Internship Report
Submitted for the degree of
Data Engineer

Universidad Politécnica de Yucatán

Analysis of the road network of the city of Merida, Yucatan, Mexico.

Alejandro De Jesus Puerto Castro

1. *External Supervisor* Gonzalo G. Peraza Mues
Data Engineering
Universidad Politécnica de Yucatán

2. *Internal Supervisor* Didier O. Gamboa Angulo
Data Engineering
Universidad Politécnica de Yucatán

May 2, 2021

Contents

List of Figures

List of Tables

Acknowledgments

Abstract

1 Introduction

1.1 Background

1.2 Problem Statement

1.3 Justification

1.4 Scope and limitations

1.5 Objectives

1.5.1 General

1.5.2 Specific

1. a)
   b)
   c)

2 Theoretical framework

On this section we present concepts to support and further explain the technologies and process found on the Methodology and Development section. We present concepts regarding spatial networks and its analysis measures based on network theory; spatial data; as well as some tools and services that we use as a base for this work.

2.1 Spatial Networks

A network (or *graph* in mathematics) is composed of nodes N connected by links, or edges, E. Nodes represent entities in a network, such as cities, people, airports, and street intersections. Edges represent relationship between nodes, such as friendships among people, flights between airports, street roads, and so on.

Graphs can be arranged nonspatially or spatially. Spatial graphs have nodes that are georeferenced, i.e. they are defined by their location in geographic space using a pair of coordinates; usually embedded in a two- or three-dimensional space [4]. Both nonspatial and spatial graphs can contain undirected, directed, unweighted, or weighted components [5].

An undirected graph has edges that can be used to represent flows of traffic on two-way streets, while edges in a directed graph represent one-way streets. A self-loop in a graph is an edge that connects a single node to itself. Two nodes can also be connected by parallel edges, in such case, graphs are called multigraphs, or multidigraphs if they are directed. The weight attribute in a weighted graph is used to quantify some value between connected nodes [6].

Network science was founded on the based on the findings by Euler [7], and gave way to the notion that graphs have different structural properties that can be discovered and cataloged using graph theory [8]. With this, early spatial network analysis focused on using graph theory and its measures to describe and catalog the properties of real systems represented as spatial networks [5].

Real spatial networks are complicated physical entities, with numerous, often complex, elements with a nontrivial configuration and structure that are neither purely fully regular nor fully random. A street network is an example of a complex spatial network with both nodes and edges embedded in space.

A spatial network is *planar* if it can be represented in two dimensions with its edges intersecting only nodes. A street network may be planar at certain small scales, but most street networks are non-planar due to overpasses, bridges, tunnels, etc. For tractability, these networks are studied as approximately planar. However, it can cause analytical problems due to the over-simplification.

Street networks can be based on two graph representations: primal graph or dual graph. In a primal graph representation, intersections are turned into nodes and street segments into edges. On the other hand, dual graphs invert the representation: streets as nodes and intersections as edges [9]. Primal graphs retain all the geographic, spatial information that are lost in a dual graph. For that reason, primal representation is the better approach for analyzing spatial networks as it faithfully represents all the spatial characteristics of a street [10].

2.2 Spatial Networks Analysis Measures

The structure and behavior of networks can be described using a variety of graph theory measures. These measures can be found in different level of detail in [4, 8, 11–14].

Each network is characterized by the **total number of nodes** N and the **total number of edges** E. We call N the **size of the network**.

The ***degree*** k of a node is its number of edges, or neighbors, and it is a local measure. We use k_i to denote the degree of node i. A node with no neighbors has degree zero ($k = 0$) and is called a ***singleton***.

The **average degree** $< k >$ is a global measure for the average degree k across all nodes N in a graph. This measure is simplified by dividing twice the number of edges E by the number of nodes N, as follows:

$$< k >= \frac{2E}{N} \tag{2.1}$$

The average streets per node measures the mean number of streets (edges in an undirected graph) that come out from each intersection or dead-end.

The **degree distribution** $P(k)$ represents the fraction of nodes in a graph with degree k, calculated by dividing the number of nodes with degree k by the total number of nodes N in the graph G. The degree distribution $P(k)$ is often plotted on a histogram and is useful for providing an overall snapshot of graph G.

The **clustering coefficient** C measures the ability of an individual node i to associate with other nodes (cliquishness). It is commonly described as the probability that ”friends” of i (i.e., nodes connected to node i) are also friends of each other: the chance that a friend of my friend is also my friend [15]. For node i of degree k_i, the clustering coefficient $C(i)$ is defined as:

$$C(i) = \frac{E_i}{k_i(k_i - 1)/2} \tag{2.2}$$

where E_i is the number of edges existing between the neighbors of i. When the local measure $C = 1$, the node v_i and its neighboring nodes are all perfectly connected. In contrast, when $C = 0$, neighbors of node i are not connected at all.

The **average clustering coefficient** $< C >$ is a global measure that determines the cliquishness of all nodes in a graph and is calculated as the average C over all individual nodes. When $< C >= 1$, the graph is perfectly connected. In contrast, when $< C >= 0$, the graph is not connected at all.

Path P is and ordered sequence, or collection, of edges that connects some ordered sequence of nodes. The collection of nodes N and edges E in a path can be defined as:

$$N_p = \{0, 1, 2, ...n\} \tag{2.3}$$

$$E_p = \{0, 1, 2, ...m\} \tag{2.4}$$

There may be many paths of varying lengths l between two nodes i and j. The **shortest path length** l_s is calculated by counting the total number of intermediate nodes or edges along the shortest path between two nodes i and j and is defined as:

$$l_s(i,j) = \min_{\text{paths}}(i \to j) \tag{2.5}$$

The **average shortest path length** $< l >$ is defined as the average shortest path length between all possible pairs of nodes in the network. The **diameter** d_G of a graph G is defined as the maximum shortest path length l_s found in the graph.

