

Topological street-network characterization through feature-vector and cluster analysis

Gabriel Spadon, Gabriel Gimenes, and Jose F. Rodrigues-Jr

University of Sao Paulo, Sao Carlos – SP, Brazil

spadon@usp.br, ggimenes@icmc.usp.br, junio@icmc.usp.br

Abstract. Complex networks provide a means to describe cities through their street mesh, expressing characteristics that refer to the structure and organization of an urban zone. Although other studies have used complex networks to model street meshes, we observed a lack of methods to characterize the relationship between cities by using their topological features. Accordingly, this paper aims to describe interactions between cities by using vectors of topological features extracted from their street meshes represented as complex networks. The methodology of this study is based on the use of digital maps. Over the computational representation of such maps, we extract global complex-network features that embody the characteristics of the cities. These vectors allow for the use of multidimensional projection and clustering techniques, enabling a similarity-based comparison of the street meshes. We experiment with 645 cities from the Brazilian state of Sao Paulo. Our results show how the joint of global features describes urban indicators that are deep-rooted in the network’s topology and how they reveal characteristics and similarities among sets of cities that are separated from each other.

Keywords: Network Topology; Feature Vector; Cluster Analysis.

1 Introduction and Related Works

Complex networks are used to shape real-world systems, *e.g.* networks of protein interaction, street meshes, and subway lines. These networks, as mathematical models, stand out due to their algebraic properties and computing potential, with analytical applicability to support cognitive processes of decision-making [1]. Through metrics and methods based on topology and/or geometry, it is possible to identify characteristics of interest that are not obvious for human inspections based on reading; this is because the networks may be wide (high number of vertices), intricate (high number of edges), or may hold non-trivial patterns and attributes whose observation depends on the application of algorithms.

In the specific case of the representation of street networks, complex networks describe factors related to the displacement of individuals, allocation of services, the improvement of tasks related to transport, and even to the study of factors from collective behavior, when the network is weighted by the associated data.

In this regard, we observed a lack of methods to characterize groups of cities by means of the features that can be extracted from their topology, which is the aim of this research. This methodology has potential to enhance the understanding of an urban space and to explain the reason why cities share properties of interest.

To this end, we developed a methodology composed of *Data Acquisition and Preparation*, *Feature Extraction and Selection*, and *Feature Vector Analysis*. We analyzed 645 cities from the state of Sao Paulo, aiming to provide comprehension of peculiarities from different cities by interpreting global network-characteristics. These cities are representations of street meshes that were extracted from digital maps, such that they were gathered and analyzed by using machine-learning methods of feature extraction, multidimensional projection, and cluster analysis. In order to demonstrate our methodology, we investigate the following hypotheses: **(A)** *the network topology is a tool-set that can reveal groups of cities with similar characteristics, potentially revealing disparities*; **(B)** *although cities may share administrative boundaries with others, they cluster with cities with no apparent geographical similarity*; and, **(C)** *there might be interesting correlations between urban and/or territorial indicators and the features extracted from the street-network topology of a given set of cities*. The answering of such assumptions allows us to render better analysis of urban agglomerations by helping in the understanding of cities by comprehending how they are arranged within the geographical extent of their territorial boundaries.

Aiming to solve questions related to the urban scenario, a vast number of studies have been conducted to explain cities considering their intense flow of vehicles [2] and collective behaviors [3], while others analyzed the accidents density in street networks [4] and the discrepancies between cities driven by their urban indicators [5]. Furthermore, some authors investigated metrical and analytical methods applied to cities [6, 7], others approached the assistance to the urban planning and design [8–10], and there are those who advanced with facility-location analysis and planning in street meshes [11]. However, although cluster analysis has been less focused [12, 13], it is still an important toolset [14].

Two state-of-the-art works used clustering techniques to analyze groups of cities, but both of them left open questions to be explored. The first one had the intention to measure the similarity among ten European cities [12], while the second one performed an eye-based cluster evaluation considering the proximity and overlap of 1,150 cities, mainly from the Anglo-Saxon America [13]. Their lack of proficiency is mainly because they do not employ clustering algorithms in the same fashion that we do, including validation metrics and analytical indicators.

In this paper, we contribute with a methodology that advances the analysis of cities modeled as complex networks. To present our contributions, this paper is organized as follows: Section 2 displays our methodology while explaining the validation of its results; Section 3 discusses the results about the applicability of the proposed methods; and, Section 4 presents the conclusions and final remarks.

2 Methodology

Our methodology is based on the intersection of methods of data **A**cquisition, **M**odeling, and **C**omputation, and it follows a process flow depicted in Figure 1.

