concepts and robustness metrics applied to spatial analysis is given in chapter 2. Then, in chapter 3 the methodology for this work is explained. Chapter 4 summarize results and major findings. Lastly, a brief conclusion is then offered in chapter 5.

1.1 Defining the research problem

1.1.1 Justification

Public transportation is a major challenge in many cities around the world. The Tijuana metropolitan area, like many other metropolises, currently faces the challenges posed by the growth of their population and economy. The design and regulation of passenger transport networks it's a complex task (Ley, 2006).

In the literature peculiarities of the local transport network have been identified, which has led to its dynamic and disorderly growth. Thus, justifying the creation of new routes. For example, in areas with high population density duplicate routes have been introduced producing high levels of congestion (Ramos, 2014).

It is not enough to know about the existence of some elements in society that indicate obstacles with the current public transport service that directly impacts the economy, education, health and the environment (Rodrigue, 2013). Through previous works, characteristics have been observed that describe deficiencies and opportunities of the transportation structure in Tijuana, but it has only been a perception of its state. The purpose of this research is not to assess perception.

Cardozo (2019) talks about the vital importance of knowing the way in which the transport system interacts with the geographical space. The interaction promotes a transformation in the geographical area that is developed to adapt to its conditions and forms the topology that spatial networks adopt (segments, nodes, hierarchies) in metropolitan areas where it is coordinated with the infrastructure that regulates mobility.

Researchers around the world are currently studying spatial networks. The transport systems studied as spatial networks can have many advantages due is easier to identify general patterns occurring in systems while their understanding can be relevant and beneficial on the economy, on the environment, performance and reduce maintenance costs (Sasidharan, 2019).

Part of the study of determining the structure of the transport network is to be able to measure the impact of a failure on the structure of the network. One of the ways to measure the impact of some failure is through robustness, which has been studied at the same time as resilience through percolation theory by various authors for example (Newman, 2010, Barabási, 2016), where resilience is most focused on measuring the ability to return to its original state after an attack, while robustness seeks to measure the ability to survive random or deliberate failures (Wang, 2017).

Wang (2017) and Frutos (2019) study this impact on subway networks through different theorical robustness metrics that emphasize alternative routes, other with the length of the routes, as well as in the robustness of a random or targeted attack; other authors such as Barabási (2016) and Barthelemy (2011) they have studied the impact of cascading failure on different networks.

Despite the fact that the design of transport systems has been a well-studied and optimized problem in many cities. As this investigation is an exploratory work, it can be argued that it is a problem not properly studied in Tijuana to our knowledge. Given that there are no previous works that analyze the structural properties of the transportation network in Tijuana, this work begins with the effort to analyze this gap as the first step in characterizing the network and measure their robustness.

The results could have an impact on the management, representation and measurement of data as a support tool for those who promote urban development plans.

1.1.2 Research questions

This study opens before us the preliminaries to understand our network, according to the above, arise questions such as: What are the structural characteristics of the transport network?, How many transfers are required for the worst case and average mobility cases?, What are the most important nodes in the network?, The network is considered robust?, these are some of the questions that this thesis addresses for the first part of its development, is worth answering to this groundwork for future work.

1.1.3 Research objectives

This research is young, so we limit ourselves to evaluating the relationship between the transport network and the city's spatial area as a first approach:

- Acquire the route data set from the Tijuana MTN authorities.
- Model the route data set as a weighted DAG.
- Elaborate the data processing and visualization pipeline.
- Analyze the global and node level properties of the urban transport network.
- Study the robustness metrics of the urban transport network.
- Write the thesis and at least one congress publication.

Chapter 2. Literature review

2.1 Networks

Firstly, what is a network? A network is a collection of points joined in pairs by lines. In mathematical terms these points are known as vertices or nodes and the lines are known as edges.

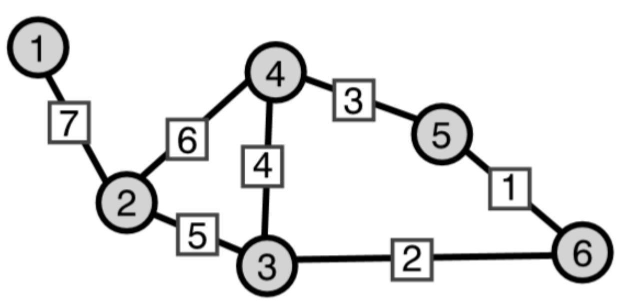

Figure 1. Network with 6 nodes (circles) joined with 7 edges (squares) (Rodrigue, 2013).

Many topics of interest from physics, biology and social sciences can be represented by networks (Newman, 2010). There are numerous systems in the real world that can be converted into nodes their elements, components or characteristics. The edges represent the relationship, the link between these nodes or as this connection requires to be called depending on the problem being explored (Figure 1).

The study of complex networks is of interest in research such as social networks, the internet, mobile networks, electrical networks, neural networks and an infinity of topics where it has a presence (Newman, 2010). The interest in these networks has been studied by countless authors who applied their approached in different ways: some study each of the nodes individually, others study only the links between the nodes, and still others focus on the interaction of both. The interaction between its nodes can be for different purposes, i. e. to understand their behavior until visualizing the complete panorama of the information which can be used later in decision-making in different scenarios according to study interests.

2.2 Spatial networks

Essentially a spatial network is described as any network where its nodes are located in a space with a metric (Barthelemy, 2011; Cardillo, 2006). A metric tells us that a variety of measurements and quantities can be calculated from the structure of a network that capture the particular characteristics of the topology of that network (Newman, 2010). In the real world, space is a bi-space or three-dimensional and the metric is the Euclidean distance.

The networks have edges that are restricted by some geometry. Spatial networks are made up of spatial data which can describe a spatial datum as everything that has a geographic reference associated with it, so that we can locate exactly where it happens within a map (Barthelemy, 2011).

According to Cardillo (2006), he indicates that many investigations that have concentrated on the topology of networks did not consider the spatial aspect. For some phenomena, the spatial distance between nodes represents a cost to pay. A important aspect in the behavior of the phenomenon is the number of edges that can be connected to a node has limitations such as physical space, for this reason the author proposes the study of spatial networks differently from complex networks.

Spatial networks can be categorized by their topology: in planar and non-planar. An example of planar networks are infrastructure networks that remain static such as road networks, trains, and other transport networks. Non-planar networks such as the airline network, the Internet network, social networks that are highly dynamic (Barthelemy, 2011).

2.3 Planar spatial network

The relationship between the term planarity and infrastructure networks is inevitable (Barthelemy, 2011), the roads, railways, among others are considered planar spatial networks where nodes and edges represent physical objects.

The importance of the distance between nodes it is an analysis factor that requires precision due to the costs that the distances between them would imply.

There are different tools that help determine the planarity of the graphs, Euler's formula, among others. In planar graphs, edges should not be crossed (Figure 2).

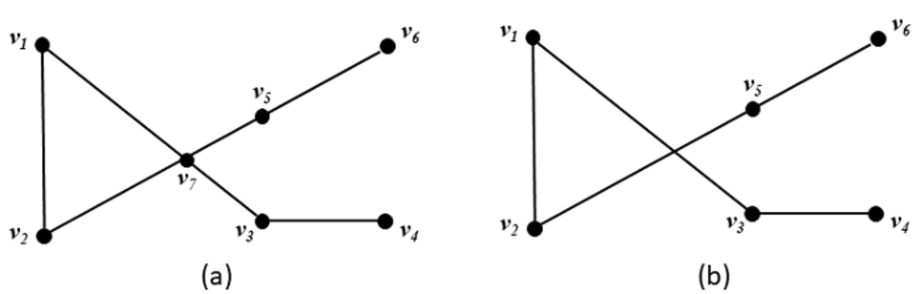

Figure 2. Network types: (a) planar and (b) non-planar (Chun, 2018).