Average street length is the mean edge length measured in meters, an example of spatial units, and indicates how fine-grained (small block size) or coarse-grained (large block size) the networks is.

Density measures provided how fine-grained the network is. **Node density** is the number of nodes divided by the area covered by the network. **Intersection density** is the node density of the set of nodes with more than one street emanating from them, excluding dead-ends. The **edge density** is the sum of all edge lengths divided by the area. The physical **street density** is the sum of all edges (in the undirected graph) divided by the area.

The **average circuity** is the circle distances between the nodes of each edge, and it is defined by the sum of all edge lengths divided by the sum of the great-circle distances between the nodes incident to each edge [16].

Eccentricity is the largest distance (the maximum of the shortest-path weighted distances) between a node and other nodes i.e., how far the node is from the node that is furthest from it [17]. The **diameter** of a network is the maximum eccentricity of any node in the network and the **radius** is the minimum eccentricity [18]. The **center** if a network is the node or set of nodes with an eccentricity equals the radius, and the **periphery** of a network is the node or set of nodes with eccentricity equals the diameter. These distances serve as indicators for network size and shape if we use length as weight.

Connectivity measures the minimum number of nodes or edges that must be removed from a connected graph to disconnect the network [17]. In the case of street networks, we use **average node connectivity** as a resilience indicator, which is the mean number of internally node-disjoint paths between each pair of nodes. This measure is more useful to represent the expected number of nodes that must removed to disconnect a randomly selected pair of non-adjacent nodes [19,20] Networks with low connectivity may have multiple points of failure, this yield to a vulnerable system.

Centrality measures indicate the most important nodes in a network [21,22]. **Betweenness centrality** g_i measures the total number of shortest paths between any two nodes in the graph that pass through node i [23,24] and is defined as:

$$g_i = \sum_{u \neq v} \frac{\sigma_{uv}(i)}{\sigma_{uv}} \tag{2.6}$$

where σ_{uv} is the number of shortest paths going from node u to note v and $\sigma_{uv}(i)$ is the number of shortest paths going from node u to node v through node i. The importance of an edge j is also measured by betweenness centrality g_j that instead calculates the total number of shortest paths between any two nodes in a graph that include edge j [25] and is defined as:

$$g_j = \sum_{u \neq v} \frac{\sigma_{uv}(j)}{\sigma_{uv}} \tag{2.7}$$

where σ_{uv} is the number of shortest paths going from node u to node v and $\sigma_{uv}(j)$ is the number of shortest path going from node u to node v through edge j. In many graphs, betweenness centrality g_i and node degree k_i correlate, where the most central node can also have the most connections. The **average betweenness centrality** is the mean of betweenness centralities of all the nodes in the network [4]. The maximum betweenness centrality in a network specifies the proportion of shortest paths that pass through the most important node. If the maximum betweenness centrality is high, the network is more susceptible to failure or inefficiency.

Closeness centraltity is another way to measure the centrality of a node by determining how close a node is to the other nodes. This can be done by averaging the sum of the distances from the node to all others. This measure gives low values for more central nodes and high values for less central ones [13]. It is defined as the inverse of the sum of distances of a node from all others:

$$g_i = \frac{1}{\sum_{j \neq i} l_{ij}} \tag{2.8}$$

where l_{ij} is the distance from i to j and the sum runs over all the nodes of the network, except i itself. An alternative formulation to discount the graph size and make the measure comparable across different networks is obtained by multiplying equation 2.8 by the constant $N-1$, which is just the number of terms in the sum at the denominator:

$$\tilde{g}_i = (N-1)g_i \tag{2.9}$$

Finally, **PageRank** is an algorithm to compute a centrality measure that aims to capture the prestige or importance of each node and it is typically used in directed networks. It ranks nodes based on the structure of incoming links and the rank of the source node. This measure can also be applied to street networks [26–29]. It is worth to mention that multiple studies use centrality measures in combination to assess street networks (e.g., [30–35]).

A graph community is defined as a set of nodes that have more connections among themselves than other nodes in the graph [36]. This feature is important in spatial networks since dense connections tend to take place between nodes that are closer in proximity. Moreover, this implies that the majority of flows between nodes occur as a function of nodes belonging to the same geographical region [37].

A community is typically identified by calculating **modularity** Q [38] and is defined as:

$$Q = \sum_{s=1}^{n_M} \frac{l_s}{E} - (\frac{d_s}{2E})^2 \tag{2.10}$$

where n_M is the number of modules of the partition, l_s is the number of edges inside module s, E is the total number of edges in the network, and d_s is the total degree of the nodes in the module s.

The above measures do not account for the distance between linked node pairs, an important measure that can be used to quantify real spatial networks embedded in geographic space. Distance can be measured in a variety of ways, the most common being Euclidean distance $d_E(i,j)$ or as the direct distance between two points. In contrast, the route distance $d_R(i,j)$ is computed by summing the geographical length of edges, which make up the shortest path between node v_i and v_j [5].

2.3 Spatial Data

As our data is embedded in space, we need to understand its properties:

A datum is a model of the Earth's shape. Sometimes the Earth is assumed to have an spherical shape who is described by two coordinates, latitude (north) and longitude (east). However the Earth is not a sphere; its shape is more like an ellipsoid. There are many possible approximations to this shape, which define their own latitude-longitude coordinate system. A coordinate system (CS) is a sequence of coordinate axes with specified units of measure, and its types are: ellipsoidal, Cartesian, affine, gravity-related, linear, spherical, polar, and cylindrical. A coordinate reference system (CRS) associates a CS with an object by mean of a datum (see Figure 2.1) [1]. Some are more accurate than others for particular regions of the Earth's surface. If our data is notated in different datums then we will need to convert them into one standard format. The most common global datum is called WGS84 (World Geodetic System, 1984) [39].