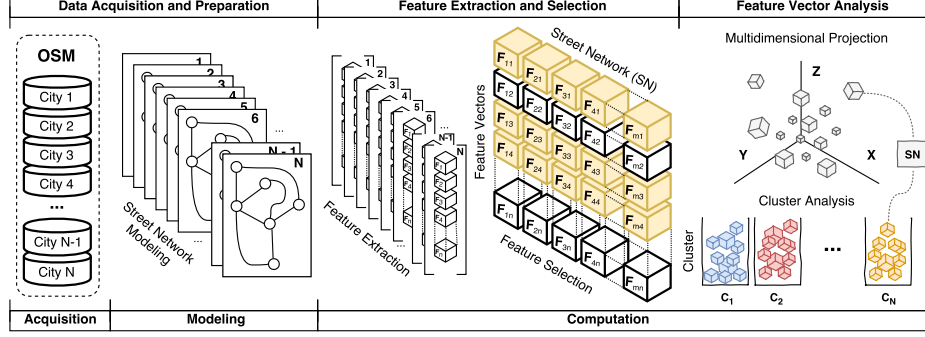


Fig. 1: Methodology for street-network characterization through feature-vector and cluster analysis based on data **A**cquisition, **M**odeling, and **C**omputation. The methodology starts by acquiring digital maps of cities from the OpenStreetMap (OSM), such maps are used for the modeling of complex networks. The resulting networks are used in the processes of extraction and selection of topological-features. These features are analyzed according to data-mining methods of multidimensional projection and cluster detection.

2.1 Preliminaries

Hereinafter, we represent complex networks as distance-weighted directed graphs. Notice that, despite different, complex networks and graphs are considered to be equivalent. A graph $\mathbf{G} = \{V, E\}$ is composed of a set of $|V|$ nodes and a set of $|E|$ edges. Furthermore, each edge $e \in E$ is known to be an ordered pair $\langle o, d \rangle$, in which $o \in V$ is named *origin* and $d \in V$ is named *destination*, $o \neq d$. We provided to the edges a double-precision floating-point weight d_{od} , which refers to the *great-circle distance* between node o and node d . The *great-circle distance* refers to the Euclidean distance between two points on the surface of a sphere; which in our case, the sphere is a projection of the Earth.

2.2 Data acquisition and preparation

For each one of the 645 cities from the Brazilian state of Sao Paulo, we got their administrative boundaries and indicators related to territorial extension and demography from the Brazilian Institute of Geography and Statistics (IBGE)¹. The boundaries served as shapefiles to crop data obtained from OpenStreetMap (OSM)², which is an open data repository and a social network of collaborative

¹ www.ibge.gov.br ² www.openstreetmap.org

street mapping. The OSM's data describe real-world abstractions represented by georeferenced objects. These objects are described by means of its relations, which, in turn, refer to the streets (edges) and crossings (nodes) of a city, which were turned into complex networks where the edges intersect only at the nodes.

2.3 Feature extraction and selection

Metrics of graphs, referred to as *features*, can be divided into local and global [15]; local metrics describe properties of individual elements that form the network, while global metrics characterize the whole network by a single value that is