We will focus as part of the scope of this work on exploring the planar space network due transport routes are static.

2.4 Network representation

Real-life geographic structures can be easily analyzed and interpreted by graph theory, allowing a simple structure of connected nodes and edges to be associated with transport networks (Cardozo, 2009).

Network analysis is also known as graph theory used to study forms and structures of transport networks abstracted from the real world (Figure 3), which compromise a large volume of data in conjunction with the use of various data analysis tools and techniques, due to the detail and the unique combination of properties in each network that is analyzed, even if they are simple descriptive measures or even more complex modeling structures (Rodrigue, 2013).

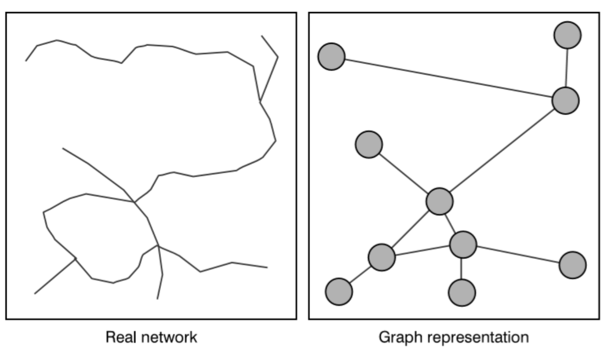

Figure 3. Graph representation of a simple real network (Rodrigue, 2013).

The literature distinguishes a variety of metrics that evaluate the transport network, i.e. in terms of its centrality, connectivity, at the node level, metrics at the level of the entire network, among others, involving in general terms a description of its nodes and relationships in order to detect similarities that guide their understanding.

2.5 Critical thresholds

The fraction of nodes that have to be removed from the network are relate to critical thresholds, such that the robustness curve is obtained after the nodes are removed, the size of the largest connected component of the remaining network is determined each time [1, V]. From the robustness curve is determined the critical threshold $f_{90\%}$ but the threshold f_c is not considered in this work.

The critical threshold $f_{90\%}$ is the first point at which the size of the largest connected component is less than 90% of the original network size, where the fraction of nodes that need to be removed.

Chapter 3. Methodology

The application of graph theory follows a descriptive analysis, using a set of algorithmic measurements. During the development of this thesis, different measurements are presented, focused on land transport routes (planar networks), capturing particular features of the network topology, in this study case, we concentrate these measures to characterize the spatial structure of Tijuana MTN.

3.1 The network model

The transport network is composed by a set of routes that aggregate geodesic coordinates and roads that routes traverse. Two kinds of roads are distinguished: bidirectional, which allow traffic in two directions, and unidirectional. Road capacity is not considered. Routes are preset and run in circular paths. A geodesic coordinate model an intersection, a location within a route, or a location within a curved road. From this point onward, the term geodetic coordinate is referred to as node and road as edge.

Formally, we model the transport network as a weighted directed graph $G = (\mathcal{V}, \varepsilon)$ with $\mathcal{V}$ the set of nodes and the ε under set of edges, $\mathcal{V} \neq \emptyset$. The number of elements in $\mathcal{V}$ and $\mathcal{E}$ are denoted by V and E . Node $n_i \epsilon\ \mathcal{V}$ is defined by the tuple (φ_i, λ_i), with φ_i and λ_i the geodetic coordinates (latitude and longitude). We will refer to a node by its order i in the set $\mathcal{V}$. G is assumed connected.

Nodes i and j are adjacent if $e_{ij} \epsilon\ \mathcal{E}$. A matricial representation of node adjacency is given by A a NxN square matrix with entries a_{ij} equaling 1 if the edge $e_{ij} \epsilon\ \mathcal{E}$, otherwise zero. The edge weight w_{ij} represents the length of the road segment between nodes i and j. Such quantity is measured in meters. The Haversine formula (Chopde and Nichat, 2013) is used to compute the great-circle length between nodes i and j.

The degree k_i of node i is the number of incident edges with the node defined as $k_i = \sum_{j\in\mathcal{V}} a_{i,j}$. For directed graphs, the node ***in-degree*** is de-noted as $k_i^{in} = \sum_{j\in\mathcal{V}} a_{j,i}$ and the ***out-degree*** as $k_i^{out} = \sum_{j\in\mathcal{V}} a_{i,j}$ (Boccaletti, Latora, Moreno, Chavez & Hwang, 2006).

The set of incident nodes of i is denoted as $\mathcal{N}_i$. The immediate predecessor and successor node sets are defined as $\mathcal{N}_i^{in}$ and $\mathcal{N}_i^{out}$ correspondingly.

Let d_{ij} be the shortest path (or geodesic) between nodes i and j, such that i, j. Further, let $L = \frac{1}{V(V-1)} \sum_{i,j\in\mathcal{V}, i\neq j} d_{ij}$ the ***average shortest path length*** defined as the mean geodesic lengths among all pair of nodes (Boccaletti, *et al.,* 2006).

Let n_{jk} be the number of all shortest paths across nodes j and k. Further, let $n_{jk}(i)$ be the number of shortest paths between nodes j and k that traverse through node i.

Lastly, the number of triangles in the graph is defined as $\delta(G)$ and triplets (possible triangles, chains of nodes) of two directed ties between three nodes are defined as $\tau(G)$. (Boccaletti, *et al.,* 2006).

In order to apply the robustness metrics, we model the transport network as a undirected graph and all of them were normalized to values from 0 to 1 to be possible to compare the results with other networks. The normalized robustness indicator $\overline{r^T}$ is obtained dividing by $r^T = \frac{ln(E_{max} - V + 2)}{V}$ with $E_{max} = \frac{V(V-1)}{2}$.

The effective graph resistance R_G study the number of parallel paths and their length $R_G = V \sum_{i=1}^{V-1} \frac{1}{\mu_i}$ where μ_i is the ith non-zero eigenvalue of the Laplacian matrix. The effective graph resistance R_G is normalized as effective graph conductance C_G.

The normalized version of $CC_i = \frac{2E_i}{k_i(k_i - 1)}$ where E_i is the number of edges connecting neighbors of node i and k_i is the degree of node i.

The average degree is normalized denoted as $\overline{k_G}$, dividing by $N - 1$ which is the maximal degree. See (Wang, 2017) for an exhaustive summary of network metrics.

3.2 Qualifying the network properties

Inadequate planning of urban transport systems leads to congestion, highly redundant routes, and from the perspective of route ownership, an unequal distribution of wealth. Thorough planning requires assessing whether applied transport policies result in a sustainable, resilient, and fair mobility infrastructure. A quantitative approach towards such end entails modeling the transport network, evaluating its structural and dynamic properties, and assessment of observed results.

3.2.1 Network and node level metrics

At the highest level of granularity, network global metrics analyze complex relations among data with the aim of characterizing and finding similarities between discovered features and known graph models. Finer grain metrics reveal node and linking properties of the directed graph i.e. node clustering degree, betweenness, among others. In this part, we focus on characterizing the topology at global and node level. For the reader interested in major concepts, metrics, and results see Boccaletti et al work on the study of structural and dynamic properties of networks (Boccaletti*, et al.,* 2006).

Table 1. Network and node level metrics.The reader is referred to references for the definition of pagerank and assortativity coefficient.

