Characterisation of Street Networks

Andreu Camprubí i Peiró and Andreu Giménez i Larred

Universitat Politècnica de Catalunya, Jordi Girona, 31 08034 Barcelona

Abstract. This study characterises the street networks of 43 Catalan cities using network science to reveal urban structures and diagnose their current state. The modelling as a network allows the computation of network metrics such as node degree, betweenness centrality, orientation entropy, and detour index. Street segment length distributions were modelled using six probability density functions, with the best fit for each city determined by maximum likelihood estimation and AICc. Results reveal significant variations in network connectivity, orientation order, and street length distributions across cities, reflecting urban form and scale differences.

Keywords: Network Science · Spatial Networks · Street Networks.

1 Introduction

Cities and urban systems are composed of intricate networks of streets and paths that form the backbone of human interaction and travel behaviour. By representing these street networks as graphs—where intersections and dead-ends are vertices and street segments are edges—we can analyse their structural properties to gain insights into urban design and transportation efficiency.

This study focuses on the street networks of Catalonia's diverse cities. We analyse the capital of each of the 43 *comarques* (regions) of Catalonia[1] to obtain a varied set of street networks. Using metrics such as street entropy, we classify and compare these networks to uncover differences in urban structure.

Our research aims to characterise street networks and link these characterisations to urban planning and design concepts. We compare the Catalan cities using established analytical techniques and benchmarks (see [2], [3], [4], and [6]). Additionally, we model the distribution of street lengths for the four provincial capitals.

This analysis applies network analysis techniques to the unique context of Catalan street networks, contributing empirical findings that may inform urban science and planning.

[1] There are 42 *comarques* and one unique territorial entity, Val d'Aran. Still, for simplicity, we treat them all as *comarques*.

2 Methods

2.1 Data

To extract all the graphs of the different cities of Catalonia, we used the Python library 'OSMnx' (see [5]), which is a free, open-source, Python-based toolkit to automatically download spatial data via the 'Nominatim' and 'Overpass' APIs (including municipal boundaries and streets) from OpenStreetMap and construct graph-theoretic objects for network analysis. We can get a graph querying the city's name using the function `graph_from_place()`. We can perform all the calculations and analyses explained in the following sections from this graph.

2.2 Measures of Street Network Structure

Metrics A total of 10 graph metrics were chosen to be analysed. Depending on the metric we will use a directed or an undirected version of the graph. Directed graphs most-faithfully represent constraints on flows (such as vehicular traffic on a one-way street), undirected graphs better model urban form by corresponding 1:1 with street segments.

- **N**: Number of Nodes
$$N = |V|$$
where V is the set of nodes in the graph.
- **D**: Average Node Degree
$$D = \frac{1}{N} \sum_{v \in V} \deg(v)$$
where $\deg(v)$ is the degree of node v.
- **L**: Total Length (m)
$$L = \sum_{e \in E} \text{length}(e)$$
where E is the set of edges in the graph and $\text{length}(e)$ is the length of edge e.
- **S**: Average Street Length (m)
$$S = \frac{L}{|E|}$$
where L is the total length and $|E|$ is the number of edges.
- $\mathbf{B_N}$: Average Betweenness Centrality (Nodes)
$$B_N = \frac{1}{N} \sum_{v \in V} \text{betweenness}(v)$$
where $\text{betweenness}(v)$ is the betweenness centrality of node v.

- $\mathbf{B_E}$: Average Betweenness Centrality (Edges)

$$B_E = \frac{1}{|E|} \sum_{e \in E} \text{betweenness}(e)$$

where betweenness(e) is the betweenness centrality of edge e.

- **OE**: Orientation Entropy

$$\text{OE} = -\sum_i p_i \log(p_i)$$

where p_i is the probability of orientation i, and we have a total of 36 possible orientations (one every 10º).

- ϕ: Normalized Measure of Orientation-Order

$$\phi = 1 - \left(\frac{\text{OE} - H_g}{H_{\max} - H_g} \right)^2$$

where H_g is the entropy of a grid and $H_{\max}$ is the maximum entropy.

$$H_{\max} = ln(36)$$

$$H_g = ln(4)$$

- $\mathbf{P_D}$: Proportion of Dead-ends

$$P_D = \frac{\text{Number of dead-ends}}{N}$$

where a dead end are all the nodes with degree 1

- $\mathbf{P_4}$: Proportion of k=4 Intersections

$$P_4 = \frac{\text{Number of 4-way intersections}}{N}$$

where a 4-way intersection are all the nodes with degree 4

- **I**: Detour Index (Average Circuity)

$$I = \frac{L_{\text{net}}}{L_{\text{gc}}}$$

where L_{net} represents the sum of all edge lengths in the graph and L_{gc} represents the sum of all great-circle distances between all pairs of adjacent nodes. Thus, it represents how much more circuitous a city's street network is than it would be if all its edges were straight-line paths between nodes (see [2]).

Correlation Between Metrics To calculate the correlation between variables, we used the function 'pearsonr' from the 'Scipy' library, returning the correlation value of each pair of metrics and the Pearson correlation test p-value. The correlation value goes from -1 to 1, and the closer a value is to 0, the less correlated are the two metrics. The p-value is calculated to determine if the observed correlation is statistically significant.

- p-value < 0.05: Reject the null hypothesis, indicating a statistically significant linear relationship.
- p-value ≥ 0.05: Fail to reject the null hypothesis, indicating no statistically significant linear relationship.

Polar histograms In our study, the bearing of edge $\mathrm{e_{uv}}$ equals the compass heading from u to v and its reciprocal (e.g., if the bearing from u to v is 90° then we additionally add a bearing of 270° since the one-dimensional street centreline points in both directions).

Once we have calculated all of the bearings (and their reciprocals) for all the edges in a city, we divide them into 36 equal-sized bins (i.e., each bin represents 10°). To avoid extreme bin-edge effects around common values like 0° and 90°, we shift each bin by 5° so that these values sit at the centres of their bins rather than at their edges. This allows similar common bearings, such as 359.9° and 0.1°, to fall in the same bin (see [4]).

Provinces and *Vegueries* division to create density functions Catalonia is divided into geographic zones with several types of divisions. The administrative divisions are the four provinces (Tarragona, Barcelona, Lleida and Girona), but they encompass a very varied territory, and their cities have different characteristics.

We used another type of division to create more homogeneous divisions: *Vegueries*. According to the Institut Cartogràfic i Geològic de Catalunya (ICGC), the last update of the *Vegueries* map (see [9]) consists of 9 different divisions: Girona, Barcelona, Lleida, Camp de Tarragona, Catalunya Central, Penedès, Terres de l'Ebre, Alt Pirineu and Aran. As Aran had been considered as part of Alt Pirineu but due to its unique territorial entity status it became a different one and it has a similar orography as Alt Pirineu, we will join them as one to perform the study.

We then used the 'kdeplot' function from the library 'Seaborn' to create the kernel density plots.

Clustering (see [8])

Finally, to systematically interpret city similarities and differences, we cluster the study sites in two different ways. First, using all the metrics calculated and then using only a four-dimensional feature space of the key indicators that aren't affected by the size of the city and also are key indicators: average node degree, orientation indicator, average street segment length (m) and average circuitry.