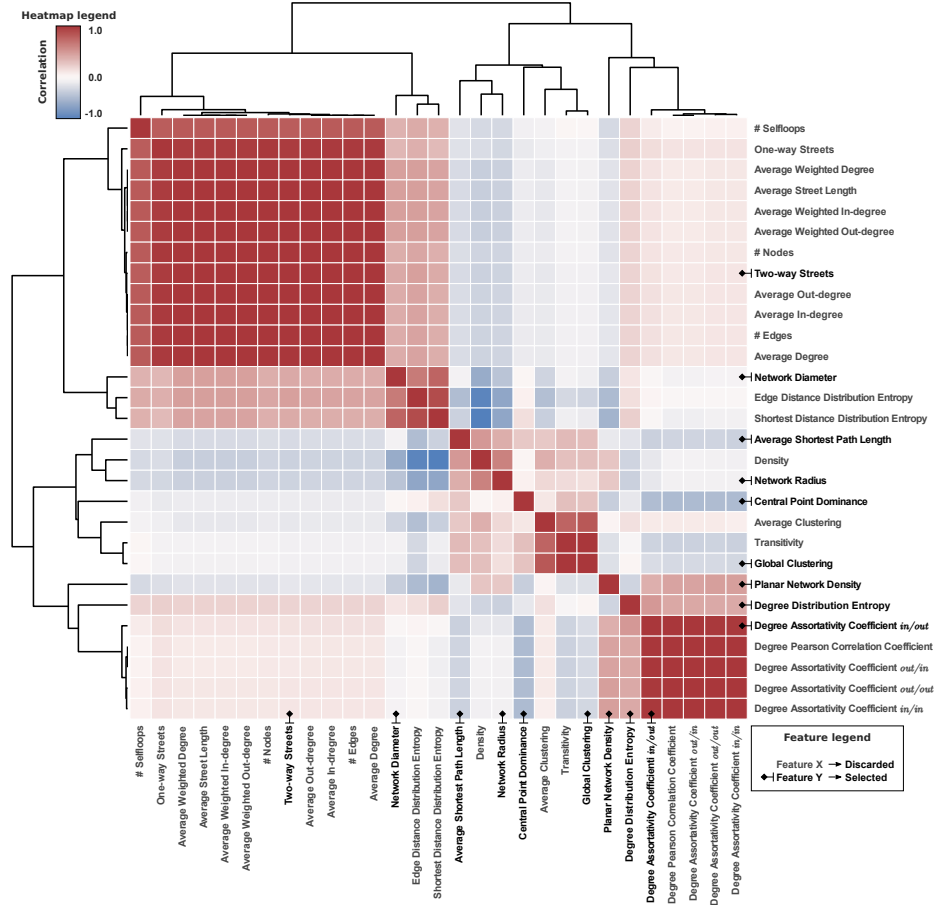


Fig. 2: A visualization of the mutual-correlation matrix of all the metrics we considered. The color describes the correlation between pairs of features. The metrics were hierarchically grouped through a dendrogram by means of the correlation of their values. Consequently, correlated metrics tend to stay in the same group; non-correlated metrics tend to be in separated groups. Additionally, the metrics we selected are colored in black and highlighted by a diamond marker.

computed considering all of their elements. We rather use global metrics than local ones because they allow straightforward comparison between cities.

In order to gather these metrics, we designed a feature extractor to calculate a feature vector from any complex network given as input. First, we selected various metrics as candidates to render characteristics about cities, from which 29 metrics were chosen by their potential in providing insights about a given street network (see Figure 2 for details). Such metrics were selected because they are linked to the network topology, which describes the streets of the cities.

After collecting all the metrics, we removed the non-relevant ones based on their mutual correlation. We computed the Pearson correlation coefficient [16] for each pair of metrics; such coefficient is defined in the interval $[-1.0, 1.0]$ where the extreme values indicate, respectively, the maximum negative and positive correlation, while 0.0 indicates no linear correlation at all. Following, we removed all the metrics with strong mutual correlation as indicated by the Pearson correlation in the interval $[-0.5, 0.5]$. In cases where any two metrics are outside this interval, one of the metrics was randomly discarded. Such process of metrics selection ensures that just metrics that are unique and non-related with the others will be used to describe the cities. Other processes of feature selection can be used in this step; even the multidimensional projection by itself can provide reasonable results. Notice that, the reduction of the dimensionality of the data was not our main priority, but rather to find the most complete set of metrics, that is the one that better characterizes the networks, without including redundant information; and, to this end, features correlation plays an import role. All metrics are depicted in Figure 2; the ones that remained, 9 out of 29, were highlighted and are defined according to Costa *et al.* [17], as follows:

Degree Distribution Entropy (\mathcal{H}). The degree distribution of a network describes the probability of finding a vertex with a given degree. Whereas, the entropy represents the amount of uncertainty and randomness in a certain piece of information. By using the entropy in a city degree distribution, we can measure the uncertainty between street connections. Equation 1 describes such metric, where P_k represents the ratio of nodes with degree k .

Average Shortest Path (\mathcal{L}). It quantifies the average of all shortest paths (d_{ij}^S) that link all the pairs of nodes in a complex network (Equation 2), it is used to quantify the capacity of locomotion through the shortest paths of a city.

$$\mathcal{H} = - \sum_{k=0}^{\infty} P_k \times \log(P_k) \quad (1) \quad \mathcal{L} = \frac{\sum_{i=1}^{|V|} \sum_{j=1}^{|V|} d_{ij}^S}{|V|(|V| - 1)} \quad (2)$$

Degree Assortativity Coefficient (\mathcal{R}). It refers to the *in and/or out* degree correlation between pairs of nodes. That is, positive values indicate that nodes with similar degrees tend to connect to each other, while negative values indicate the same, but regarding nodes with different degrees. It can be understood as the probability of moving from an unimportant street to an important one based only on the number of adjacent streets to both of them. Equation 3 uses e_{xy} to

refer to the fraction of edges that join together vertices with degree x and y , a_x and b_y to the fractions of edges that start and end at vertices with degree x and y ; and, σ_a and σ_b to the standard deviations of the distributions a_x and b_y .

Eccentricity (\mathcal{E}). This metric is a local one, measuring for a vertex the longest shortest distance between all the other vertices of a given graph [18] (see Equation 4). In a global perspective, the greatest eccentricity from a graph is known to be the *network diameter*, while the smallest one is regarded as the *network radius*. They can reveal cities that may suffer from access issues by being sparse if the radius of a network is too small when compared to its diameter.

$$\mathcal{R} = \frac{\sum_{xy} xy(e_{xy} - a_x b_y)}{\sigma_a \sigma_b} \quad (3) \quad \mathcal{E}_i = \frac{1}{\max\{d_{ij}^S | \forall j \in V\}} \quad (4)$$

Planar Network Density (\mathcal{D}). The density of a planar graph is defined as the ratio between the number of edges E and the number of all possible edges in a network with N nodes with no intersecting edges. It can be used to describe how dense is the street mesh of a city or a neighborhood. The metric is unique to each network, once the position of the nodes interferes in the number of edges. It is an algorithmic adaptation of the graph density [19], described in Equation 5.