	Metric	**Expression**
Centrality	Degree	$\mathbf{k_i} = \sum_{j \in \mathcal{V}} \mathbf{a_{i,j}}$
	Knn	$\mathbf{k_{nn,i}} = \frac{\mathbf{1}}{\mathbf{k_i}} \sum_{j \in \mathcal{N}_i} \mathbf{k_j}$
	Pagerank	$\boldsymbol{pr}_\mathbf{i}$, see (Langville, 2005; Page, 1999)
	Betweenness	$\mathbf{b_i} = \sum_{j,k \in \mathcal{V}, j \neq k} \frac{\mathbf{n_{jk}(i)}}{\mathbf{n_{jk}}}$
	Closeness	$\mathbf{c_i} = \frac{\mathbf{1}}{\sum_{j \in \mathcal{V}; i \neq j} \mathbf{d_{ij}}}$
Network	Assort. Coeff.	$\boldsymbol{r}$, see (Newman, 2014)
	Diameter	$\boldsymbol{d} = \mathbf{max}_{j \in \mathcal{V}} \mathbf{d_{ij}}$
	Transitivity	$\boldsymbol{t} = \frac{\mathbf{3}\boldsymbol{\delta}(\boldsymbol{G})}{\boldsymbol{\tau}(\boldsymbol{G})}$

In traditional network analysis, influential nodes are those that either maximize data dissemination or prevent it, maximize the network robustness, among other properties. In the context of transport networks, influential nodes are often those with high degree, minimize proximity to other nodes, or tend to concentrate many passing routes through them.

Node importance based on the node degree assumes an uncorrelated network. That is, were node linkage is independent of the node degree (Boccaletti, *et al.,* 2006). A metric for correlated networks is the node ***average nearest neighbor degree*** or ***knn***, such criterion uses knowledge of its immediate neighbor's degree to compute a given node importance (Barrat, Barthelemy, Pastor, & Vespignani, 2004). Thus, hub nodes that connect to other hub nodes are more likely to be important. Knn does not distinguish if neighboring nodes are in-edge/out-edge nodes.

Another approach extends the previous two criterions by restricting the neighboring node set to contain in-edge nodes. Such metric is re-ferred to as ***Pagerank*** (Langville and Meyer, 2005; Page, Brin, Motwani & Winograd, 1999). The principal notion behind Pagerank is that important nodes are those with many in-edges from important nodes. The method iteratively gives an equal share of its current rank (importance) to all nodes it links too. To deal with small strongly connected components at each time step chooses, with probability alpha, the heuristic chooses to continue on a previously selected random walk or start a new walk at anew randomly selected location. Default values for alpha range between [0.8 and 0.9]. Pagerank and knn both assign greater importance to hubs that aggregate routes from other hub nodes. The major distinction between them is that in pagerank all nodes iteratively agree which nodes are designated important, while in knn only neighboring nodes do.

Metrics based on distance criterion evaluate the mean shortest path length to all reachable nodes or analyze in what ratio does a given node is on the shortest path between all shortest paths in the network. The first criterion is termed ***closeness centrality,*** whereas the later ***betweenness centrality*** [Boccaletti, *et al.,* 2006]. Nodes that minimize closeness centrality are important because they minimize the number of edges needed to get to them from any reachable node. These nodes are equivalent to the authority class nodes described previously. Closeness centrality requires the graph to be connected. The betweenness centrality identifies nodes that act as hubs (bridges or brokers) between connected components in the network. Removing these may result in a network partition. Hubs are important because they connect authoritative nodes. A betweenness value of on is indicative of a hub node.

Global network properties include assortative mixing, clustering coefficient or transitivity, and diameter. ***Assortative mixing*** characterizes the network based on node linking preferences (Newman, 2003; Foster, Foster, Grassberger & Paczuski, 2010). In an assortative network, nodes tend to connect to adjacent peers. Whereas, in a disassortative network low degree nodes are more likely to connect with high degree ones. Assortative mixing values are in the range of [0,1], with 1 an assortative network and 0 otherwise. ***Clustering coefficient*** or ***transitivity*** is a measure of association (Barrat, *et al.,* 2004). It evaluates how likely nodes that are connected form part of a larger connected component. Formally, it is a measure of the number of triad closures relative to the connected triples in the graph (Boccaletti, *et al.,* 2006). A network containing many loops will have a transitive ratio close to one, whereas a network without loops will have a value of zero. The network ***diameter*** is the maximum shortest path (***eccentricity***) of any node in the graph. It represents the shortest distance between the two most distanced nodes in the Directed Graph. From a scalability perspective, it is desirable that the diameter remains constant as the network size increases. Diameter assumes the graph is connected. Table 1 summarize the previously discussed metrics.

3.2.2 Network robustness metrics

3.2.2.1 Theorical robustness metrics

The most important theorical robustness measures are presented to show the current condition of the network topology to take alternative routes under random failures or deliberate attacks over the nodes and offer different aspects of robustness. The first criterion is termed **robustness indicator**, measures the alternative paths in a graph is, this measure increase when alternative paths are offered to reach a destination, and it decreases in larger systems, which are arguably more difficult to upkeep. The **meshedness coefficient** measure the potential number of cycles in a planar graph and considers the maximal number of faces in a planar graph. From a different aspect of robustness, the **effective graph conductance** study the number of parallel paths (i.e., redundancy) and their length between two nodes. **Global Average efficiency** quantifies the efficiency of transportation by the number of connections in the shortest path of any node in the network. In other terms the **average clustering coefficient** measure the connectivity of the whole network since they assess how the neighbors of any node are connected to with another. While the **average degree** obtains the number of average connections of a node, a higher connectivity can be thought of as more robust network. Finally **degree diversity** is another connectivity measure that computes the critical fraction of nodes that are necessary to be removed from the network in order to disconnect it, the network is more robust if more nodes have to be removed from the network to disintegrate it.

Table 2 summarize the previously discussed metrics where the expressions were normalized to scale the value in the interval [0,1] the results can be compared to networks in other cities.

Table 2. Network robustness metrics.

	Metric	**Expression**
Theorical measures	Robustness indicator	$r^T = \frac{ln(E - V + 2)}{V}$
	Effective graph conductance	$C_G = \frac{V - 1}{R_G}$
	Global Average efficiency	$EF\left[\frac{1}{d}\right] = \frac{2}{V(V-1)} \sum_{i=1}^{V} \sum_{j=1}^{V} \frac{1}{d_{ij}}$
	Average clustering coefficient	$CC_G = \frac{1}{V} \sum_{i=1}^{V} CC_i$
	Average degree	$k_G = \frac{\sum_{i=1}^{V} k_i}{V}$
	Degree diversity	$k = \frac{\sum_{i=1}^{V} k_i^2}{\sum_{i=1}^{V} k_i}$
	Meshedness coefficient	$M_G = \frac{E - V + P}{2V - 5}$

The theoretical robustness metrics capture two distinct aspects of the robustness networks: A first aspect deals with the number of alternative paths, suggesting that more alternative paths is more desirable as captured in $\boldsymbol{r^T}$, $\boldsymbol{CC_G}$ and $\boldsymbol{M_G}$ favors large networks. Second aspect deals with ‘‘resistance’’, suggesting that longer lines with no shorter alternative paths perform poorly, as captured in $\boldsymbol{C_G}$, $\boldsymbol{EF\left[\frac{1}{d}\right]}$, $\boldsymbol{k}$ and $\boldsymbol{k_G}$ focuses on the length of the routes and favors small networks.

3.2.2.2 Numerical robustness metrics

As opposed to the theoretical metrics the numerical robustness metrics are obtained through simulations under three strategies for node removal: random node removal, betweenness node removal and target node removal. The numerical metrics consider the MTN for random failures or targeted attacks the critical threshold $f_{90\%}$.

- Random node removal: the nodes are removed one by one chosen randomly from the network until the network is completely disintegrated from 4461 to 1 node. 30 simulations were run ensuring randomness in each simulation, the average of their simulations is calculated to obtain the final random set of nodes.
- Target node removal: the nodes are removed progressively in the network based on the highest importance of pareto fronts order until the network is completely disintegrated.
- Betweenness node removal: the nodes are removed by selecting the highest betweenness degree b_i in the network. First are removed the node with the highest betweenness degree and continue selecting and removing nodes in decreasing order of their degree.