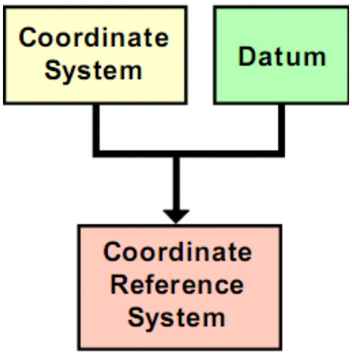


Figure 2.1: A coordinate reference system combines a coordinate system with a datum, which gives the relationship of the coordinate system to the surface and shape of the Earth. Retrieved from [1].

A projection is the change of the representation of locations from one coordinate system to another. Sometimes it is more convenient to work with a flattened 2D projection of a datum rather than its spherical coordinates. With this, we project the coordinates into Cartesian x and y meters. We take x = Easting and y = Northing, in the order (x, y), in meters from some origin. When we do a projection, we must make some compromise because it is not possible to make a perfect flat version of an ellipsoid surface. All projections of locations on the Earth into a two-dimensional plane are distortions as something always will be distorted [40]. A good projection to use is one of the Universal Transverse Mercator (UTM) family. UTM splits the Earth's surface into state-sized regions, and defines separate projection for each one, to minimize the distortion there (see Figure 2.2).

With geographic data it is common to work in only two of the three dimensions. Two dimensional space support three basic types of spatial entity [41]:

- points - having a location
- lines - comprising two or more locations in an ordered sequence
- polygons - areas defined by three or more vertex locations in an ordered sequence

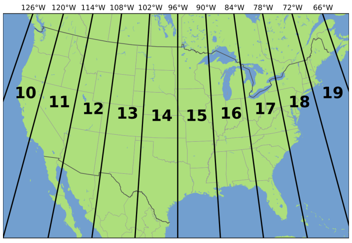


Figure 2.2: UTM zones across the continental United States. Source: [2].

2.4 Geographic Information Systems (GIS)

A Geographic Information System (GIS) is any system that is specifically designed to work with spatial data. GIS systems are implementations of some standard tasks, which may be present in as programming language libraries or functions, and/or as graphical user interfaces [39]. Standard tasks include:

- converting datums and projections
- searching quickly for entities in particular regions
- searching quickly for entities with spatial relationships to other entities
- handling lots of data at large and small scales
- geographical data visualizations
- converting between standard spatial data file formats

2.5 Spatial files

Shapefiles are a storage format for the Open Geo-spatial Consortium (OGC) definition: points, lines, and polygons. They are small collections of related files, usually stored together in a directory. The main file has a *.shp* extension and stores the actual feature geometries. Other files that may appear along with it include *.dbf* (associated non-spatial properties data), *.shx* (indexing structure), and *.prj* (datum/projection information) [39].

GeoJSON and Well-Known Text (WKT) are alternatives for storing such data. Most GIS systems can convert between those formats.

2.6 OpenStreetMap

OpenStreetMap (OSM) is a collaborative worldwide mapping project that provides a free and publicly editable map of the world. In February 2021, there were over 7M registered contributors, as outlined on the OSM wiki [42].

In Mexico, OSM imported the 2015 INEGI's Marco Geoestadístico Nacional (MGN) and the Red Nacional de Caminos (RNC) road data during the years 2015 and 2016 as part of the two OSM projects: Mexico Main Road Network Import Project [43], and Mexico's Administrative Divisions Import Project [44]. Additionally, individual contributions have been made from OSM members to keep updated and accurate the data. Acquiring the spatial data from OSM is made via an API, called Overpass, to retrieve any data in the database. However, its usage and syntax are somewhat difficult and there are other services available that can simplify the process.

2.7 Network Analysis Tools

All the tools here presented are Python language packages or libraries. Python was chosen because it is a popular language, free, open-source, easy for beginners, powerful, and it gives us the ability to work interactively and easy integration with other Python libraries.

2.7.1 NetworkX

NetworkX is a free, open-source Python language package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks [45,46]. It provides data structures for graphs, digraphs, and multigraph, and implementations of many of the algorithms used in network science. These algorithms are implemented for structure and analysis measures, such as shortest paths, betweenness centrality, clustering, degree distribution and many more.

In addition NetworkX can read and write various graphs formats (e.g., adjacency list, edge list, GEXF, GML, pickle, GraphML, JSON, LEDA, YAML, SparseGraph6, Pajek, and GIS Shapefile), and provides generators for classic graphs, random graphs, and synthetic networks.

2.7.2 GeoPandas

Python's spatial data frame library is called GeoPandas [47]. GeoPandas is an open source project that allows easier manipulation of geospatial data. It extends the datatypes used by pandas (data series and data frames) [48,49] to allow spatial operations on geometric types (points, lines, polygons). It uses shapely for geometric operations [50], fiona for spatial data file access [51], as well as descartes [52] and matplotlib [53,54] for plotting.

2.7.3 OSMnx

OSMnx [6] is a free, open-source Python package to download spatial data from OpenStreetMap and model, project, visualize, and analyze real-world street networks (e.g., walking, driving, or biking network) including node elevations and street grades. It also allow us to save the street network as shapefiles, GeoPackages, and GraphML files for later use.

OSMnx is built on top of GeoPandas, NetworkX, and matplotlib and interacts with OpenStreetMap's APIs. Thanks to that we can conduct topological and spatial analyses that OSMnx automatizes for calculating dozens of indicators, as well as calculating and visualizing the street network, street bearings and orientations, shortest-path routes.

