# Spatial network analysis of Tijuana's metropolitan transport network

Valencia-García, B.[1][0000−0003−2088−6390] and Hirales-Carbajal, A.[1][0000−0002−5067−7250]

CETYS Universidad, Calz Cetys 813, Lago Sur, Tijuana, B.C. México
belgica.valencia@cetys.edu.mx, adan.hirales@cetys.mx

**Abstract.** 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 properties have not been analyzed. In order to understand its structural properties and identify the locations of important nodes, we model and evaluate the transport network as weighted directed graph and analyzed its overall structural properties. We locate important nodes by analyzing each node in function of three criterion's: degree, distance and rank based attributes. We use non-dominated sorting to find nodes in the Pareto 1 front that maximize all criterion's. Among 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. Showing that the network needs restructuring.

**Keywords:** Tijuana transport network · structural properties · spatial analysis.

## 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 [10]. Immigration phenomenon have accelerated the city growth, with more than 50% of its settlements of irregular origin, mostly in periphery [1, 10]. 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 [2], quality of service [4, 11], network supply [8], relationship between urban form and the environment [15], and public policy and governance of the transit network [4, 11]. In this study, our focus is on analyzing the network structural and node properties.

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. Route data was provided by the offices of the Municipal Public Transport Directorate.

## 2 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 $G = (\mathcal{V}, \mathcal{E})$, with $\mathcal{V}$ the set of nodes and the $\mathcal{E}$ the set of edges. The number of elements in $\mathcal{V}$ and $\mathcal{E}$ are denoted by $V$ and $E$. Node $n_i \in \mathcal{V}$ is defined by the tuple $(\varphi_i, \lambda_i)$, with $\varphi_i$ and $\lambda_i$ the geodetic coordinates (latitude and longitude). A matricial representation of node adjacency is given by $A$ a $NxN$ square matrix with entries $a_{ij}$ equaling 1 if the edge $e_{ij} \in \mathcal{E}$, 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 [5] 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 denoted as $k_i^{in} = \sum_{j \in \mathcal{V}} a_{j,i}$ and the out-degree as $k_i^{out} = \sum_{j \in \mathcal{V}} a_{i,j}$. 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 $i \neq j$. 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. Let $n_{jk}$ be the number of all shortest paths between 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 $\tau(G)$. See [3] for an exhaustive summary of network metrics.

## 3 The data processing pipeline

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.

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 DAG i.e. node clustering degree, betweenness, among others. In this work, we focus on characterizing the topology at global and node level. We evaluate the network quality by applying metrics summarized in Table 1. Due to space constrains we refer the reader to [3,7,12,13,16] for their description and interpretation.

**Table 1.** Network and node level metrics. Source: Own elaboration.

| | **Metric** | **Expression** |
|---|---|---|
| **Centrality** | Degree | $k_i = \sum_{j \in \mathcal{V}} a_{i,j}$ |
| | Knn | $k_{nn,i} = 1/k_i \sum_{j \in \mathcal{N}_i} k_j$ |
| | Pagerank | $pr_i$, see [9,13] |
| | Betweenness | $b_i = \sum_{j,k \in \mathcal{V}, j \neq k} \frac{n_{jk}(i)}{n_{jk}}$ |
| | Closeness | $c_i = \frac{1}{\sum_{j \in \mathcal{V}; i \neq j} d_{ij}}$ |
| **Network** | Assort. Coeff. | $r$, see [12] |
| | Diameter | $d = \max_{j \in \mathcal{V}} d_{ij}$ |
| | Transitivity | $t = \frac{3\delta(G)}{\tau(G)}$ |

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 [1].

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. Thus, node importance may decrease throughout nodes these routes traverse in our results. The size of $\mathcal{V}$ 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 $\mathcal{V}$ reduced to 4461 non-redundant nodes. The transport network was digitally edited with JOSM Ver. 15238, an extensible editor for Open Street Map. The resulting set of routes was stored in geojson format and is available for download in the project site.

[1] See *https://github.com/ahiralesc/tijuana_vehicle_network_analysis.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.

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 $\bar{k}_i, k_i^{in}, k_i^{out}, k_{nn}$, 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 front [6].

# 4 Results

## 4.1 Network properties

Network level properties are summarize in Table 2. A diameter, 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 [14]. 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 [8]. 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 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 [18]. These results confirm that the diameter is large.

**Table 2.** Global network properties. Source: Own elaboration.

| $r$ | $d$ | $t$ |
|---|---|---|
| 0.37 | 190 | 0.01 |

### 4.2 Node importance

Fig 1 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).

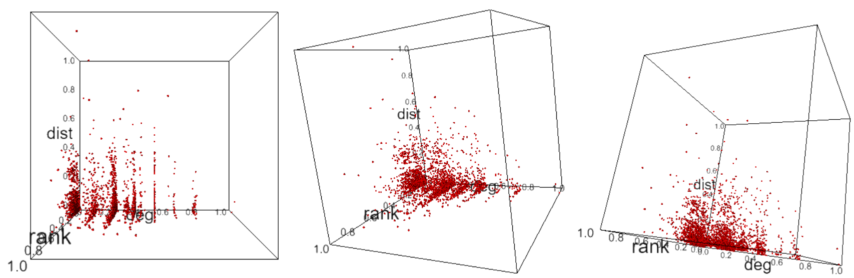


**Figure 1.** Linearization of metrics into 3D space. Source: Own elaboration based on [6].

The 10 most important nodes are depicted with blue markers. See Fig. 2. 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. 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.