Central Point Dominance (C_D^P). This metric assesses the global centrality of a whole network by means of its network's betweenness deviation, which is a distance-based centrality metric. Values close to 0 indicate plenty of distance-efficient routes similar to the shortest one; whereas, values close to 1 indicate that the network might become vulnerable without its central node because the node might be used to connect different components, serving as an access point (*e.g.* bridges and tunnels). In Equation 6, \bar{v} is the node with the highest betweenness and $\mathcal{B}(v)$ is the normalized betweenness of the node v that lies in the range $[0, 1]$.

$$\mathcal{D} = \frac{|E|}{|N|(|N| - 1)} \quad (5) \quad C_D^P = \frac{\sum_v^{|V|} \mathcal{B}_{\bar{v}} - \mathcal{B}_v}{|V|(|V| - 1)} \quad (6)$$

Two-way Streets (\mathcal{T}_w). It refers to the number of double edges in a network, which are edges that provide two-way routes between the same pair of nodes. This metric follows Equation 8, in which f_{ij} is a clause-based auxiliary function.

Global Clustering (\mathcal{G}_c). The metric, which is described by Equation 8, consists of the *fraction* of the number of triangles \mathbb{N}_Δ and triples \mathbb{N}_3 of the network. It refers to how the streets tend to cluster in the crossings of a given city, such that the greater the value the more possibilities of locomotion in fewer steps.

$$\mathcal{T}_w = \frac{\sum_{\langle i,j \rangle}^E f_{ij}}{2}, \quad f_{ij} = \begin{cases} 1, & \langle j, i \rangle \in E \\ 0, & \text{otherwise} \end{cases} \quad (7) \quad \mathcal{G}_c = \frac{(3 \times \mathbb{N}_\Delta)}{\mathbb{N}_3} \quad (8)$$

2.4 Feature vector analysis

In this step, we focused on two methods from the data mining literature, the first one of multidimensional projection and the second one of clustering detection. Multidimensional projection allows the visualization of data by reducing its dimensional space, revealing particularities and behaviors to be explored through cluster-based analysis. Cluster analysis, in turn, focuses on the study of data interactions, inferring that two elements are similar because they are in the same cluster or dissimilar because they are in different ones. Consequently, the combination of these two methods contributes to the assessment of cities by their potential to reveal patterns that are not evident through an eye-based analysis.

Regarding multidimensional projection, our methodology consists of using two techniques [20]; the first one is named Isomap and the second one is known as Principal Components Analysis (PCA). Isomap is a nonlinear dimensionality reduction technique, which provides an embedding in a lower dimension while maintaining the geodesic distance between the data elements. Contrarily, PCA is a linear technique, which uses orthogonal conversions to turn a set of variables into linearly uncorrelated values with the largest possible variance. To choose both techniques, we used knowledge about the domain; we have kept track of some already-known dissimilar cities, seeking for approaches to distinguish them.

In the cluster analysis part, we used the technique KMeans [21], which splits the data into groups of equal variance, minimizing the sum-of-squares distance within clusters. The KMeans algorithm assumes that (i) the distribution of features within each cluster resembles spheres, which means that all features have equal variance and they are independent of each other; (ii) regarding the cluster size, the dataset is balanced; and, (iii) the density of the clusters is similar. The dataset we used consists of uncorrelated values and balanced instances of feature vectors, all of which have quasi-equal variance, meeting the algorithm requirements. In addition, KMeans is widely used in the related literature due to its robustness, versatility, and scalability. To validate our results we considered cluster quality metrics [22]. Their focus is to analyze the similarity between elements that have been assigned to the same cluster. We used a combination of the Silhouette score [23] and the Dunn index [24]; both of which are known to be internal-quality metrics, not requiring a pre-labeled dataset. The Silhouette is defined between $[-1, 1]$ for each cluster, the closer to 1 the better; it measures the cohesion and separation of clusters by evaluating how similar an element is in its own cluster when contrasted to other clusters. To further enhance the reliability of our analysis, we applied the Dunn index, which is a cluster distance-based quality metric that measures the separation among clusters, whose values are in between of $[0, \infty[$. In cases when the Dunn's index distance is greater than one, there is little or none cluster overlapping. Using both together, we have a double validation of quality by means of cohesion and separation of our set of clusters.