2.7.4 PySAL

Python Spatial Analysis Library (PySAL) is an open-source, cross-platform library for geospatial data science and spatial analysis [55].

PySAL is a family of packages and is divided into four components:

- Explore – It includes modules to conduct exploratory analysis of spatial and spatio-temporal data, focused on enabling better understanding of patterns in the data.
- Model – It focuses on confirmatory analysis to model spatial relationships in data with a variety of models.
- Viz – It supports the creation of geovisualizations and visual representations of outputs from a variety of spatial analyses.
- Lib – It help us to solve a wide variety of computational geometry problems including graph construction from polygonal lattices, lines, and points, construction and interactive editing of spatial weights matrices and graphs, computation of spatial relationships, and reading and writing of spatial vector data.

2.8 Related work

Spatial networks have been subject of study in many forms through the years. These forms, such as locations, flows of people and goods, activities, etc. are commonly studied involving time and space to make and answer questions in the complex system field to discuss the importance and evolution of networks.

Barthélemy's work [4,37] provides an important, comprehensive review of spatial networks properties, models and measures for their analysis. The information presented in his work explains in very detail the constraints and effects of spatial networks in complex systems and its processes.

Other authors have tried to contribute by complementing the work made by Barthélemy. O'Sullivan [56] describes some important concepts and definitions placing particular emphasis on high-level structure in networks. On the other hand, Anderson [5] evaluates the integration of geographic information systems (GIS) and complex spatial networks to explore the development and applications of geographic automata systems (GAS), which are network-based automata models.

In recent years, Geoff Boeing has made relevant contributions to the field of spatial network analysis, specifically for street networks. He has developed a tool previously mentioned in this work: OSMnx [6]. With such tool, he has done multiple studies regarding the analysis of urban street networks [57]. From developing two new indicators (spatial planarity ratio, and the edge length ratio) for measuring planarity and describing infrastructure and urbanization [58] to analyzing and comparing thousands of urban street networks [59] and exploring patterns and configuration through visualization methods [60].

3 Methodology and development

3.1 The Repository

This work was done in Jupyter notebook format in a GitHub repository (`https://github.com/gperaza/road-network`). This repository's root contains an environment definition file and a notebooks folder. Within that folder, the repository contains one folder for the work done in Mérida, Yucatán, and other folder for the different cities in México that are beyond the scope of this report. The following notebooks are included in the Mérida folder:

1. Data preparation, downloading/modeling and calculating network stats of Mérida's road network and its urban AGEBs.
2. Analysis of the road network of Mérida and its urban AGEBs.
3. Cluster analysis of the urban AGEBs.

To run the code examples in this resource repository, we simply run everything in a pre-built Anaconda environment. This process is detailed in the following section.

3.2 The Environment

This project's repository contains an Anaconda environment file (i.e., .yml) for running the Jupyter notebooks on any computer. Anaconda [61] is a data science platform that facilitates package management and deployment. It is available for Windows, Linux and macOS. We use the Individual Edition, which is the open-source distribution of Anaconda.

First, download and install Anaconda Individual Edition. Once it is installed and running on your computer, run the following code in the terminal window:

```
$ conda config --add channels conda-forge
$ conda env create --file road-network-project.yml
$ conda activate road-network-project
$ jupyter lab
```

Once you are in the active environment, open your computer's web browser and visit `http://localhost:8888` to access Jupyter Lab and open this work's notebook files.

3.3 Data Collection

This work uses OSMnx to download the street network of Mérida and its AGEBs, construct a graph model of it (using NetworkX), correct, analyze, and visualize at municipal, and neighborhood (AGEB) scales.

In order to get the street network of Mérida, we define a function to download and save the graph of the municipality, or, if already present, load it as an NetworkX graph object:

```
def get_roads_osmnx(places, update=False, proj=False, crs=None):

        dirpath = pathlib.Path(’./data/networks/’)
        filepath = dirpath/’merida-road.graphml’
        logpath = dirpath/’log’

        if filepath.exists() and not update:
                G = ox.load_graphml(filepath)
        else:
                # get drivable public streets network, aka road network,
                    ↪ without service roads,
                # e.g. private, parking lots, etc.
                # use retain_all if you want to keep all disconnected
                    ↪ subgraphs (e.g. when your places aren’t adjacent)
                G = ox.graph_from_place(places, network_type=’drive’)
                ox.save_graphml(G, filepath=filepath, gephi=False)

        if proj:
                G = ox.project_graph(G, to_crs=crs)

        print(f"Graph␣created␣at:␣{G.graph[’created_date’]}")
        return G, *ox.graph_to_gdfs(G)

places = [{’county’ : ’Merida’, ’state’ : ’Yucatan’, ’country’ : ’Mexico’}]
G_proj, nodes_proj, edges_proj = get_roads_osmnx(places, update=False, proj
    ↪ =True, crs=3857)
```

OSMnx geocodes the query ”Merida, Yucatan, Mexico” to retrieve the place boundaries of that city from the Nominatim API, retrieves the drivable street network data within those boundaries from the Overpass API, constructs a graph model (via NetworkX), then simplifies/corrects its topology such that nodes represent intersections and dead-ends, and edges represent the street segments linking them. It also saves the constructed graph as a GraphML file to not re-download the same data again.

OSMnx models all networks as NetworkX MultiDiGraph objects. You can convert to:

- Undirected MultiGraphs.
- DiGraphs without (possible) parallel edges.
- GeoPandas node/edge GeoDataFrames.

In the function, we also convert the graph to node and edge GeoDataFrames. Additionally, we project the graph to the WGS84 Pseudo-Mercator CRS.