Table 3 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 [17]. 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 cen-

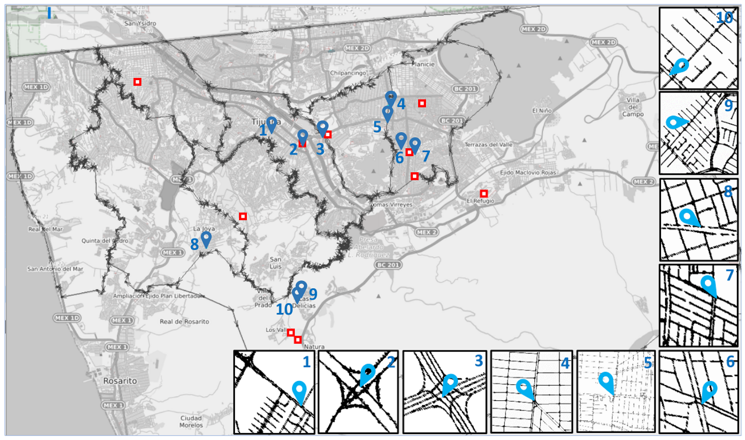

**Figure 2.** First Pareto front nodes. Source: Own elaboration based on [6].

trality 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 3.** Mean metrics values. Source: Own elaboration.

| **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 closeness 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 follow circuits in suburbs.

## 5 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.

To our knowledge, this is the first study to address the analysis of the MTN 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. We also plan to evaluate the robustness with the aim of identifying the locations and events that may render the network unable to operate.

## References


1. Tito Alegria and Gerardo Ordóñez Barba. *Legalizando la ciudad: asentamientos informales y procesos de regularización en Tijuana.* El Colegio de la Frontera Norte, March 2005.
2. Miguel Angel Avelar-Dueñas. Sistema de transporte de tijuana: Análisis de su proceso de modernización e implicaciones en la movilidad urbana. Un acercamiento desde la sustentabilidad. mathesis, El Colegio de la Frontera Norte, Tijuana, B.C., México, 2014.
3. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, February 2006.
4. Perla Rocío Castro-Cebreros. Procesos de concesión del transporte público en Tijuana, y efectos en la calidad del servicio. mathesis, El Colegio de la Frontera Norte, Tijuana, B.C., Mexico, 2012.
5. Nitin R. Chopde and Mangesh K. Nichat. 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, April 2013.
6. Carlos A. Coello Coello, Gary B. Lamont, and David A. Van Veldhuizen. *Evolutionary algorithms for solving multi-objective problems (Genetic and Evolutionary Computation).* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. Evolutionary.
7. J. G. Foster, D. V. Foster, P. Grassberger, and M. Paczuski. Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences*, 107(24):10815–10820, May 2010.
8. Instituto Metropolitano de Planeación IMPLAN. Plan municipal de desarrollo 2017-2019. Tijuana, B.C.: México. p.129, 2017.
9. Amy N. Langville and Carl D. Meyer. A survey of eigenvector methods for web information retrieval. *SIAM Review*, 47(1):135–161, January 2005.
10. Judith Ley García. Movilidad y transporte público: El arreglo urbano en Tijuana. *Revista Universitaria UABC*, 4:8–13, January 2006.

11. Lilia Cristina Millán-Velázquez. ¿Transporte público o mi automóvil? Factores socioculturales determinantes en la elección modal de desplazamiento en Tijuana, B.C. mathesis, El Colegio de la Frontera Norte, Tijuana, B.C., México, 2018.
12. M. E. J. Newman. Mixing patterns in networks. *Physical Review E*, 67(2), February 2003.
13. Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999.
14. Louise Pumar, Rafael Barbastefano, and Diego Carvalho. Bus transport network analysis in Rio de Janeiro based on topological models using social networks. *CoRR*, abs/1808.05692, 2018.
15. Dalila Ramos-Ortiz. Relación entre forma urbana, movilidad a través de transporte motorizado y emisiones de Gei en Tijuana, B. C. mathesis, El Colegio de la Frontera Norte, Tijuana, B.C., México, 2014.
16. Thomas Schank and Dorothea Wagner. Approximating clustering coefficient and transitivity. *J. Graph Algorithms Appl.*, 9(2):265–275, 2005.
17. J. Scheurer, Carey Curtis, and Sergio Porta. Spatial network analysis of public transport systems : developing a strategic planning tool to assess the congruence of movement and urban structure in Australian cities. *Australasian Transport Research Forum*, 2007.
18. Tanuja Shanmukhappa, Ivan Wang-Hei Ho, and Chi Kong Tse. Spatial analysis of bus transport networks using network theory. *Physica A: Statistical Mechanics and its Applications*, 502:295–314, jul 2018.