We first standardise the features for appropriate scaling, then perform hierarchical agglomerative clustering using the Ward linkage method. We then used 4 different methods to determine the ideal number of clusters to use:

- **Dendrogram observation**: By observing the dendrogram, we can try to figure a horizontal line in a place where there is some distance between division and division. The number of lines intersecting with the imaginary drawn horizontal line will be the optimal number of clusters.
- **Elbow method**: It works by finding the within-cluster sum of square (WCSS) on the y-axis corresponding to the different values of K on the x-axis. If the graphic shows an elbow at some point, it means that the value of the x-axis of that point (k) is the optimal.
- **Silhouette method**: this method measures how similar an object is to its cluster compared to others. It ranges from -1 to 1, where a value close to 1 indicates that the object is well-matched to its cluster and poorly matched to neighbouring clusters, while a value close to -1 indicates the opposite.
- **The Davies-Bouldin Index**: this method evaluates the average similarity ratio of each cluster with the cluster that is most similar to it. It ranges from 0 to infinity, where a lower value indicates better clustering performance, meaning clusters are compact and well-separated.

We finally used t-SNE (see [11]), a dimensionality reduction technique for visualising high-dimensional data by mapping it to a lower-dimensional space (typically 2D or 3D). This technique emphasises preserving the local structure of the data, making it useful for visualising clusters.

2.3 Street Segment Length Analysis

To obtain the street segment lengths for each provincial capital, we employ a Python script utilising the OSMnx library. The corresponding street network graph is loaded from each provincial capital's previously generated GraphML file. The directed graph is converted into an undirected graph so that its edges correspond 1:1 with street segments. Next, it is projected onto a suitable coordinate system for more precise length measurements. The undirected, projected graph is transformed into a GeoDataFrame of its edges using OSMnx utilities. From this GeoDataFrame, the length attribute of each edge, representing the length of each street segment in meters, is extracted and saved to a CSV file.

We consider an ensemble of six probability density functions on which to perform model selection. Our choice is based on the street segment length distribution proposed for London in [13]. Their proposal is:

$$f(l) \propto \exp\left[-\frac{145}{l} - \frac{l}{2000}\right] l^{-3.36}$$

Our ensemble consists of the following density functions:

- **Model 1:**
$$f_1(l) = C\, l^{-\gamma}$$
where $\gamma > 0$ so that the power-law term $l^{-\gamma}$ is monotonically decreasing with respect to l. This behaviour aligns with what we observe in the data: as the street length l increases, the density $f_1(l)$ decreases. This will be our null model since many empirical studies of real-world networks consistently observe power-law behaviour in various properties.
- **Model 2:**
$$f_2(l) = C\, l^{-\gamma} e^{-\beta l}$$
where $\beta, \gamma > 0$. The combination of a power-law $l^{-\gamma}$ and an exponential decay $e^{-\beta l}$ ensures a rapid decrease of the density as l increases. The exponential component dominates for large l, accelerating the decay. This rapid decline at larger lengths would be consistent with observed data where very long street segments become increasingly rare.
- **Model 3:**
$$f_3(l) = C\, e^{-\alpha/l}\, l^{-\gamma}$$
where $\alpha, \gamma > 0$. The term $e^{-\alpha/l}$ increases from 0 to 1 as l increases, while $l^{-\gamma}$ decreases. The product yields a decreasing density function tempered at small l by the factor $e^{-\alpha/l}$. This tempered decrease would reflect observed data where the drop-off is less abrupt for very short segments.
- **Model 4:**
$$f_4(l) = C\, e^{-\alpha/l-\beta l}$$
where $\alpha, \beta > 0$. The combined exponential terms $e^{-\alpha/l}$ and $e^{-\beta l}$ ensure that the density decreases as l becomes very small and large. This model typically yields a unimodal distribution with a peak at an intermediate l. Such a unimodal shape would align with data that show a most common street length, with fewer very short or very long streets.
- **Model 5:**
$$f_5(l) = C\, e^{-\beta l}$$
where $\beta > 0$. The pure exponential decay $e^{-\beta l}$ results in a strictly decreasing density as l increases. This behaviour would mirror observed data where longer street segments are progressively less common without the additional influence of other factors.
- **Model 6:**
$$f_6(l) = C\, e^{-\alpha/l-\beta l}\, l^{-\gamma}$$
where $\alpha, \beta, \gamma > 0$. This model combines power-law and exponential behaviours. The term $l^{-\gamma}$ decays as l increases, while $e^{-\alpha/l}$ rises from 0 to 1 and $e^{-\beta l}$ decays. The overall density decreases for large l, with complex behaviour at small and intermediate values due to the interplay of these components. This complexity would match data where street length distributions exhibit a non-trivial mix of frequent short segments, a peak at intermediate lengths, and a steep decline for longer segments.

To estimate the model's parameters and select the best model, we follow the approach described in [12]. For each model, dedicated functions in our code compute the normalisation constant and the negative log-likelihood. Since integration routines may fail or return non-finite values for the normalisation constant computation, these cases are caught using `try` or `tryCatch` constructs. When an integration error occurs, or the result is non-finite, the functions return a large penalty value (e.g., 1×10^{12}). This approach discourages the selection of parameter sets that lead to unstable or invalid computations. Moreover, each model specification includes lower bounds for parameters (the ones specified above when we defined the models) and a grid of starting values. This decision reflects a deliberate strategy to explore a variety of initial conditions for the optimisation, increasing the likelihood of finding a global optimum.

The parameter estimation uses maximum likelihood estimation via the `mle` function from the `stats4` package. The routine captures errors during optimisation, filtering out failed attempts. Among successful estimations, the one with the smallest negative log-likelihood is selected as the best result for each model. Next, AIC values are computed for each model using the sample-size-corrected formula:

$$AIC_c = -2\mathcal{L} + 2K\frac{N}{N-K-1}$$

where $\mathcal{L}$ denotes the model's log-likelihood, K represents the number of parameters, and N is the sample size. For every city, the best model is selected based on this metric. Finally, plots of the best model superposed to the real data are generated for every city.

3 Results

3.1 Measures of Street Network Structure

After getting the graphs from all the *comarques* capitals, we calculated all the metrics from every city and presented them in Table 1 and Table 2.

Finally, Table 3 shows the orientation entropy values normalised in two different ways. Normalisation 1 corresponds to min-max scaling, and normalisation 2 to the square of the min-max scaling (1- ϕ).

Betweenness Centrality After calculating the metrics, we plotted the edge betweenness centrality of Vic to understand which edges were the most likely to be busy and central to the city (Figure 1).

Polar histograms We then calculated the street angle orientation creating polar histograms. In Figure 2 all the polar histograms of all the cities are displayed in increasing order of entropy. You can see that the cities on top have more defined polarity and the ones on the bottom are more distributed among all the directions.