Remember that one of the most commonly used CRS is the WGS84 latitude-longitude projection. This can be referred to using the authority code ”EPSG:4326”. However, such EPSG is in degree units. For that reason, we will use an alternative option which is the ”EPSG:3857” that is measured in meters. This projected coordinate system is the one that Google, OpenStreetMap, Bing, ArcGIS, ESRI, etc. use for rendering their maps [62].

In the case of the Mérida's urban AGEBs, we collect geographic data from the Institute of Statistics and Geography's (INEGI) National Geoestatistical Framework (MG) and socio-demographic data from INEGI's 2020 Population and Housing Census (2020 Census) conducted from March 2 to March 27, 2020 [63].

The MG is a mexican unique national system designed by INEGI to correctly reference statistical information from censuses and surveys with the corresponding geographic locations [3]. It is conformed by geostatistical areas divided into three dissaggregation areas (see Figure 3.1):

- State geoestatistical areas (AGEE).
- Municipal geoestatistical areas (AGEM).
- Basic geoestatistical areas (AGEB).
  - Rural AGEB.
  - Urban AGEB.

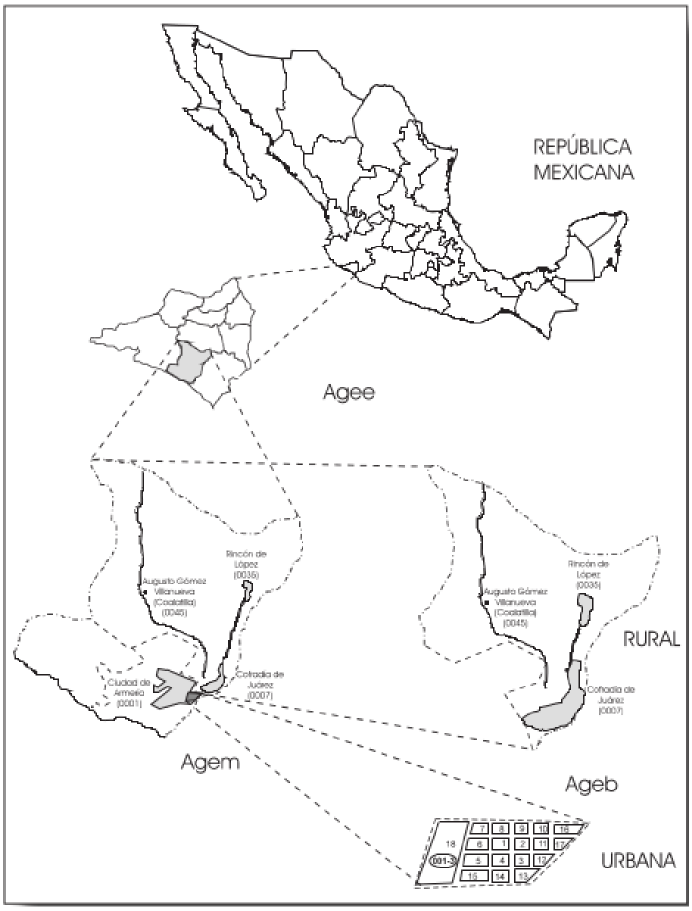


Figure 3.1: MG dissaggregation areas. Retrieved from: [3].

Urban AGEBs are the geographic area, subdivision of municipal areas, occupied by a set of blocks, generally ranging from 1 to 50, perfectly delimited by streets, avenues, walkways or

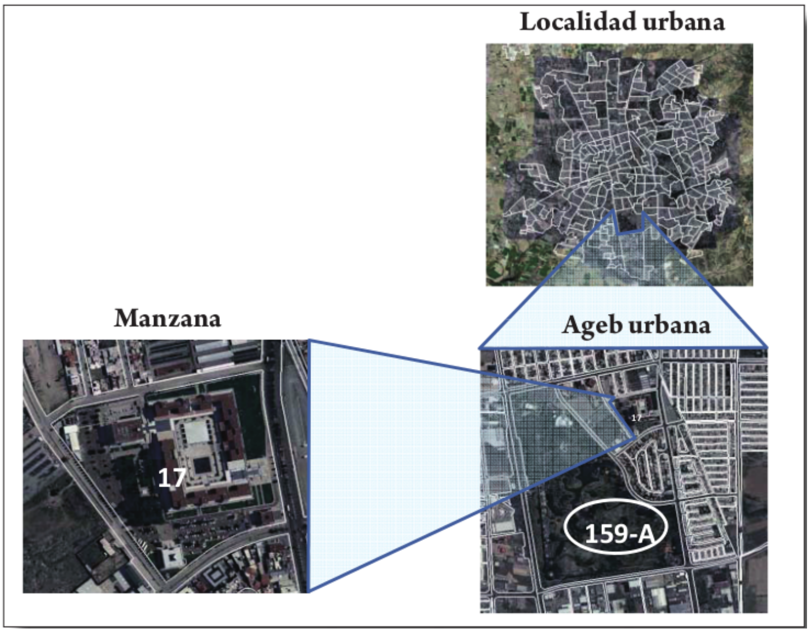
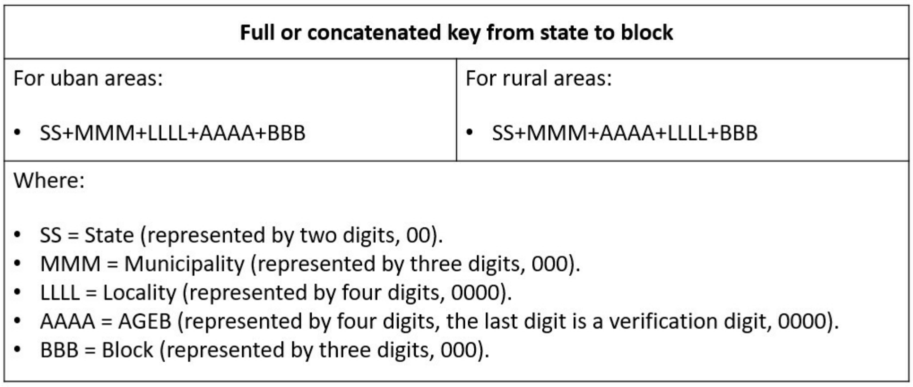


Figure 3.2: Urban AGEB dissaggregation areas. Retrieved from: [3].

any other easily identifiable feature on the ground and whose land use is mainly residential, industrial, services, commercial, etc., only assigned within urban localities (see Figure 3.2).