3.3 The data processing pipeline

In order to facilitate reproducible and verifiable results the data set, metrics, data processing pipeline, and the project anaconda development environment are made available via the project site.

The transport network was built by concatenating 104 out of the 129 independent routes. The remaining 25 routes that were not included correspond to incomplete routes. The size of ν is of 20,411 nodes. Routes share up to 80% of nodes. Nodes that share a location or are close to each other, i.e. withing the street width, were merged. The size of ν reduced to 4461 non-redundant nodes. The transport network was digitally edited with JOSM Ver. 15238, an extensible editor for Open Street Map. The node merging process was done with JOSM. The resulting set of routes was stored in .geojson format and is available for download in the project site (https://github.com/ahiralesc/tijuanavehiclenetworkanalysis.git).

The data processing pipeline consists of the following phases: *data preparation phase*, which load and merges the transport routes. The greater circle distance between nodes is also computed during this phase. Nodes and edge data are modeled as dictionaries; *graph preparation phase*, the dictionaries are transformed into a directed graph; *metric application phase*, user selected global and node level metrics are applied; and *statistical analysis* and *data visualization*. Statistical analysis was done by using simple descriptive statistics. All processed data is stored in CSV (Comma-Separated Values) and pickle format.

The merged transport network covers nearly 637 km^2 of the metropolitan area. The network aggregates regular lattices, modeling city sections that were regulated, and irregular geometries, modeling human settlements of unregulated origin (Figure 4).

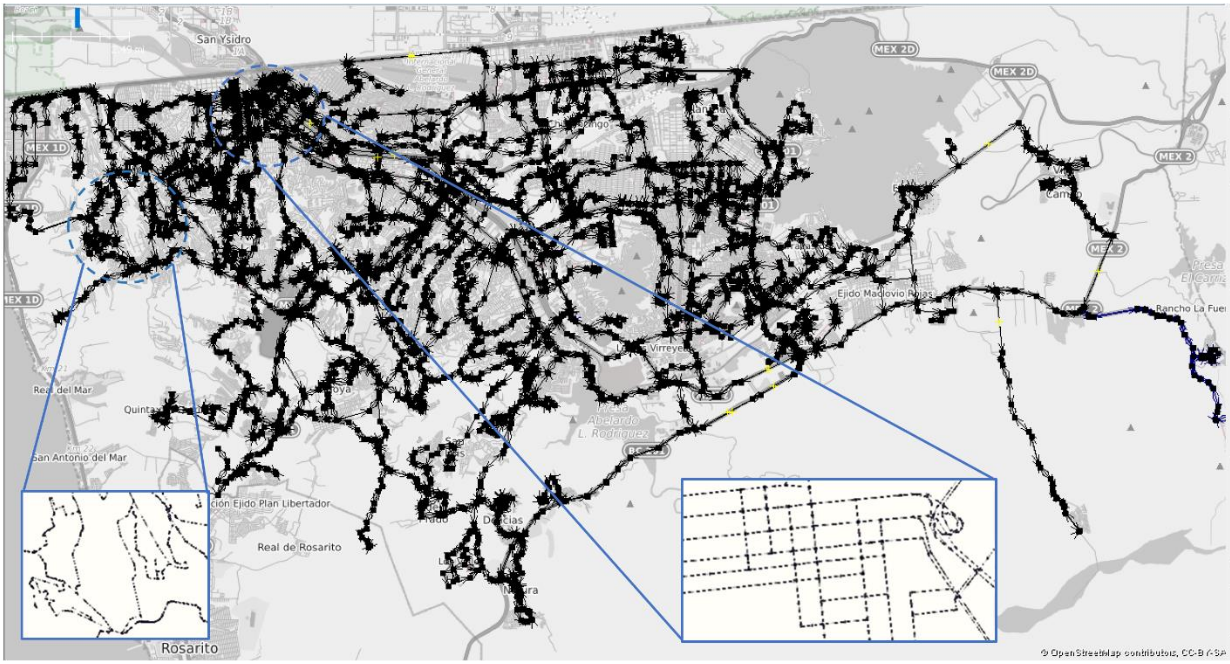

Figure 4. City of Tijuana mass transport network.

Node importance is evaluated by categorizing node metric into three classes, namely: *degree* based (deg); based on *distance* and neighborhood based (dist); and *rank* base (rank). Metrics in the first class are $\overline{k_i}, k_i^{in}, k_i^{out}, k_n n$, in the second b_i and c_i, and in the third pr_i. Metrics in each class are equally weighted and linearly combined. The non-dominated sorting technique is used to extract points in each Pareto (Coello, 2006).

Chapter 4. Results

4.1 Network properties

Network level properties are summarized in Table 3. A diameter d, or maximum eccentricity of 190 was obtained. In average a commuter traveling along the diameter requires 16 route transfers. Such quantity is four times larger than that in Rio de Janeiro's bus transfer network (Pumar, 2018). The number of transfers can be minimized to 10 however it increases the length of the route. The minimum eccentricity (or radius) contains 99 edges. A total of 4461 pairs of nodes length equals the radius. In average 10 bus transfers are needed to traverse along a path with length equaling the radius. Once again, the number transfers can be minimized to 6 at the expense of increasing the length of the path. The Tijuana 2017 municipal development plan report that in average 4 bus transfers are made by 46% of the MTN commuters (IMPLAN, 2017). No statement is made regarding the length of the trip. An additional two transfers would be required if the length of those trips equals the radios. Note the above analysis assumes bus transfers occur locations where routes intersect.

The MTN assortativity r is of 0.37. Such value suggest that the network is not centralized in a few large hubs and that these are interconnected. Hub nodes often had a degree of 7-9 and constituted approximately 1.2% of the network nodes. The network transitivity t is low (0.01). In general, a network with poor transitivity have large path lengths (Shanmukhappa, 2018). These results confirm that the diameter is large.

Table 3. Global network properties.

$\boldsymbol{r}$	$\boldsymbol{d}$	$\boldsymbol{t}$
0.37	190	0.01

4.2 Node centrality

Figure 5 displays the weighed node importance scores according to the applied linearization. Nodes with less importance cluster near the origin, whereas those with highest importance tend to the unit origin. Results show that most of the nodes cluster near the origin with a range of [0,0.4]. A cloud like pattern is formed withing this region. However, as the nodes span out they begin to lean towards the degree coordinate axis. Suggesting that node importance is skewed by the node degree. We found a total of 20 non-dominated nodes in the first Pareto front. These correspond to the most important nodes in the transport network (according to the proposed technique).

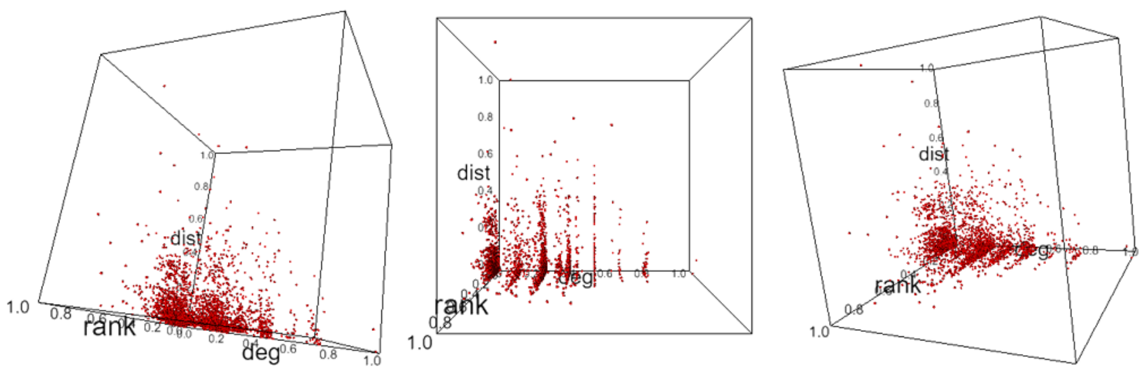

Figure 5. Linearization of metrics into 3D space.