3 Results

3.1 Relationship of population size and topological features

With regard to the population density — see Figure 3 for details —, the majority of the cities in our dataset is of tiny or small size, but the dataset has a substantial number of medium-sized cities and a small number of large-sized ones, including Sao Paulo — the biggest Brazilian city. Prior analyses can be done by observing Figure 4, where cities (depicted as points) were sized by their number of nodes.

A first evidence that the topological features we selected can describe relevant knowledge about cities is the fact that Sao Paulo is isolated from the other ones in the PCA projection. A similar fact can be observed on a small scale considering the large-sized city of Campinas and the medium-sized cities of Marilia and Piracicaba, which are apart from the main group of cities located on the left part of the image. We believe that such behavior is connected to the demographics of the cities.

On a large scale, topological features can predict demographic characteristics of a city, whereas, on a small scale, they can reflect the neighborhoods that are densely or sparsely populated. For a less unbalanced view, we removed Sao

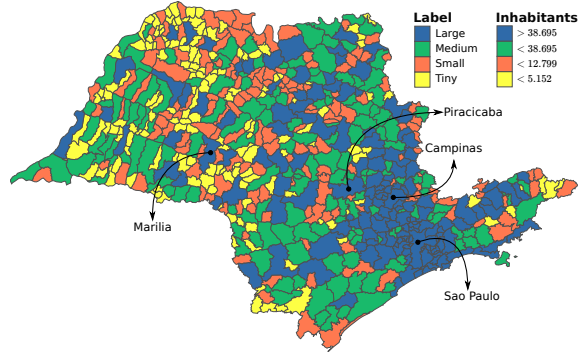


Fig. 3: Urban indicator related to the population density of the cities of the state of Sao Paulo. The cities were divided into four classes that describe the number of inhabitants of each one of them.

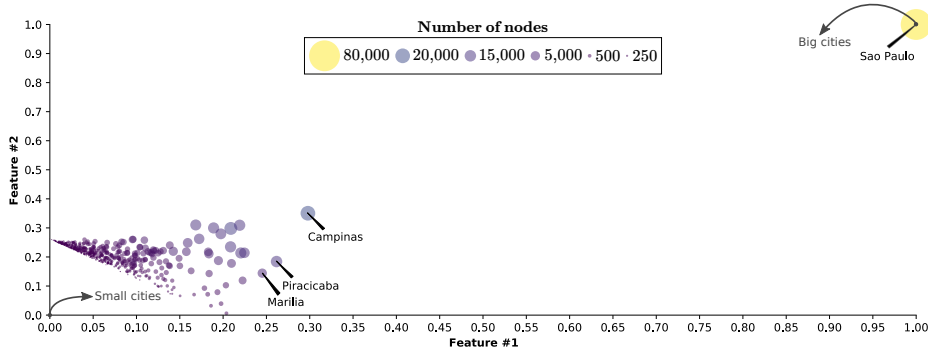


Fig. 4: Projection of feature vectors in two dimensions by using PCA; the size of the points refers to the number of nodes (intersections) in the cities' complex-network. The projected features reveal that the city of Sao Paulo (on the right-hand side) is an outlier when compared to the others (on the left-hand side).

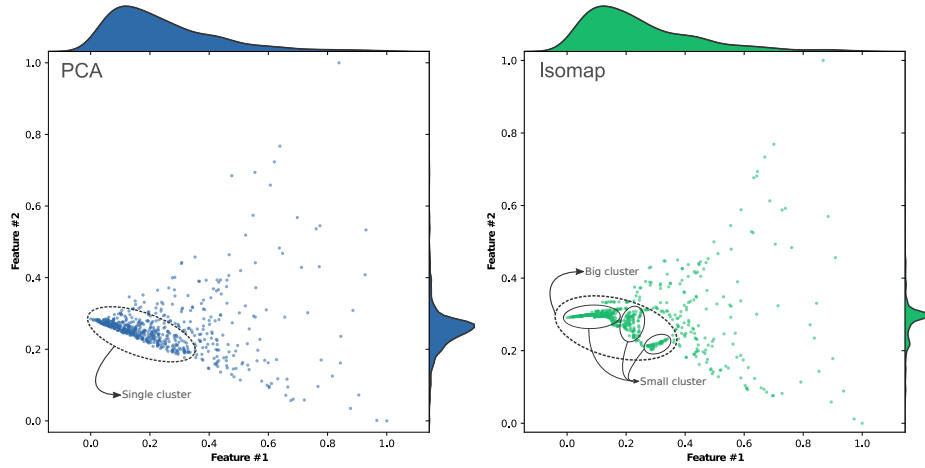


Fig. 5: Projection of the features, excluding Sao Paulo, using PCA and Isomap. PCA shows a single dense area with many sparse data, while Isomap shows multiple dense areas together with several sparse data. As a consequence, PCA implies a single cluster while Isomap points to an inherent hierarchy of clusters.

Paulo from the dataset, depicting in Figure 5 the normalized values of the feature vectors of the cities that remained using both PCA and Isomap techniques.