City	N	D	L	S	B_N
Barcelona	8885	3.71	1392727.88	99.84	0.005
Girona	2104	4.17	268903.54	87.88	0.0167
Lleida	3240	4.18	626246.07	128.75	0.0122
Tarragona	2768	3.89	431954.88	107.75	0.0129
Mataró	1718	3.92	220399.81	82.21	0.0166
Sabadell	3144	4.0	398494.7	79.37	0.0111
Terrassa	3654	3.84	501453.83	87.45	0.0094
Manresa	1524	3.85	221078.4	99.32	0.0164
Vic	1295	4.01	168060.07	86.27	0.0219
Igualada	982	4.23	113025.93	73.63	0.0235
Vilafranca del Penedès	787	4.09	106141.63	88.9	0.0257
Vilanova i la Geltrú	1614	3.73	211632.59	86.42	0.0185
el Vendrell	2153	4.47	275759.86	86.47	0.0163
Reus	2791	3.66	347803.8	82.79	0.0129
Tortosa	1525	4.46	287357.63	131.15	0.0252
Amposta	1232	4.17	265724.09	138.69	0.019
Gandesa	290	4.97	47792.26	113.52	0.0415
Falset	204	4.75	32458.73	113.89	0.0489
Montblanc	509	4.53	77613.15	106.47	0.0376
Valls	1231	4.09	165571.94	96.43	0.0217
Balaguer	672	4.28	111340.8	113.73	0.0284
Cervera	654	4.62	121495.92	128.3	0.0295
Solsona	585	4.36	67920.44	81.34	0.0319
la Seu d'Urgell	483	4.14	55277.18	78.52	0.0379
Sort	389	4.44	63737.27	127.73	0.0728
Tremp	226	4.15	28660.14	90.13	0.0588
el Pont de Suert	189	4.95	74928.3	312.2	0.0725
Mollerussa	571	4.11	57859.78	67.44	0.0333
les Borges Blanques	456	4.5	78329.75	117.44	0.0319
Tàrrega	943	4.26	126662.5	91.98	0.0278
Olot	1395	4.53	168018.53	81.76	0.0206
Ripoll	453	4.37	60080.94	95.22	0.0479
Puigcerdà	426	4.42	51818.3	86.22	0.0395
Banyoles	888	4.11	102103.47	76.14	0.0261
la Bisbal d'Empordà	299	5.04	39152.43	91.05	0.0412
Santa Coloma de Farners	557	4.37	104498.53	128.06	0.0448
Figueres	1213	3.92	134082.07	74.0	0.0234
Vielha e Mijaran	388	4.13	73830.31	146.78	0.0642
Berga	526	4.05	81881.67	110.8	0.0365
Sant Feliu de Llobregat	445	4.04	55273.69	82.13	0.0368
Granollers	1012	4.04	136761.01	90.15	0.0232
Móra d'Ebre	410	4.7	68224.71	117.43	0.0352
Prats de Lluçanès	153	5.12	20213.62	93.15	0.0619
Moià	326	5.16	57847.44	128.26	0.0403

Table 1. N: Number of Nodes, **D**: Average Node Degree, **L**: Total Length (m), **S**: Average Street Length (m), $\mathbf{B_N}$: Average Betweenness Centrality (Nodes)

City	B_E	ϕ	P_D	P_4	I
Barcelona	0.0028	0.1072	0.0724	0.293	1.05
Girona	0.0082	0.0351	0.0984	0.1516	1.07
Lleida	0.006	0.031	0.0664	0.1519	1.05
Tarragona	0.0068	0.0165	0.0968	0.1077	1.08
Mataró	0.0087	0.0508	0.0477	0.2159	1.06
Sabadell	0.0057	0.1538	0.0407	0.2888	1.04
Terrassa	0.005	0.1202	0.0575	0.2635	1.04
Manresa	0.0088	0.0062	0.0978	0.1365	1.08
Vic	0.0113	0.0507	0.0556	0.1521	1.04
Igualada	0.0116	0.1378	0.0346	0.2057	1.02
Vilafranca del Penedès	0.0131	0.0913	0.0457	0.1804	1.04
Vilanova i la Geltrú	0.0102	0.0726	0.0539	0.1679	1.04
el Vendrell	0.0075	0.0126	0.092	0.1737	1.04
Reus	0.0072	0.0609	0.0473	0.125	1.04
Tortosa	0.0116	0.0389	0.1272	0.1593	1.12
Amposta	0.0095	0.1207	0.0641	0.2589	1.04
Gandesa	0.0179	0.1434	0.0966	0.1517	1.11
Falset	0.0224	0.0632	0.1373	0.1078	1.12
Montblanc	0.0174	0.0888	0.1198	0.1336	1.09
Valls	0.011	0.0418	0.1413	0.1007	1.08
Balaguer	0.014	0.0879	0.0997	0.1086	1.06
Cervera	0.0134	0.0671	0.0963	0.1162	1.07
Solsona	0.0154	0.0231	0.1128	0.0991	1.09
la Seu d'Urgell	0.0192	0.1471	0.1118	0.1615	1.07
Sort	0.0338	0.0322	0.2314	0.0591	1.33
Tremp	0.0302	0.0751	0.1372	0.1327	1.16
el Pont de Suert	0.0311	0.0735	0.2487	0.0582	1.37
Mollerussa	0.017	0.1546	0.0595	0.1611	1.02
les Borges Blanques	0.0151	0.1037	0.0855	0.114	1.05
Tàrrega	0.0135	0.0787	0.1198	0.1527	1.06
Olot	0.0094	0.0508	0.1097	0.1821	1.06
Ripoll	0.0228	0.0389	0.1501	0.0927	1.15
Puigcerdà	0.0189	0.0282	0.1315	0.1174	1.06
Banyoles	0.0132	0.1087	0.0833	0.214	1.06
la Bisbal d'Empordà	0.0175	0.1118	0.1104	0.1572	1.05
Santa Coloma de Farners	0.0212	0.0943	0.1167	0.1741	1.11
Figueres	0.0123	0.1066	0.0635	0.1509	1.04
Vielha e Mijaran	0.0322	0.0506	0.2139	0.0258	1.38
Berga	0.0189	0.0579	0.1255	0.1046	1.18
Sant Feliu de Llobregat	0.0192	0.2034	0.0494	0.2022	1.04
Granollers	0.0119	0.2737	0.0583	0.1887	1.04
Móra d'Ebre	0.0159	0.0727	0.1439	0.1488	1.07
Prats de Lluçanès	0.0264	0.1721	0.1307	0.1307	1.12
Moià	0.0167	0.0495	0.1718	0.1196	1.13

Table 2. $\mathbf{B_E}$: Average Betweenness Centrality (Edges), ϕ: Normalized Measure of Orientation-Order,$\mathbf{P_D}$: Proportion of Dead-ends, $\mathbf{P_4}$: Proportion of k=4 Intersections, **I**: Detour Index

Capital	OE (normalisation 1)	OE (normalisation 2)
Barcelona	0.944867	0.892773
Girona	0.982301	0.964915
Lleida	0.984366	0.968977
Tarragona	0.991736	0.983541
Mataró	0.974279	0.949220
Sabadell	0.919909	0.846232
Terrassa	0.937970	0.879788
Manresa	0.996912	0.993833
Vic	0.974322	0.949304
Igualada	0.928531	0.862170
Vilafranca del Penedès	0.953239	0.908665
Vilanova i la Geltrú	0.963038	0.927441
el Vendrell	0.993702	0.987444
Reus	0.969091	0.939137
Tortosa	0.980339	0.961064
Amposta	0.937733	0.879343
Gandesa	0.925540	0.856624
Falset	0.967872	0.936777
Montblanc	0.954587	0.911237
Valls	0.978861	0.958169
Balaguer	0.955061	0.912142
Cervera	0.965873	0.932911
Solsona	0.988387	0.976909
la Seu d'Urgell	0.923541	0.852927
Sort	0.983744	0.967753
Tremp	0.961726	0.924918
el Pont de Suert	0.962548	0.926498
Mollerussa	0.919468	0.845421
les Borges Blanques	0.946756	0.896347
Tàrrega	0.959842	0.921296
Olot	0.974270	0.949201
Ripoll	0.980337	0.961061
Puigcerdà	0.985821	0.971843
Banyoles	0.944109	0.891342
la Bisbal d'Empordà	0.942469	0.888248
Santa Coloma de Farners	0.951685	0.905704
Figueres	0.945176	0.893359
Vielha e Mijaran	0.974364	0.949384
Berga	0.970617	0.942097
Sant Feliu de Llobregat	0.892511	0.796576
Granollers	0.852245	0.726322
Móra d'Ebre	0.962964	0.927300
Prats de Lluçanès	0.909916	0.827947
Moià	0.974948	0.950523

Table 3. Comparison between two methods of calculating the normalized value of orientation entropy (OE) for all the cities.