We download the MG data from [64], which contains the shapefiles of every dissaggregation area of every mexican state. It is made up of 32 folders, each one named by the geoestatistical key of the federal entity (from 1 to 32), with a national total of 2,469 municipal geoestatistical areas, 45,397 polygons of rural localities, and 4,911 polygons of urban localities, 295,779 points of rural localities, 350 polygons of island territory, 17,469 basic rural geostatistical areas, 63,982 basic urban geostatistical areas and 2,513,853 urban and rural blocks (including scattered hamlets); the information maintains associated names and geostatistical keys as attributes.

The MG references every dissaggregation area with a unique numeric key. The structure of such geostatistical key is represented in Figure 3.3.



Full or concatenated key from state to block	
For uban areas: • SS+MMM+LLLL+AAAA+BBB	For rural areas: • SS+MMM+AAAA+LLLL+BBB

Where:

- SS = State (represented by two digits, 00).
- MMM = Municipality (represented by three digits, 000).
- LLLL = Locality (represented by four digits, 0000).
- AAAA = AGEB (represented by four digits, the last digit is a verification digit, 0000).
- BBB = Block (represented by three digits, 000).

Figure 3.3: MG's geostatistical key structure. Retrieved from: [3].

Every state folder of the MG is composed of three subfolders:

- Catalogs (catálogos): contains the product catalogs and documentation.
- Data set (conjunto_de_datos): contains 32 folders, each one corresponding to the state geoestatistical key.
- Metadata (metadatos): contains 32 files, each one with the corresponding state geostatistical key, in xml and txt format, and a generic metadata with national information.

The file names are formed with the state geoestatistical key and the following suffixes of the file content:

Where **ee** corresponds to the state geoestatistical key (from 01 to 32).

ee**ent**	State geoestatistical areas
ee**mun**	Municipal geoestatistical areas
ee**ar**	Basic rural geoestatistical areas
ee**l**	Polygon of urban and rural localities
ee**lpr**	Rural point locations
ee**ti**	Island territory
ee**a**	Basic urban geoestatistical areas
ee**m**	Block polygons
ee**fm**	Block fronts
ee**e**	Road axes
ee**cd**	Scattered hamlet
ee**sia**	Complementary area-type services and information (green areas, medians, traffic circles)
ee**sil**	Complementary line-type services and information (rivers, railroads, streams)
ee**sip**	Complementary point-type services and information (municipal palaces, parks or gardens, etc.)
ee**pe**	External polygon
ee**pem**	External polygon of blocks

Layers with suffix **ti**, **cd**, **pe**, **pem**, **sia**, **sil**, **sip**, are included only if the locality has this type of information.

INEGI's 2020 census data was downloaded from [63] on the main results by AGEB and urban block subsection. In this subsection we can download data from every mexican state in a CSV (comma-separated values) file.

In this work, we are only using the MG Yucatán folder (31_yucatan) and the file of the basic urban geoestatistical areas (31a.shp), as well as census data from Yucatán.

3.4 Data Exploration and Preparation

3.5 Data Modeling

4 Results

Table 4.1: Selected measures of all Mérida street network.

measure	
Area (km^2)	1032.365
Avg of the avg neighborhood degree	2.843
Avg of the avg weighted neighborhood degree	0.045
Avg circuity	0.964
Avg clustering coefficient	0.030
Avg weighted clustering coefficient	0.001
Intersection count	31837
Avg degree centrality	<0.001
Edge density (km/km^2)	9013.845
Avg edge length (m)	99.662
Total edge length (km)	9305581.721
Proportion of dead-ends	0.091
Proportion of 3-way intersections	0.592
Proportion of 4-way intersections	0.312
Intersection density (per km^2)	30.839
m	93371
n	35031
Node density (per km^2)	33.933
Max PageRank value	<0.001
Min PageRank value	<0.001
Self-loop proportion	0.001
Street density (km/km^2)	5259.333
Average street segment length (m)	99.177
Total street length (km)	5429553.502
Street segment count	54746
Average streets per node	3.130

Table 4.2: Central tendency and statistical dispersion for selected measures of all Mérida urban AGEB's street networks: μ is the mean, σ is the standard deviation, and D is the dispersion index $\frac{\sigma^2}{\mu}$.