The two techniques show us that the majority of the data is concentrated in a small region, while the rest of it is sparse and distributed along the axes. The main difference between both of them is that Isomap implies multiple areas with considerable density, while PCA has a single dense area and many sparse data. This is evidence that tiny and small-sized cities tend to cluster isolating medium and large-sized cities that are too different from them. Despite the fact that such cities tend to cluster, Isomap shows that they have particularities

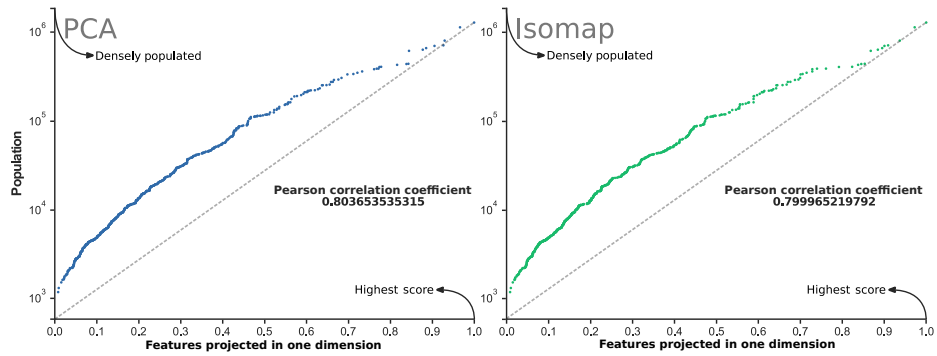


Fig. 6: Correlation test between the population density and a one-dimension projection of the cities' topological-features regarding both PCA and Isomap. Both images show a strong correlation, revealing that, on a large scale, the topological features of the cities can indicate, or even predict, their demography.

that make them split into smaller clusters inside a bigger one. Also, it is safe to infer that by being scattered, medium-sized and large-sized cities have no clear pattern, but still, they may share common characteristics to be further explored with clustering algorithms. Even so, we can show, by using correlation, that the network’s demography can be inferred from the city’s topology – see Figure 6.

To prove that the network’s demography can be inferred from the city’s topology, we measured the relationship between the topological features and the demography by means of correlation. To this end, we reduced the dimensionality of the feature vectors of each city to one, using both techniques, PCA and Isomap, resulting in one single value for each one of the 645 cities. Next, we correlated such values with the size of their population. As a result, we got 0.803 and 0.799 of correlation for PCA and Isomap, respectively. Both values indicate that the data has a strong correlation, allowing us to state that in the case of the Brazilian state of Sao Paulo, topological features and demographics are strongly correlated. Such pattern opens doors for new investigations, as the ones placed by the dynamics of the social behavior; as in the case of criminality and mobility.

3.2 Relationship of cluster assignment and territorial extension

The cluster analysis aimed at the identification of the best number of clusters to describe our dataset. Consequently, we exhaustively tested the KMeans’ cluster-quantity parameter from 2 to 644 clusters — the total number of cities without considering Sao Paulo. During the test, *we were seeking for the greatest average Silhouette score (AVG) only when the Dunn index (DNN) was larger than one.*

The previous experiment suggests that the best way to split our data is into two clusters. Such configuration has an AVG of 0.59 and a DNN of 1.10 (see Figure 7). When dividing the data into two, the clusters are better balanced rather than when considering Sao Paulo — a big outlier — as part of the dataset.

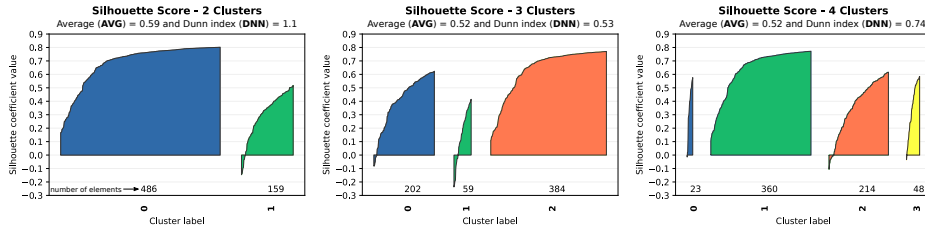


Fig. 7: Silhouette analysis of the subset of our data without Sao Paulo, in which clusters are represented as color-coded polygons. In each scenario that we have tested, the results were validated according to the Dunn index together with the Silhouette score. Although we have depicted the first three tested scenarios, which are also the best ones, the experiment considers a total of 643 scenarios.