Fig. 1. Edge betweenness of Vic represented as the street thickness.

In Figure 3 we show the most entropic city (Manresa) and its polar histogram. The map helps us understand how the polar histogram works. It can be further seen in Figure 4 Where we show the least entropic city (Granollers). Looking at the street map we can clearly see that is a very checkered city.

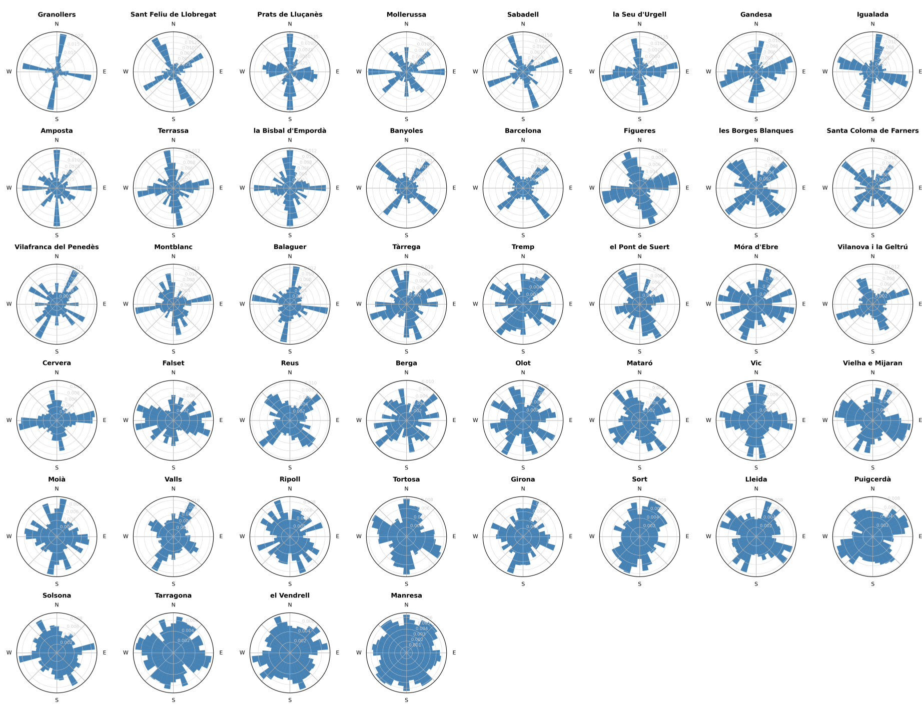

Fig. 2. Polar histograms of all the cities ordered by increasing entropy

Correlations between metrics The next step was to calculate the correlations between the different metrics. In Figure 5 the resultss are shown. After calculating the correlation matrix we also calculated the p-values of the Pearson Correlation Test. The result of the p-values are shown in Figure 6

Aggregations to create density functions We performed two different types of aggregations: one by province and another by *vegueria*. The meaning and justification of the elections are explained in the Methods section.

We choose two metrics to show the distribution of both aggregations. In Figures 7 and 8 we can see the Betweenness Centrality distribution of the provinces and *vegueries* respectively. In Figures 9 and 10 the distributions of Normalized Measure of Orientation-Order of province and *vegueries* respectively.

Fig. 3. Polar histogram of Manresa beside its street map, the one with most entropy.

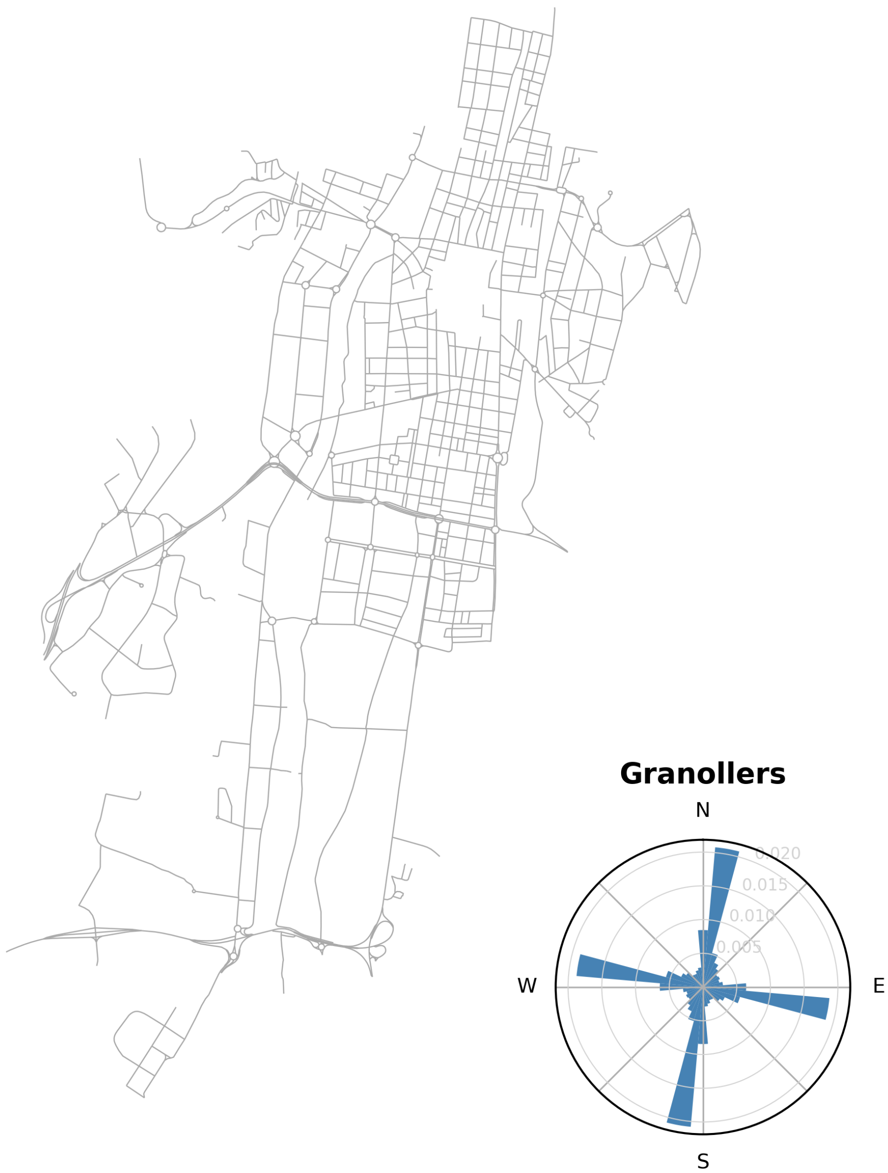

Fig. 4. Polar histogram of Granollers beside its street map, the one with less entropy.