measure	μ	σ	min	median	max	D
Area (km^2)	0.582	0.527	0.014	0.475	7.613	0.477
Avg of the avg neighborhood degree	2.477	0.470	0.500	2.556	3.414	0.089
Avg of the avg weighted neighborhood degree	0.039	0.012	0.005	0.039	0.097	0.004
Avg circuity	0.960	0.066	0.932	0.940	1.741	0.005
Avg clustering coefficient	0.027	0.044	0	0.017	0.583	0.072
Avg weighted clustering coefficient	0.007	0.02	0	0.003	0.397	0.057
Intersection count	53.393	29.065	1	52	204	15.664
Avg degree centrality	0.156	0.270	0.015	0.088	2	0.467
Edge density (km/km^2)	23657.180	8967.089	1543.392	23866.890	47341.220	3398.913
Avg edge length (m)	92.625	23.339	44.488	88.761	237.350	5.881
Total edge length (km)	12152.490	7014.459	88.976	11467.030	49217.050	4048.771
Proportion of dead-ends	0.069	0.089	0	0.031	0.500	0.115
Proportion of 3-way intersections	0.566	0.179	0	0.585	1	0.057
Proportion of 4-way intersections	0.373	0.190	0.006	0.349	1	0.097
Intersection density (per km^2)	112.849	54.826	8.403	108.692	410.011	26.636
Average node degree	1.348	0.488	0.163	1.333	2.772	0.177
m	136.035	11.534	2	130	528	44.191
n	58.193	33.431	2	54	264	19.206
Node density (per km^2)	120.451	57.043	9.226	116.253	410.011	27.014
Max PageRank value	0.063	0.082	0.001	0.039	0.500	0.107
Min PageRank value	0.016	0.060	0.001	0.003	0.500	0.225
Self-loop proportion	0.001	0.004	0	0	0.046	0.016
Street density (km/km^2)	13743.660	4999.535	774.708	13920.640	28345.930	1818.683
Average street segment length (m)	92.472	22.521	44.488	88.585	198.010	5.485
Total street length (km)	7133.770	4115.884	44.488	6748.335	32542.370	2374.691
Street segment count	79.803	45.775	1	76	321	26.257
Average streets per node	3.225	0.320	2	3.25	4	0.032

Table 4.3: Median values, aggregated by towns, of selected measures of the neighborhood-scale street network Mérida's urban AGEBs.

Town	Intersect density (per km^2)	Avg streets per node	Avg circuity	Avg street segment length
Mérida	112.24	3.27	0.94	87.65
Caucel	37.02	3.17	0.94	142.02
Chablekal	30.71	3	0.93	144.79
Cholul	55	2.84	0.98	109.71
Komchén	34.68	2.95	1	131.11
San José Tzal	28.21	2.62	0.94	134.50

5 Conclusions and recommendations

5.1 Conclusion

5.2 Recommendations

5.3 Future work

Bibliography

[1] Pal Nikolli and Bashkim Idrizi. Coordinate Reference Systems Used in Albania to Date. FIG Working Week 2011, Morocco, 05 2011.

[2] Chrismurf. Utm-zones-USA.svg. English Wikipedia, 2009. Retrieved from https://commons.wikimedia.org/wiki/File:Utm-zones-USA.svg.

[3] Instituto Nacional de Estadística y Geografía. Manual of Geostatistical Cartography. INEGI, 2010. Retrieved from: https://www.inegi.org.mx/contenidos/temas/mapas/mg/metadatos/manual_cartografia_censal.pdf.

[4] Marc Barthélemy. Spatial networks. Physics Reports, 499(1):1–101, February 2011.

[5] Taylor Anderson and Suzana Dragićević. Complex spatial networks: Theory and geospatial applications. Geography Compass, 14, July 2020.

[6] Geoff Boeing. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems, 65:126–139, September 2017.

[7] Leonhard Euler. Solutio problematis ad geometriam situs pertinentis (the solution to a problem relating to the geometry position). Commentarii Academie Scientiarum Imperialis Petropolitanae, 8:128–140, 1741.

[8] Albert-László Barabasi. Network science. Cambridge University Press, Glasgow, UK, 2016.

[9] Sergio Porta, Paolo Crucitti, and Vito Latora. The network analysis of urban streets: A primal approach. Environment and Planning B, 33(5):705–725, 2006.

[10] Carlo Ratti. Space syntax: Some inconsistencies. Environment and Planning B, 31(4):487–499, 2004.

[11] Ted G. Lewis. Network science: Theory and applications. John Wiley & Sons, Hoboken, NJ, September 2011.

[12] Mark. E. J. Newman. Analysis of weighted networks. Physical Review E, 70(5), Nov 2004.

[13] Mark E. J. Newman. Networks: An introduction. Oxford University Press, Oxford, England, 2010.

[14] Filippo Menczer, Santo Fortunato, and Clayton A. Davis. A First Course in Network Science. Cambridge University Press, 2020.

[15] D. J. Watts and S. H. Strogatz. Collective dynamics of ’small-world’ networks. Nature, 393(6684):440–442, 1998.

[16] David J Giacomin and David M Levinson. Road network circuity in metropolitan areas. Environment and Planning B: Planning and Design, 42(6):1040–1053, 2015.

[17] Dean Urban and Timothy Keitt. Landscape connectivity: A graph-theoretic perspective. Ecology, 82:1205–1218, 05 2001.

[18] Per Hage and Frank Harary. Eccentricity and centrality in networks. Social Networks, 17(1):57–63, 1995.

[19] Lowell Beineke, Ortrud Oellermann, and Raymond Pippert. The average connectivity of a graph. Discrete Mathematics, 252:31–45, 05 2002.

[20] Peter Dankelmann and Ortrud R. Oellermann. Bounds on the average connectivity of a graph. Discrete Applied Mathematics, 129(2):305–318, 2003.

[21] Xiaoke Huang, Ye Zhao, Chao Ma, Jing Yang, Xinyue Ye, and Chong Zhang. Trajgraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data. IEEE Transactions on Visualization and Computer Graphics, 22(1):160–169, January 2016.

[22] Chen Zhong, Markus Schläpfer, Stefan Arisona, Michael Batty, Carlo Ratti, and Gerhard Schmitt. Revealing centrality in the spatial structure of cities from human activity patterns. Urban Studies, 54, 10 2015.

[23] L. C. Freeman. A set of measures of centrality based on betweenness. Sociometry, 40(1):35–41, 1997.

[24] Alireza Ermagun and David Levinson. An introduction to the network weight matrix: Introduction to the network weight matrix. Geographical Analysis, 50, 07 2017.