Subsequently, we investigated the reason why the cities were better arranged into only two clusters. By analyzing indicators related to population and territory extension, we found that 61.20% of the state’s population is in the first cluster

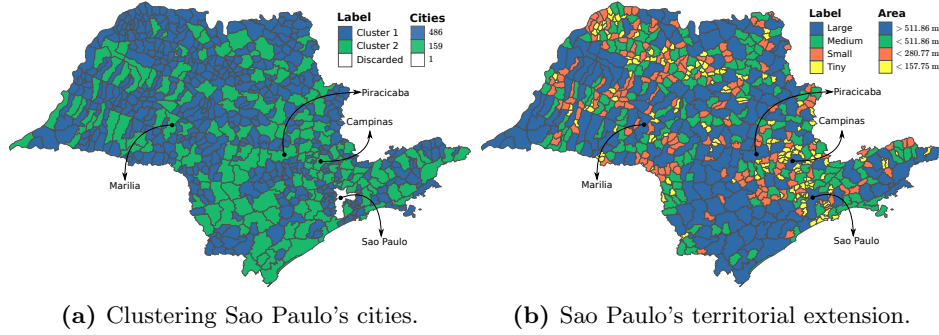


Fig. 8: Investigating cities through clustering techniques; Figure 8a shows the results of the clustering of topological features when removing the Sao Paulo city from the dataset; this layout has an average Silhouette of 0.59 and Dunn Index of 1.1. Figure 8b describe the area within the cities' administrative boundaries.

and 38.80% is in the second one (see Figure 8a), and that the first cluster is mainly populated with cities that are considered to be of tiny or small territorial extension (see Figure 8b), while the second cluster has the opposite behavior. Bearing in mind that our dataset does not imply any relationship between indicators of territorial extension and population density, we concluded that the relation that favored two clusters, as the arrangement with best values of Silhouette and Dunn index, was the territorial extension of the cities. Hence, we found evidence that there is a significant relationship between topological features, territorial extension, and demographics of the cities of Sao Paulo state.

The relationship between the cluster arrangement and territorial extension can be understood as the way cities organize within their available space. In fact, regarding the territorial extension, 30.51% of the cities from the first cluster are tiny-sized, 31.13% are small-sized, 25.78% are medium-sized, and 12.58% are large-sized; whereas, 7.59% of the ones from the second cluster are tiny-sized, 6.32% are small-sized, 22.78% are medium-sized, and 63.29% are large-sized. Therefore, cities in the first cluster can be considered smaller and heavily populated, while the ones in the second cluster are larger and less populated.

3.3 Discussions on results generalization

We have chosen to present a joint of direct findings and analytical conclusions in our results section. This was done so that one can follow the practical application of the proposed methodology in a way that can be adapted and generalized for different domains and scenarios. Our methodology can also be used in non-urban applications, such as in the characterization of the topology in any group of complex networks, however, depending on the specificities of the domain, it may be necessary to use different network metrics and features to be more effective.

Additionally, while our findings cannot be generalized for any set of cities, we believe that the proposed methodology can be used to find non-trivial properties in different urban scenarios and not only to the cities that shape the state of Sao

Paulo. Comprising a straightforward framework of analysis that can be useful to the academic community and cities' governing body, *e.g.* planners and designers.

Finally, our proposal has intricacies that can be explored in further studies: **(1)** using hierarchical clustering to reveal additional knowledge, which may also demand prior expertise about the cities (*e.g.* history and geography); and, **(2)** using more complex feature selection techniques such as fractal-dimension based methods or by applying ones related to mutual information. Notice that, this refinement might reveal other patterns of the data, but will not change the ones we discussed; and, **(3)** including non-topological features to capture different characteristics of cities, enhancing our methods capabilities and its versatility.

4 Conclusion

In this paper we proposed a three-folded method encompassing the data Acquisition, Modeling, and Computation. Furthermore, our methodology comprises the following phases: *Data Acquisition and Preparation*, *Feature Extraction and Selection*, and *Feature Vector Analysis*; culminating in the use of multidimensional projection and cluster analysis algorithms to assess feature vectors of complex-network metrics. To validate our proposal, we investigated the following hypotheses: **(A)** *the network topology is a tool-set that can reveal groups of cities with similar characteristics, potentially revealing disparities*; **(B)** *although cities may share administrative boundaries with others, they cluster with cities with no apparent geographical similarity*; and, **(C)** *there might be interesting correlations between urban and/or territorial indicators and the features extracted from the street-network topology of a given set of cities*. Such hypotheses were investigated by analyzing relations between 645 cities that constitute the Brazilian state of Sao Paulo. Our main findings confirm the hypotheses of our work, allowing us to state that, on a large scale, the topological features of the cities can indicate, or even predict, their demography and that cities group themselves by means of their territorial extension, which describes the way that cities organize within their available space. Therefore, our main contributions are: **(i)** *the description of how the network topology is capable of revealing groups of cities with similar characteristics*; **(ii)** *the correlation analysis between the demography of the cities and their features*; and, **(iii)** *the discussion of why cities cluster with other cities distant apart instead of with those that they share boundaries with*. As a future work, we will measure the similarity between cities by means of non-topological features, looking for discrepancies in the collective behavior that emerges from this same set of cities.