The 10 most important nodes are depicted with blue markers. Nodes color in red correspond to the later 10 nodes with second most importance. These two sets of nodes are in the first Pareto front (Figure 6). The first set of nodes are concentrated in 4 of the 9 delegations the city is composed of, namely: La Mesa, La Presa Abelardo. L. Rodriguez, La Presa Este, and San Antonio de los Buenos. Areas containing these nodes are magnified for better visualization. Intuitively, one would expect large intersections to concentrate many routes, i.e. areas 7 and 3. Visual inspection of the remaining areas show that large route concentration occurs in main pathways in regions with lattice like topology.

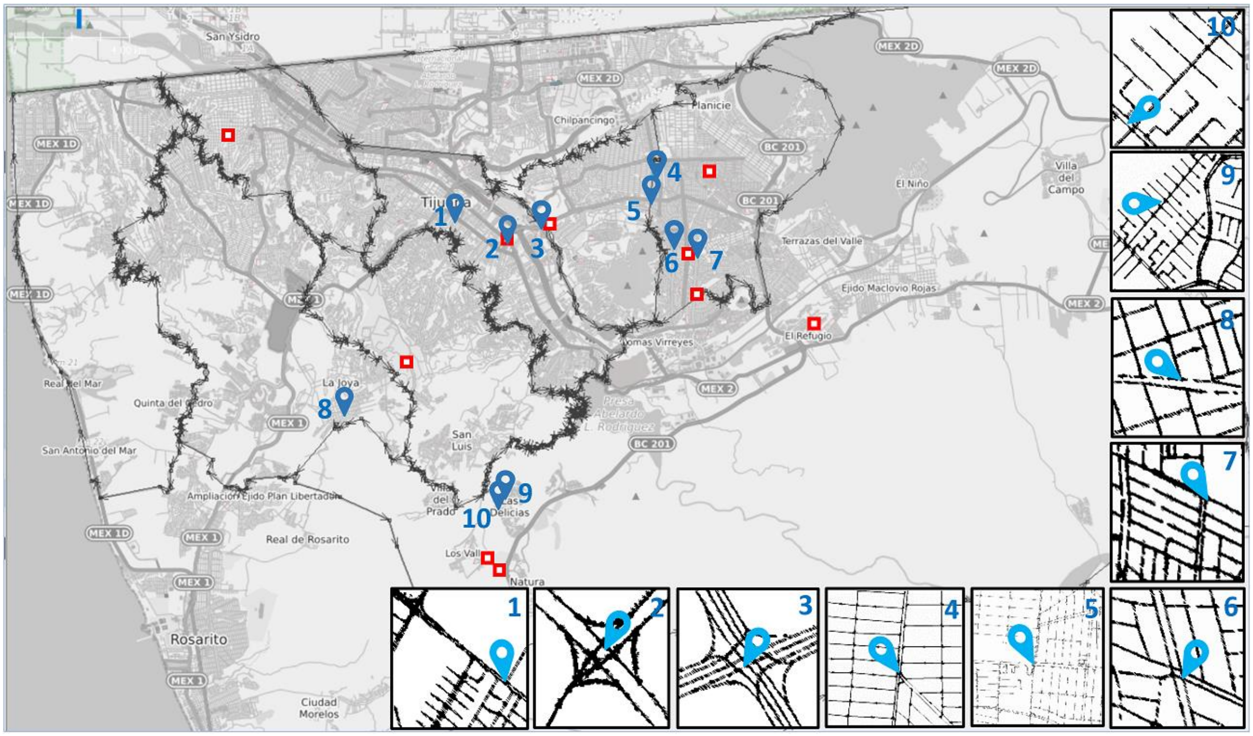

Figure 6. First Pareto front nodes.

Table 4 lists the range of mean measures corresponding to the first 10 Pareto fronts. Quantitatively, nodes in the first four Pareto fronts are very similar. Thus, intuitively, they share a similar level of importance and cumulatively increase the number of nodes in this class to 164 nodes.

The mean path length from any reachable node to any node in the Pareto 1 front is 54. Such length is smaller than the radius by 45 edges. Nodes in the front 1 may be good candidates for bus stops and attract a greater share of trips increasing the centrality of the network (Scheurer, 2007). Nevertheless, their feasibility needs to be assess as these locations currently model locations with high route intersection.

Mean betweenness centrality is low. Approximately 6% of all shortest paths go through a node in Pareto 1 and 2 nodes sets. A maximum betweenness centrality of 0.21 was found. However, this occurs in only one location. Even if we extend the range from 0.1 to 0.21 only 1% of the nodes categorized as important have this degree of centrality.

Table 4. Mean metrics values

Front	k_i	b_i	c_i	$k_{nn,i}$	pr_i
1	5.350	0.065	0.017	2.950	0.001
2	4.974	0.060	0.017	2.893	0.000
3	4.905	0.053	0.017	2.734	0.000
4	4.943	0.047	0.017	2.421	0.000
5	4.721	0.040	0.016	2.197	0.000
6	4.375	0.036	0.016	2.270	0.000
7	4.264	0.032	0.016	2.267	0.000
8	4.151	0.030	0.016	2.181	0.000
9	4.020	0.026	0.016	2.104	0.000
10	4.010	0.023	0.016	2.021	0.000

A mean clustering coefficient is not significantly high, 0.017 and 0.016, this might be explained because: routes tend to run in parallel along main avenues; they transit through roundabouts; or they follow circuits in suburbs.

4.3 Robustness Analysis

The robustness of networks reflects the ability of a network to maintain its general structural properties connected when it faces random failures or target attacks like removal nodes or edges, possible solve by offering alternative routes in some cases.
The theorical robustness metrics described in section 3.2.2.1 were computed for MTN and briefly analyzed, see all results in table 5. Then the numerical robustness metrics under random failures and targeted attacks were obtained through simulations see in section 4.3.2.

4.3.1 Theorical robustness metrics results

Table 5. Theorical robustness metrics in Tijuana MTN.

Metrics	r^T (NLRI)	C_G (NLEC)	$EF\left[\frac{1}{d}\right]$ (NLAE)	CC_G (NLCC)	k_G (NLAD)	$\frac{1}{K}$ (NLDD)	M_G (NLMC)
Normalized data	0.4097	0.0002	0.0267	0.0100	0.0006	0.2760	0.0825

Figure 7. Two types of radar diagrams of theorical robustness metrics results.

The robustness indicator $\overline{r^T}$ = 0.40, this value has relative high robustness level indicating a Tijuana MTN with many alternative paths between any pair of nodes which that confirm the low transitivity with a large path lengths.

The effective graph conductance is C_G = 0.0002, thus, according to this value, the network is not robust due to this measure accounts not only the number of alternative paths but also the length of each alternative path, favoring networks with the smallest length of the shortest paths like star topology which is not the case of Tijuana MTN.

The global average efficiency $\boldsymbol{EF}\left[\frac{1}{d}\right]$ = 0.0267, this value describes the network efficiency is quite poor, that can be explain due this measure favors the opposite topology of Tijuana MTN.

The average clustering coefficient $\boldsymbol{CC_G}$ = 0.0100, the value is low, thus it is quite easy to disintegrate Tijuana MTN into components due its poor connectivity.

The average degree $\boldsymbol{k_G}$ = 0.0006, thus, according to this value, the network is not robust and estimate low connectivity due the same explanation of C_G and $\boldsymbol{EF}\left[\frac{1}{d}\right]$

The degree diversity $\frac{1}{K}$ = 0.2760, thus, the low value suggests a few nodes need to be removed to disintegrate the Tijuana MTN.