[25] Mark E. Newman. The structure and function of complex networks. SIAM Review, 45(2):167–256, 2003.

[26] Bin Jiang. Ranking space for predicting human movement in an urban environment. International Journal of Geographical Information Science, 23, 12 2006.

[27] Taras Agryzkov, José Oliver, Leandro Tortosa, and José-Francisco Vicent. An algorithm for ranking the nodes of an urban network based on the concept of PageRank vector. Applied Mathematics and Computation, 219:2186–2193, 11 2012.

[28] David Gleich. PageRank Beyond the Web. SIAM Review, 57, 07 2014.

[29] Wei Chien Benny Chin and Tzai-Hung Wen. Geographically Modified PageRank Algorithms: Identifying the Spatial Concentration of Human Movement in a Geospatial Network. PLoS ONE, 10, 10 2015.

[30] Sergio Porta, Paolo Crucitti, and Vito Latora. The network analysis of urban streets: A dual approach. 11 2004.

[31] Sergio Porta, Paolo Crucitti, and Vito Latora. The network analysis of urban streets: A primal approach. Environment and Planning B: Planning and Design, 33:705–725, 02 2006.

[32] S. Porta, V. Latora, and E. Strano. Networks in urban design. six years of research in multiple centrality assessment, pages 107–130. Springer, September 2010.

[33] Paolo Crucitti, Vito Latora, and Sergio Porta. Centrality in Network of Urban Streets. Chaos (Woodbury, N.Y.), 16:015113, 04 2006.

[34] Paolo Crucitti, Vito Latora, and Sergio Porta. Centrality Measures in Spatial Networks of Urban Streets. Physical review. E, Statistical, nonlinear, and soft matter physics, 73:036125, 04 2006.

[35] Andres Sevtsuk and Michael Mekonnen. Urban network analysis. A new toolbox for ArcGIS. Revue internationale de géomatique, 22:287–305, 06 2012.

[36] Santo Fortunato. Community detection in graphs. Physics Reports, 486(3-5):75–174, 2010.

[37] Marc Bathélemy. Morphogenesis of spatial netwoks. Springer International Publishing, Cham, Switzerland, 2018.

[38] Mark. E. J. Newman and Michelle Girvan. Finding and evaluating community structure in networks. Physical Review E, 69(2), February 2004.

[39] Charles Fox. Spatial Data. In Data Science for Transport, pages 57–74. Springer International Publishing, Cham, 2018.

[40] Miljenko Lapaine and E. Lynn Usery, editors. Choosing a Map Projection. Lecture Notes in Geoinformation and Cartography. Springer International Publishing, Cham, 2017.

[41] Geographic information - Simple feature access - Part 1: Common architecture. Standard ISO 19125-1:2004, International Organization for Standardization, 2004.

[42] OpenStreetMap Wiki. Retrieved from `https://wiki.openstreetmap.org/wiki`.

[43] Mexico Main Road Network Import Project - OpenStreetMap Wiki. Retrieved from `https://wiki.openstreetmap.org/wiki/Mexico_Main_Road_Network_Import_Project`.

[44] Mexico's Administrative Divisions Import Project - OpenStreetMap Wiki. Retrieved from `https://wiki.openstreetmap.org/wiki/Mexico%27s_Administrative_Divisions_Import_Project`.

[45] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring Network Structure, Dynamics, and Function using NetworkX. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, Proceedings of the 7th Python in Science Conference, pages 11 – 15, Pasadena, CA USA, 2008.

[46] NetworkX developers. NetworkX.

[47] The GeoPandas development team. geopandas/geopandas: v0.8.2, January 2021.

[48] The pandas development team. pandas-dev/pandas: Pandas, February 2020.

[49] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, Proceedings of the 9th Python in Science Conference, pages 56 – 61, 2010.

[50] Sean Gillies et al. Shapely: manipulation and analysis of geometric objects, 2007–2020.

[51] Sean Gillies et al. Fiona is OGR's neat, nimble, no-nonsense API, 2011–2020.

[52] Sean Gillies et al. descartes 1.1.0, 2017.

[53] The Matplotlib development team. matplotlib/matplotlib: Rel: v3.4.0rc1, February 2021.

[54] J. D. Hunter. Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3):90–95, 2007.

[55] Sergio J. Rey and Luc Anselin. PySAL: A Python Library of Spatial Analytical Methods. The Review of Regional Studies, 37(1):5–27, 2007.

[56] David O'Sullivan. Spatial Network Analysis, pages 1253–1273. 07 2014.

[57] Geoff Boeing. The Morphology and Circuity of Walkable and Drivable Street Networks, pages 271–287. Springer International Publishing, Cham, 2019.

[58] Geoff Boeing. Planarity and street network representation in urban form analysis. Environment and Planning B: Urban Analytics and City Science, 47(5):855–869, 2020.

[59] Geoff Boeing. A multi-scale analysis of 27,000 urban street networks: Every US city, town, urbanized area, and Zillow neighborhood, August 2018.

[60] Geoff Boeing. Exploring Urban Form through OpenStreetMap Data: A Visual Introduction, pages 167–184. Routledge, Abingdon, England, 2020.

[61] Anaconda Inc. Anaconda Software Distribution. Anaconda Documentation, 2021.

[62] MapTiler Team. WGS 84 / Pseudo-Mercator – Spherical Mercator, Google Maps, OpenStreetMap, Bing, ArcGIS, ESRI. epsg.io, 2019.

[63] Instituto Nacional de Estadística y Geografía. Censo de Población y Vivienda 2020. INEGI, 2020.

[64] Instituto Nacional de Estadística y Geografía. Marco Geoestadístico. Censo de Población y Vivienda 2020. INEGI, 2020. Retrieved from: `https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=889463807469`.