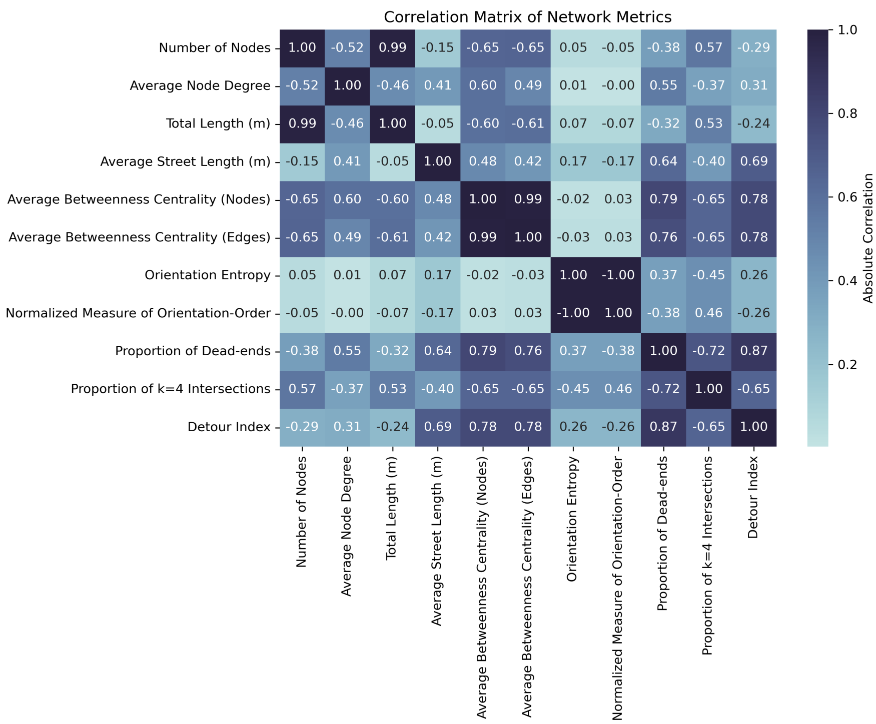

Fig. 5. Correlation matrix of the metrics.

Fig. 6. P-Values of the Pearson Correlation test of all the metrics.

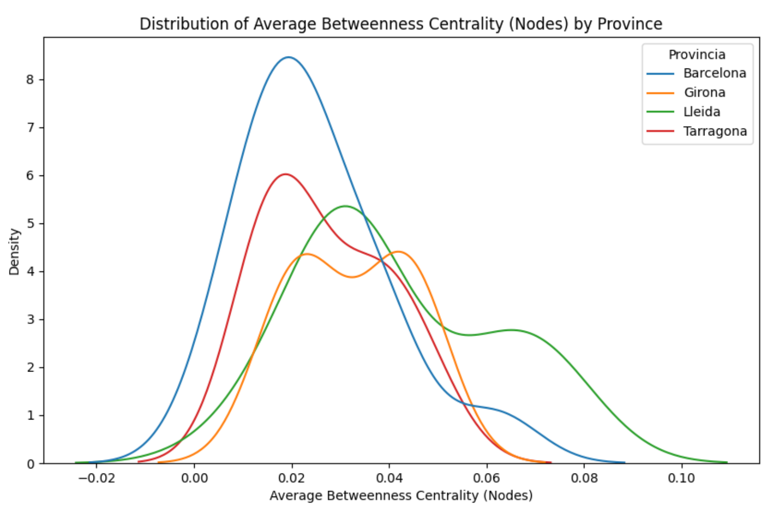

Fig. 7. Average Betweenness Centrality distribution grouped by province.

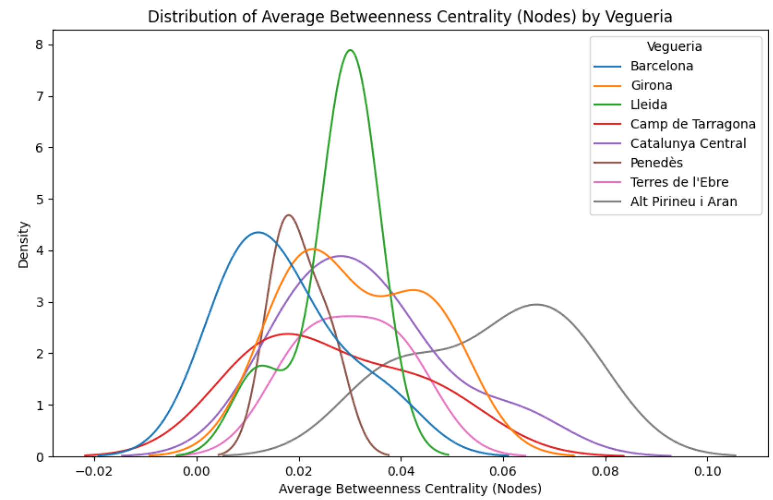

Fig. 8. Average Betweenness Centrality distribution grouped by *vegueria*.

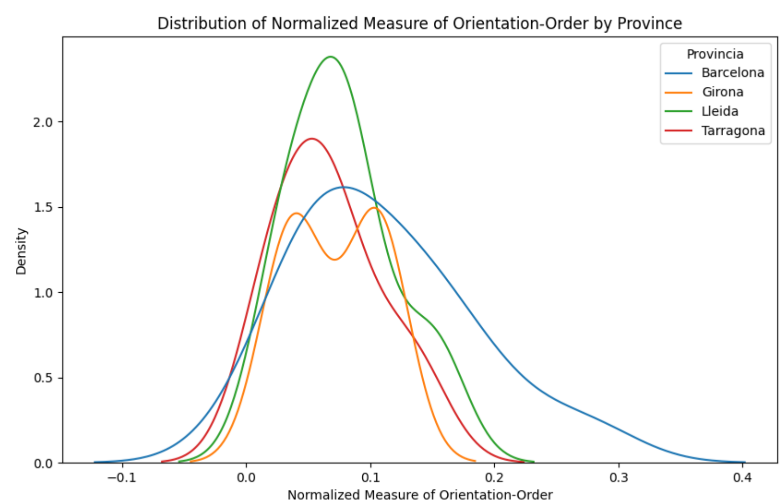

Fig. 9. Normalized measure of orientation order distribution grouped by province.

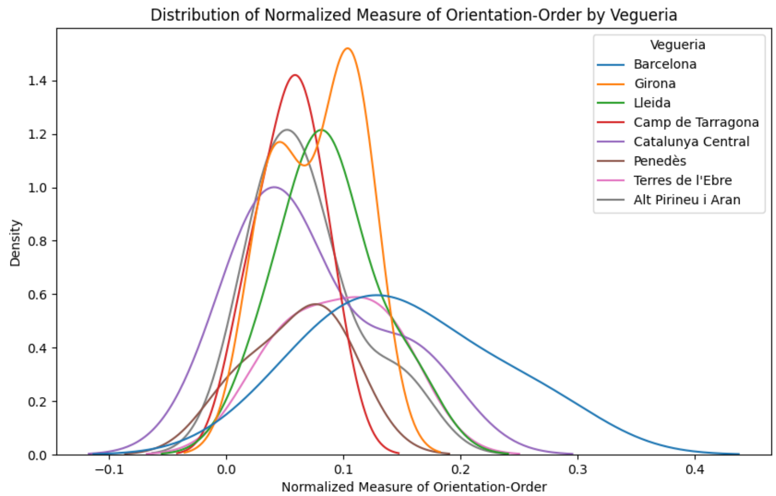

Fig. 10. Normalized measure of orientation order distribution grouped by *vegueria*.