Acknowledgment

We are thankful to CNPq (grant 167967/2017-7), FAPESP (grants 2014/25337-0, 2016/17078-0 and 2017/08376-0) and CAPES that supported this research.

Bibliography

- [1] Boccaletti, S; Latora, V; Moreno, Y; Chavez, M; Hwang, D: Complex networks: Structure and dynamics. *Physics Reports* **424**(4-5) (feb 2006) 175–308
- [2] Masucci, AP; Stanilov, K; Batty, M: Limited Urban Growth: London’s Street Network Dynamics since the 18th Century. *PLoS ONE* **8**(8) (aug 2013)
- [3] Blumer, H: Social problems as collective behavior. *Soc Probl* **18**(3)(1971) 298–306
- [4] Anderson, TK: Kernel density estimation and k-means clustering to profile road accident hotspots. *Accident Analysis & Prevention* **41**(3) (2009) 359–364
- [5] Grauwin, S; Sobolevsky, S; Moritz, S; Gódor, I; Ratti, C: Towards a comparative science of cities: Using mobile traffic records in new york, london, and hong kong. In: *Computational approaches for urban environments*. Springer (2015) 363–387
- [6] Crucitti, P; Latora, V; Porta, S: Centrality measures in spatial networks of urban streets. *Phys Rev E: Stat Nonlinear Soft Matter Phys* **73**(3) (mar 2006)
- [7] Costa, LF; Travençolo, BAN; Viana, MP; Strano, E: On the efficiency of transportation systems in large cities. *EPL (Europhysics Letters)* **91**(1) (jul 2010)
- [8] Porta, S; Latora, V; Wang, F; Strano, E; Cardillo, A; Scellato, S; Iacoviello, V; Messori, R: Street Centrality and Densities of Retail and Services in Bologna, Italy. *Environment and Planning B: Planning and Design* **36**(3) (2009) 450–465
- [9] Strano, E; Nicosia, V; Latora, V; Porta, S; Barthélemy, M: Elementary processes governing the evolution of road networks. *Scientific Reports* **2** (mar 2012)
- [10] Spadon, G; Gimenes, G; Rodrigues-Jr, JF: Identifying Urban Inconsistencies via Street Networks. Volume 108; Elsevier BV (2017) 18 – 27 International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland.
- [11] Li, X; Parrott, L: An improved genetic algorithm for spatial optimization of multi-objective and multi-site land use allocation. *Comput Environ Urban Syst* **59**(2016)
- [12] Strano, E; Viana, M; Costa, LF; Cardillo, A; Porta, S; Latora, V: Urban Street Networks, a Comparative Analysis of Ten European Cities. *Environment and Planning B: Planning and Design* **40**(6) (dec 2013) 1071–1086
- [13] Domingues, GS; Silva, FN; Comin, CH; Costa, LF: Topological characterization of world cities. *arXiv preprint arXiv:1709.08244* (2017)
- [14] Pan, G; Qi, G; Zhang, W; Li, S; Wu, Z; Yang, LT: Trace analysis and mining for smart cities: issues, methods, and applications. *IEEE Commun Mag* **51**(6) (2013)
- [15] Scripps, J; Nussbaum, R; Tan, PN; Esfahanian, AH: Link-Based Network Mining. In: *Structural Analysis of Complex Networks*. Springer Nature (sep 2010) 403–419
- [16] Chiang, C: *Statistical Methods of Analysis*. World Scientific (2003)
- [17] Costa, LF; Rodrigues, FA; Travieso, G; Boas, PRV: Characterization of complex networks: A survey of measurements. *Adv Phys* **56**(1) (2007) 167–242
- [18] Hage, P; Harary, F: Eccentricity and centrality in networks. *Social networks*, **17**(1), (1995), 57-63.
- [19] Brath, R; and Jonker, D: *Graph Analysis and Visualization: Discovering Business Opportunity in Linked Data*. John Wiley & Sons (2015).
- [20] Spiwok, V; Oborský, P; Pazúriková, J; Křenek, A; Králová, B: Nonlinear vs. linear biasing in trp-cage folding simulations. *J Chem Phys* **142**(11) (2015)
- [21] MacQueen, J; et al: Some methods for classification and analysis of multivariate observations. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Volume 1. (1967) 281–297

- [22] Kremer, H; Kranen, P; Jansen, T; Seidl, T; Bifet, A; Holmes, G; Pfahringer, B: An effective evaluation measure for clustering on evolving data streams. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '11, New York, NY, USA, ACM (2011) 868–876
- [23] Rousseeuw, Peter J Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math, **20**(1), (1987), p. 53-65.
- [24] Dunn, Joseph C: Well-separated clusters and optimal fuzzy partitions. Journal of cybernetics **142**(1), (1974), 95-104.