Finally, the meshedness coefficient M_G = 0.0825, this might be explained because the paths are large, and the cycles are scarce due the structure of Tijuana MTN.

4.3.2 Numerical robustness metrics results

Random node removal: The average of the 30 simulations indicate after removing just 1 node the size of the largest connected component is less than 90% of the original network size, see dashed blue line in Fig. 8, obtaining the critical threshold $f_{90\%}$ which reflect a poor robustness with respect to random failures. The curve color cyan shows the size of the largest connected component is almost around 1100 nodes (25% residual network) after removed less of 40 nodes randomly (0.9% attack), see Fig. 10.

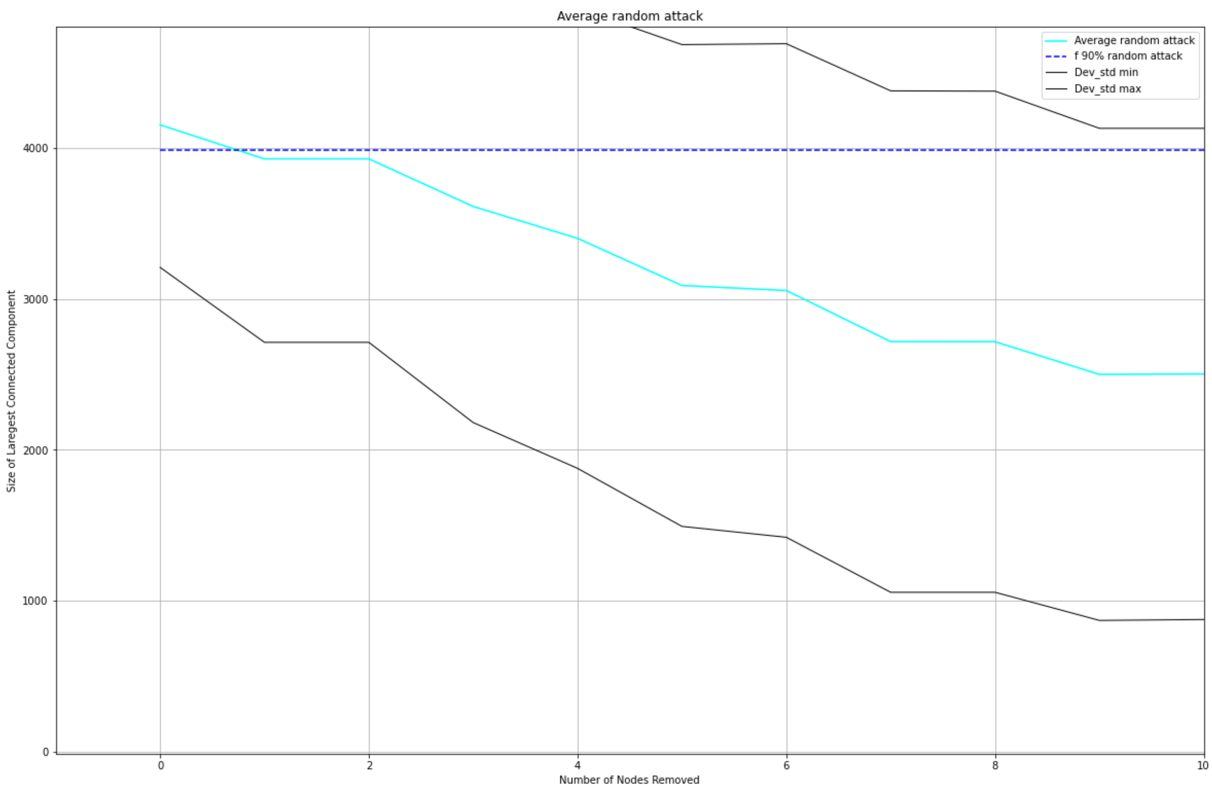

Figure 8. Average random attack curve with visualization of first 10 nodes.

Target node removal: Computing the size of the largest connected component for removed nodes from groups of pareto fronts previously created and described in section 4.2, results in a robustness curve color magenta shows the size of the largest connected component is 961 nodes after pareto front 1 and 2 attack the network with a group of 58 nodes in total (1.3% attack), obtaining a residual network of 21.5%. The node v2122 with high node degree of 8 is considering one of most vulnerable locations by this analysis, see depicted with magenta mark in Fig. 11.

Betweenness node removal: results in a robustness curve color green shows with just the two higher betweenness degree (0.04% attack), obtaining a residual network of around 1100 nodes (25%), see Fig. 9 and Fig 10 for more details. The node v1534 with the higher betweenness degree and node degree as well, is considering one of most vulnerable locations by this analysis, see depicted with green mark in Fig. 11.

All $f_{90\%}$ are remarkably similar for the three strategies of node removal. According with the graph in the Fig. 10, just a few nodes removed from the network generate a disconnected network which support the degree diversity $\frac{\mathbf{1}}{\boldsymbol{K}}$ value.

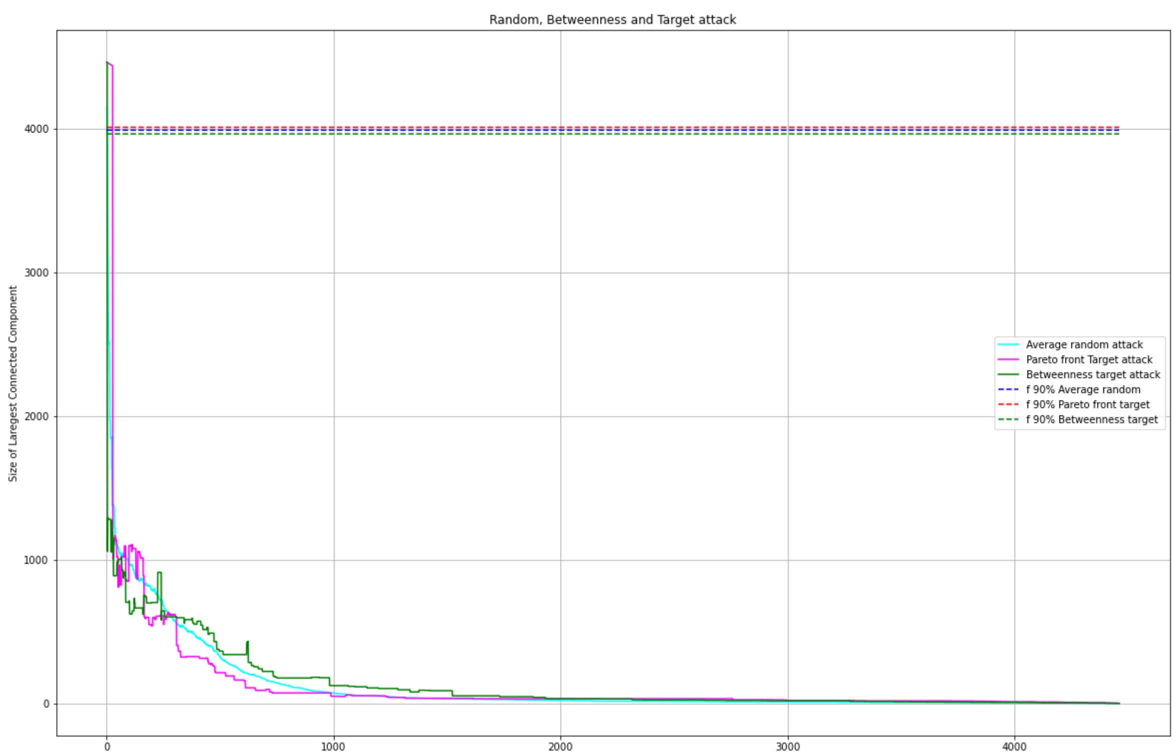

Figure 9. Robustness curve for Tijuana MTN with visualization of all 4461 nodes.