Mean and median values of the metrics In tables 4, 5, 6, 7, 8, 9, 10 and 11 all the results of the grouped metrics can be found.

Vegueria	**N**	**D**	**L**	**S**	$\mathbf{B_N}$
Alt Pirineu i Aran	388.50	4.28	59507.22	108.93	0.06
Barcelona	2431.00	3.96	309447.26	84.83	0.01
Camp de Tarragona	1231.00	4.09	165571.94	106.47	0.02
Catalunya Central	585.00	4.23	81881.67	93.15	0.03
Girona	888.00	4.37	104498.53	87.88	0.03
Lleida	663.00	4.27	116418.36	115.58	0.03
Penedès	1614.00	4.09	211632.59	86.47	0.02
Terres de l'Ebre	821.00	4.58	166974.40	124.29	0.03

Table 4. Metrics by Vegueria (Median) - Part 1

Vegueria	$\mathbf{B_E}$	**OE**	ϕ	$\mathbf{P_D}$	$\mathbf{P_4}$	**I**
Alt Pirineu i Aran	0.03	3.51	0.06	0.18	0.09	1.24
Barcelona	0.01	3.43	0.14	0.05	0.24	1.04
Camp de Tarragona	0.01	3.52	0.06	0.12	0.11	1.08
Catalunya Central	0.02	3.53	0.05	0.11	0.13	1.09
Girona	0.01	3.48	0.09	0.11	0.16	1.06
Lleida	0.01	3.49	0.08	0.09	0.13	1.05
Penedès	0.01	3.50	0.07	0.05	0.17	1.04
Terres de l'Ebre	0.01	3.47	0.10	0.11	0.16	1.09

Table 5. Metrics by Vegueria (Median) - Part 2

Vegueria	**N**	**D**	**L**	**S**	$\mathbf{B_N}$
Alt Pirineu i Aran	350.17	4.37	58041.92	140.26	0.06
Barcelona	3143.00	3.93	450851.82	86.86	0.02
Camp de Tarragona	1500.60	4.18	211080.50	101.46	0.03
Catalunya Central	770.14	4.40	104289.65	96.11	0.03
Girona	987.00	4.36	125262.79	90.59	0.03
Lleida	1089.33	4.32	186989.14	107.94	0.03
Penedès	1518.00	4.10	197844.69	87.26	0.02
Terres de l'Ebre	864.25	4.58	167274.67	125.20	0.03

Table 6. Metrics by Vegueria (Mean) - Part 1

Vegueria	$\mathbf{B_E}$	**OE**	ϕ	$\mathbf{P_D}$	$\mathbf{P_4}$	**I**
Alt Pirineu i Aran	0.03	3.51	0.07	0.18	0.09	1.23
Barcelona	0.01	3.41	0.15	0.05	0.24	1.05
Camp de Tarragona	0.01	3.52	0.05	0.11	0.11	1.08
Catalunya Central	0.02	3.50	0.07	0.10	0.14	1.09
Girona	0.01	3.50	0.08	0.10	0.16	1.08
Lleida	0.01	3.49	0.09	0.09	0.13	1.05
Penedès	0.01	3.52	0.06	0.06	0.17	1.04
Terres de l'Ebre	0.01	3.48	0.09	0.11	0.18	1.08

Table 7. Metrics by Vegueria (Mean) - Part 2

Provincia	**N**	**D**	**L**	**S**	$\mathbf{B_N}$
Barcelona	1153.50	4.02	152410.54	88.17	0.02
Girona	722.50	4.37	103301.00	87.05	0.03
Lleida	527.00	4.27	74379.30	115.58	0.03
Tarragona	1231.50	4.47	215648.01	110.63	0.02

Table 8. Metrics by Province (Median) - Part 1

Provincia	$\mathbf{B_E}$	**OE**	ϕ	$\mathbf{P_D}$	$\mathbf{P_4}$	**I**
Barcelona	0.01	3.47	0.10	0.06	0.18	1.04
Girona	0.02	3.50	0.07	0.11	0.15	1.06
Lleida	0.02	3.50	0.07	0.11	0.12	1.07
Tarragona	0.01	3.51	0.06	0.11	0.14	1.08

Table 9. Metrics by Province (Median) - Part 2

Provincia	**N**	**D**	**L**	**S**	$\mathbf{B_N}$
Barcelona	1861.79	4.13	263213.73	91.99	0.02
Girona	916.88	4.37	116082.23	90.04	0.03
Lleida	733.00	4.34	123857.37	123.69	0.04
Tarragona	1311.30	4.37	200026.10	109.46	0.03

Table 10. Metrics by Province (Mean) - Part 1

Provincia	$\mathbf{B_E}$	**OE**	ϕ	$\mathbf{P_D}$	$\mathbf{P_4}$	**I**
Barcelona	0.01	3.46	0.11	0.07	0.19	1.07
Girona	0.02	3.50	0.07	0.11	0.15	1.08
Lleida	0.02	3.50	0.08	0.13	0.11	1.14
Tarragona	0.01	3.51	0.07	0.11	0.15	1.08

Table 11. Metrics by Province (Mean) - Part 2

Clustering The next analysis we performed was Clustering. To do so we performed hierarchical agglomerative clustering using the Ward linkage method. In figure 11 we can see the resulting dendrogram when making the clustering with all the metrics. In figure 12 we used the elbow method to choose the number of clusters. Finally we performed T-SNE to plot the clusters in a two-dimensional plane (Figure 13. We performed the same again but with a selection of variables, the dendrogram is shown in Figure 14, then we used the Silhouette method to choose the number of clusters (Figure 15). Finally the resulting T-SNE with the selected variables is shown in Figure 16.

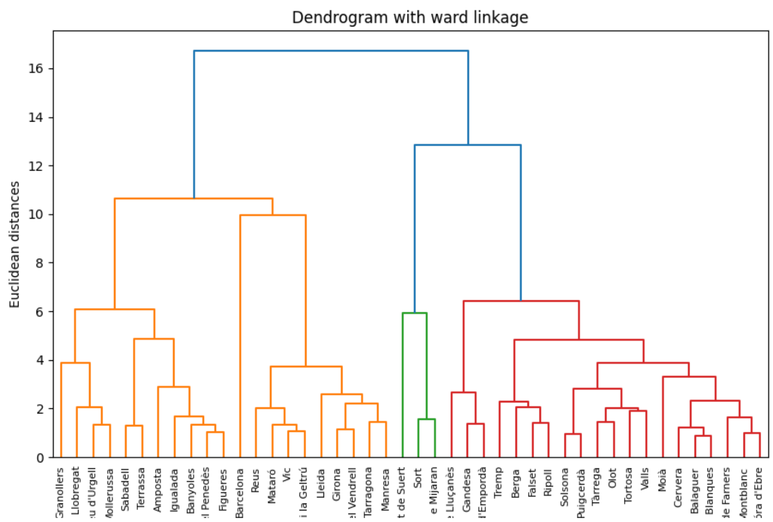

Fig. 11. Dendrogram of the hierarchical agglomerative clustering using Ward linkage method with all the metrics.

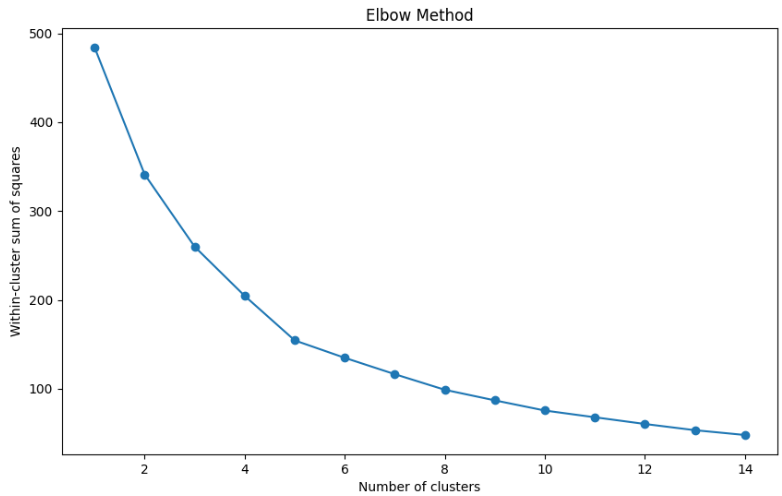

Fig. 12. Elbow method to choose the number of clusters for hierarchical agglomerative clustering using Ward linkage method with all the metrics showing that a good option is 5 clusters.

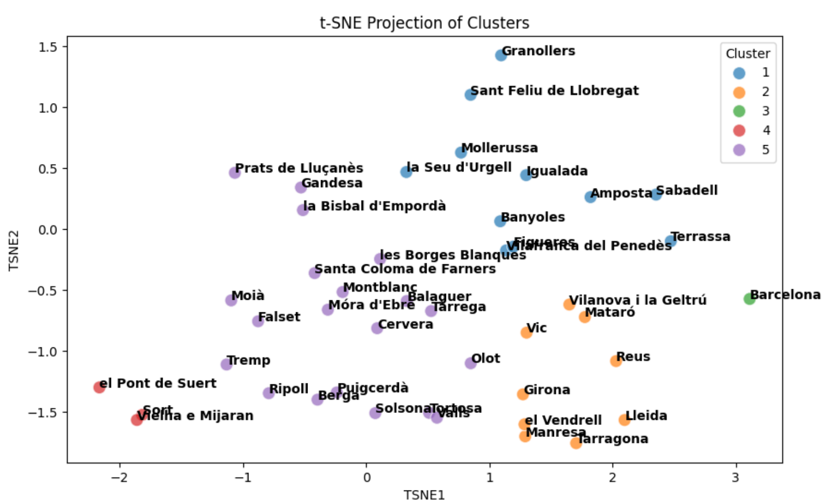

Fig. 13. T-SNE to show the result of a 5 clusters hierarchical agglomerative clustering using Ward linkage method with all the metrics.

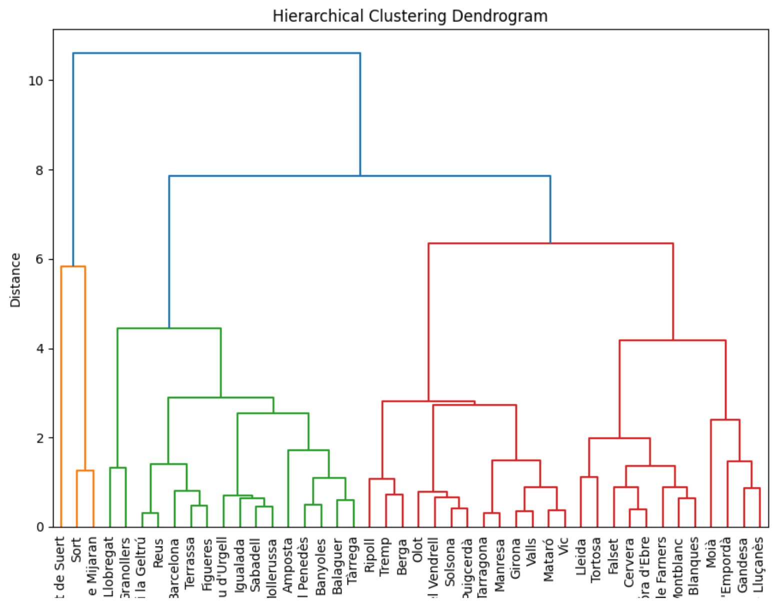

Fig. 14. Dendrogram of the hierarchical agglomerative clustering using Ward linkage method with metrics average node degree, orientation indicator, median street segment length and average circuitry.

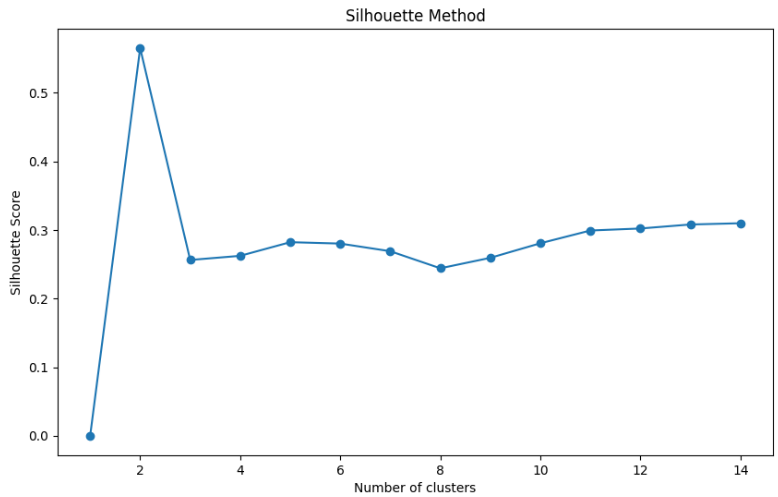

Fig. 15. Silhouette method to choose the number of clusters for hierarchical agglomerative clustering using Ward linkage method with the metrics average node degree, orientation indicator, median street segment length and average circuitry showing that a good option is 5 clusters.

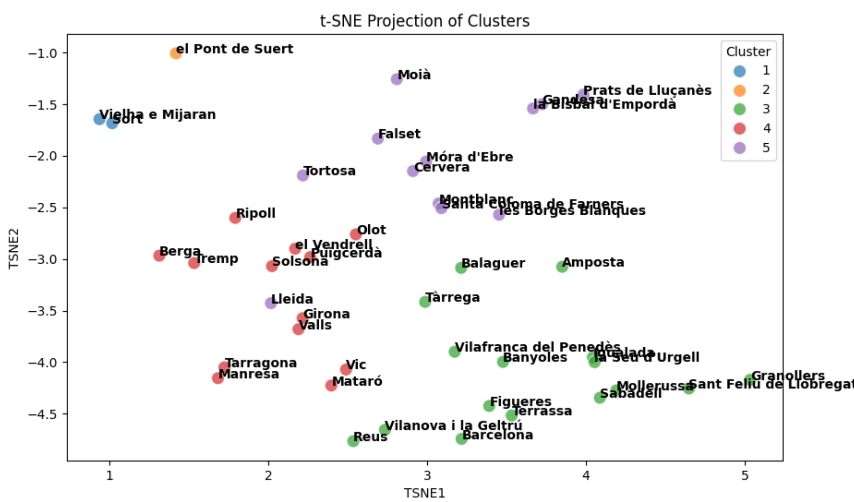

Fig. 16. T-SNE to show the result of a 5 clusters hierarchical agglomerative clustering using Ward linkage method with the metrics average node degree, orientation indicator, median street segment length and average circuitry.