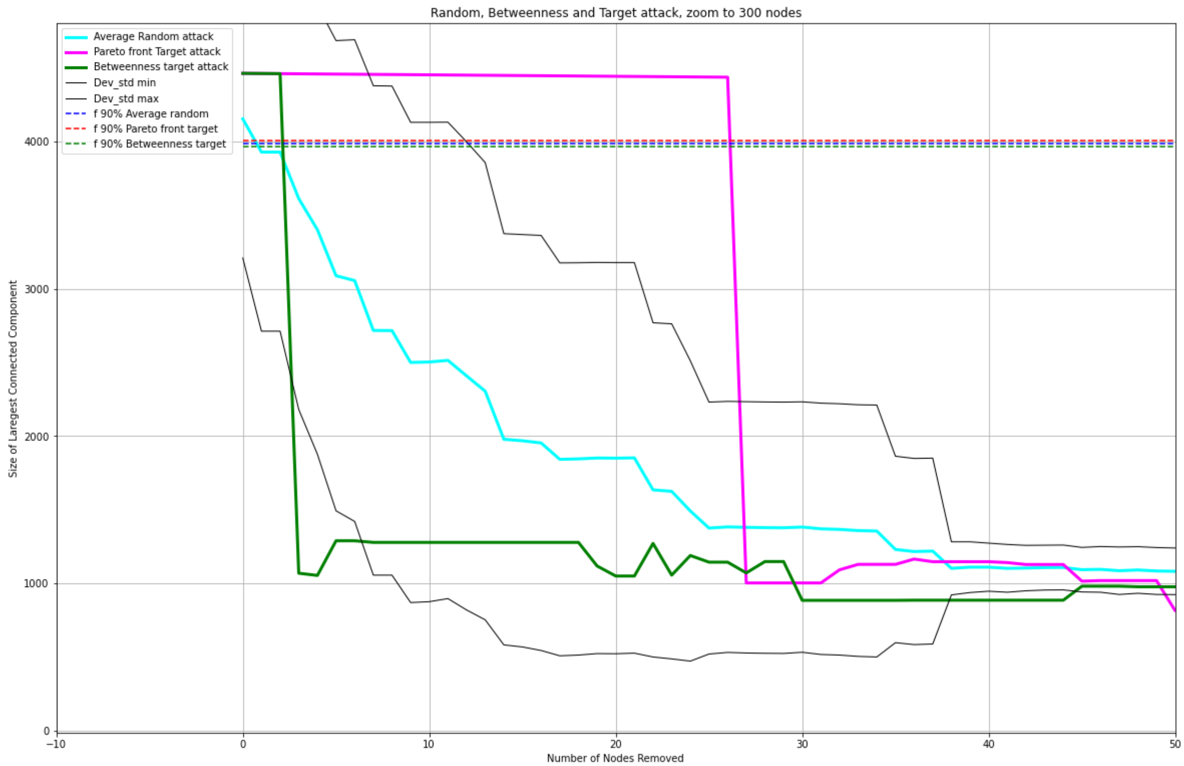

Figure 10. Robustness curve for Tijuana MTN with visualization of first 50 nodes.

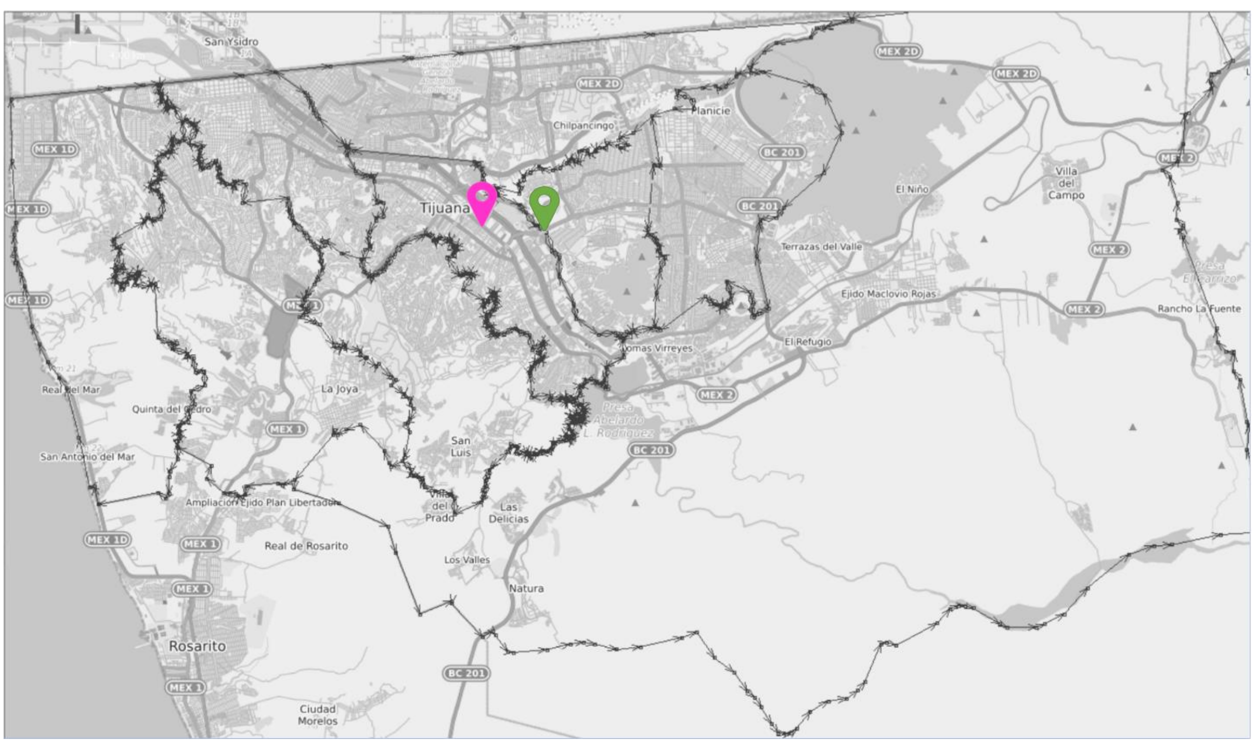

Figure 11. The most vulnerable locations. Pareto front attack: magenta. Betweenness attack: green.

4.4 Discussion

In this work, we analyzed the Tijuana Mass Transit Network (MTN) structural and node importance properties by modeling it as a weighted DAG. In our model, edge weights quantify the great-circle length between adjacent nodes, while nodes route geodetic coordinates. A four-phase data processing pipeline is proposed. Node importance is evaluated by categorizing node metrics into three classes, namely: degree; distance and neighborhood; and rank based. A weighted linear combination of metrics in each class is applied. We apply non-dominated sorting technique to find nodes in each Pareto front.

Findings show that a commuter traveling along the network diameter or radius requires in average 16 and 10 route transfers correspondingly. The network is assortative, approximately 1.2% of its nodes had a high degree of route concentration. According to the proposed node selection criterion, important nodes are concentrated in 4 out of the 9 delegations that the city is composed of. Nodes in the Pareto front 1 can reduce the radius by 45 edges.

The most vulnerable nodes are located both in La Mesa over Federico Benitez Blvd. and Bernardo O'Higgins Blvd., while the other node over Insurgentes avenue and Manuel J. Clouthier Blvd. Visual inspection of these areas show that intersections concentrate many routes, these locations are important due their low robustness under target attacks.

To our knowledge, this is the first study to address the analysis of the MTN network properties. In subsequent studies we transform the network model from a weighed DAG to a multi-graph, since we lose knowledge of independent routes with the current representation.

Chapter 5. Conclusions

The most notable findings are:

- The network is not centralized in a few large centers due to its few hubs, its long segments between intersections and its large diameter that compose it.
- A large number of transfers are required to visit the entire network diametrically and radially.
- All theorical metrics present a low robustness values except the robustness indicator $\boldsymbol{r^T}$ indicating a Tijuana MTN with many alternative paths between any pair of nodes which could be more desirable.
- 80% of the Tijuana MTN can be disconnected with 1.3% of nodes from 2 pareto front attack, very similar case happens with betweenness attack and random failures therefore can be classified as a weak network.

Based on how the network is structured according to the properties presented, a scalable solution would be one in which important nodes are not centered in few locations and the paths between the hubs would decrease. An even less centralized network design is suggested, it would be interesting to analyze the scenario with at least one hub on each delegation. Based on intuition, it is observed that it is not very good that the points are so close to each other. Although by the other previous criteria the network indicates it is not centralized and is vulnerable to attacks from different aspects according with the simulations. Specifically, it would also be interesting to consider in the future robustness studies additional information such as the number of passengers. This would provide us with a data on the number of passengers affected by the failure of a single in a single location for example.

Based on our results, what makes it really important to a node is the number of routes that pass over it, and by concentrating only on 4 of the 9 branches of the city, we can say that they are the regions with the highest concentration of redundant routes. At the moment we do not know if this result is positive or not because we do not have the data to relate it to a demand or congestion issues, perhaps it would be convenient to distribute these nodes but that is a matter of another analysis for future work.

List of references

Alegría, T. y Ordóñez, G. (2005). Legalizando la ciudad. Asentamientos informales y procesos de regulación en Tijuana, COLEF, Tijuana.

Avelar, M. (2014). "Sistema de Transporte de Tijuana: Análisis de su Proceso de Modernización e Implicaciones en la Movilidad Urbana. Un Acercamiento Desde la Sustentabilidad". Tesis de Maestría en Desarrollo Regional. El Colegio de la Frontera Norte, A.C. México. 130.

Ley, J. (2006). Movilidad y transporte público: El arreglo urbano en Tijuana. Revista Universitaria UABC, 4, 8-13.

Millán, L. C. (2018). "¿Transporte público o mi automóvil?: Factores socioculturales determinantes en la elección modal de desplazamiento en Tijuana, B.C.". Tesis de Maestría en Administración Integral del Ambiente. El Colegio de la Frontera Norte, A.C. México. 116 pp.

Newman, M. E. J. (2003). Mixing patterns in networks. Phys. Rev. E, 67, 026126. doi: 10.1103/PhysRevE.67.026126.

Page, L., Brin, S., Motwani, R. & Winograd, T. (1998). The PageRank citation ranking: Bringing order to the Web. Proceedings of the 7th International World Wide Web Conference (p./pp. 161--172), Brisbane, Australia.

Ramos, D. (2014). "Relación entre forma urbana, movilidad a través de transporte motorizado y emisiones de GEI en Tijuana, B.C.". Tesis de Maestría en Administración Integral del Ambiente. El Colegio de la Frontera Norte, A.C. México. 159 pp.

Castro, R. (2012). "Procesos de concesión del transporte público en Tijuana, y efectos en la calidad del servicio.". (Tesis de Maestría). El Colegio de la Frontera Norte, A.C. México.

Barrat, A., Barthelemy, M., Pastor, R. & Vespignani, A. (2004). The architecture of complex weighted networks. Proceedings of the National Academy of Sciences. USA, 101(11):3747–3752.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. (2006). Complex Networks: Structure and Dynamics. Phys. Rep., 424, 175-308.

Rodrigue, J.-P., Comtois, C., & Slack, B. (2013). The geography of transport systems. London: Routledge. 416 pages. ISBN: 978-0-415-82254-1

Cardozo, O., Gómez, E., y Parras, M. (2009). Teoría de Grafos y Sistemas de Información Geográfica aplicados al Transporte Público de Pasajeros en Resistencia

(Argentina). Revista Transporte y Territorio Nº 1, Universidad de Buenos Aires, 2009. ISSN: 1852-7175. pp. 89-111.

Sasidharan, M., Usman, K., Ngezahayo. E & Burrow, M. P. N. (2019). Evidence on impact evaluation of transport networks using network theory. K4D Helpdesk Report. Brighton, UK: Institute of Development Studies.

Barthelemy M. (2011). “Spatial Networks”, Physics Reports, 499:1.

Cardillo, A., Scellato, S., Latora, V., & Porta, S. (2006). Structural properties of planar graphs of urban street patterns. *Physical Review E*, *73*(6), [066107]. https://doi.org/10.1103/PhysRevE.73.066107

Newman, M. E. J. (2010). Networks: An Introduction. New York: Oxford University Press.

Chun, Y, & Kim, H. (2018). Networks Defined. *The Geographic Information Science & Technology Body of Knowledge* (4th Quarter 2018 Edition), John P. Wilson (Ed.). DOI: 10.22224/gistbok/2018.4.3

Chopde N. R. & Nichat M. K. (2013). Landmark based shortest path detection by using A* and haversine formula. International Journal of Innovative Research in Computer and Communication Engineering, 1(2):298–302.

Foster, J. G., Foster, D. V., Grassberger, P., & Paczuski, M. (2010). Edge direction and the structure of networks. Proceedings of the National Academy of Sciences of the United States of America, 107(24), 10815–10820. https://doi.org/10.1073/pnas.0912671107.

Instituto Metropolitano de Planeación IMPLAN (2017). Plan municipal de desarrollo 2017-2019. Tijuana, B.C.: México. p.129.

Kleinberg, J. M. (1999). Authoritative Sources in a Hyperlinked Environment. J. ACM, 46, 604--632. doi: 10.1145/324133.324140.

Kleinberg, J. M. (1999). Hubs, authorities, and communities. ACM Comput. Surv., 31, 5.

Langville, A. N. & Meyer, C. D. (2005). A Survey of Eigenvector Methods for Web Information Retrieval. SIAM Rev., 47, 135--161. doi: 10.1137/s0036144503424786.

Coello, C., Veldhuizen, D., & Lamont, G. (2006). Evolutionary Algorithms for Solving Multi-Objective Problems (Genetic and Evolutionary Computation). Springer-Verlag New York, Inc. Secaucus, NJ, USA.

Pumar, L., Barbastefano, R., and Carvalho, D. (2018). Bus trans-port network analysis in Rio de Janeiro based on topological modelsusing social networks. CoRR, abs/1808.05692.

Scheurer J., Curtis C., Porta S. (2007). Spatial Network Analysis of Public Transport Systems: Developing a Strategic Planning Tool to Assess the Congruence of Movement and Urban Structure in Australian Cities. ResearGate.

Shanmukhappa, T., Ho, I. W.-H. & Tse, C. K. (2018). Spatial analysis of bus transport networks using network theory.PhysicaA: Statistical Mechanics and its Applications, 502:295–314.

Wang, Xiangrong, Koç, Yakup, Derrible, Sybil, Ahmad, Sk Nasir, Pino, Willem J.A. and Kooij, Robert E., (2017), Multi-criteria robustness analysis of metro networks, Physica A: Statistical Mechanics and its Applications, 474, issue C, p. 19-31, https://EconPapers.repec.org/RePEc:eee:phsmap:v:474:y:2017:i:c:p:19-31.

Elisa Frutos Bernal & Angel Martín del Rey, 2019. "Study of the Structural and Robustness Characteristics of Madrid Metro Network," Sustainability, MDPI, Open Access Journal, vol. 11(12), pages 1-24, June. X. Barabási, A.-L., Pósfai, M. (2016). Network science. Cambridge: Cambridge University Press. ISBN: 9781107076266 1107076269

Barabási, A.-L., Pósfai, M. (2016). Network science. Cambridge: Cambridge University Press. ISBN: 9781107076266 1107076269