City	Model	α	β	γ
Barcelona	1	-	-	1.055
Barcelona	2	-	0.01005	0
Barcelona	3	42.89	-	2.08
Barcelona	4	5.776	0.01149	-
Barcelona	5	-	0.01002	-
Barcelona	6	5.751	0.01148	0
Girona	1	-	-	1.056
Girona	2	-	0.01141	0
Girona	3	33.32	-	2.039
Girona	4	3.558	0.01266	-
Girona	5	-	0.01139	-
Girona	6	7.29	0.01006	0.3452
Tarragona	1	-	-	1.056
Tarragona	2	-	0.00766	0.1737
Tarragona	3	22.17	-	1.846
Tarragona	4	0.7229	0.009571	-
Tarragona	5	-	0.009284	-
Tarragona	6	10.43	0.003323	1.037
Lleida	1	-	-	1.056
Lleida	2	-	0.005759	0.2658
Lleida	3	28.02	-	1.927
Lleida	4	0.7367	0.007997	-
Lleida	5	-	0.007809	-
Lleida	6	10.73	0.003003	0.9781

Table 12. Summary of the most likely parameters.

City	Model	Δ
Barcelona	1	70879.64
Barcelona	2	1162.29
Barcelona	3	4925.60
Barcelona	4	0.00
Barcelona	5	1160.14
Barcelona	6	1.96
Girona	1	15066.64
Girona	2	220.32
Girona	3	817.25
Girona	4	31.99
Girona	5	218.30
Girona	6	0.00
Tarragona	1	18169.93
Tarragona	2	741.37
Tarragona	3	603.09
Tarragona	4	798.76
Tarragona	5	844.91
Tarragona	6	0.00
Lleida	1	21940.09
Lleida	2	1531.72
Lleida	3	0.00
Lleida	4	1834.12
Lleida	5	1881.80
Lleida	6	3.59

Table 13. The AIC difference (Δ) of a model on a given city.

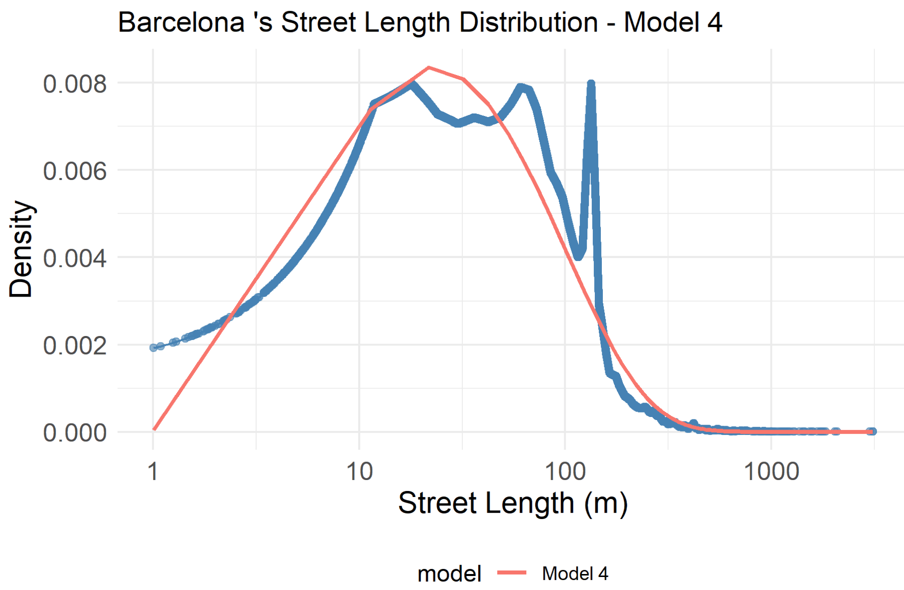

Fig. 17. Street length distribution in Barcelona (blue dots) with the best model fit (red line).

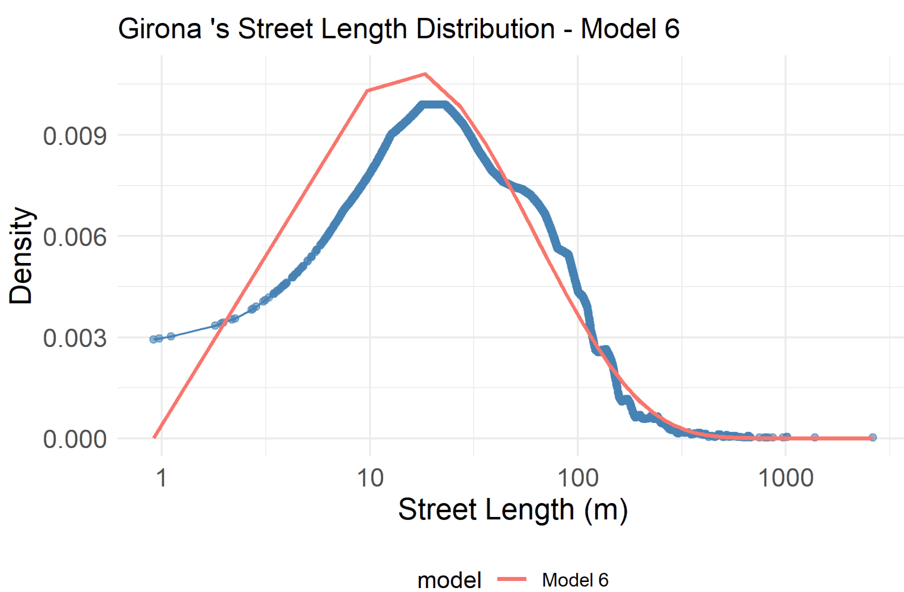

Fig. 18. Street length distribution in Girona (blue dots) with the best model fit (red line).

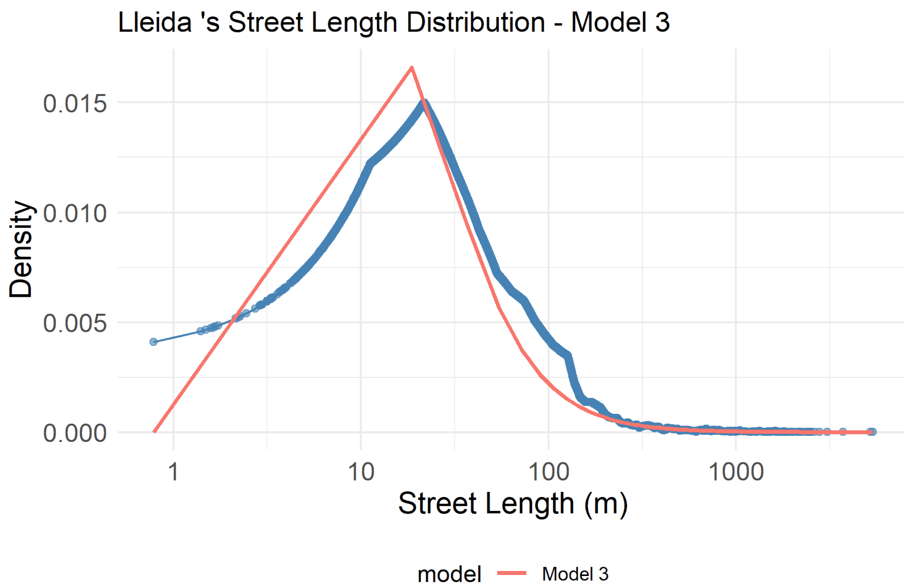

Fig. 19. Street length distribution in Lleida (blue dots) with the best model fit (red line).

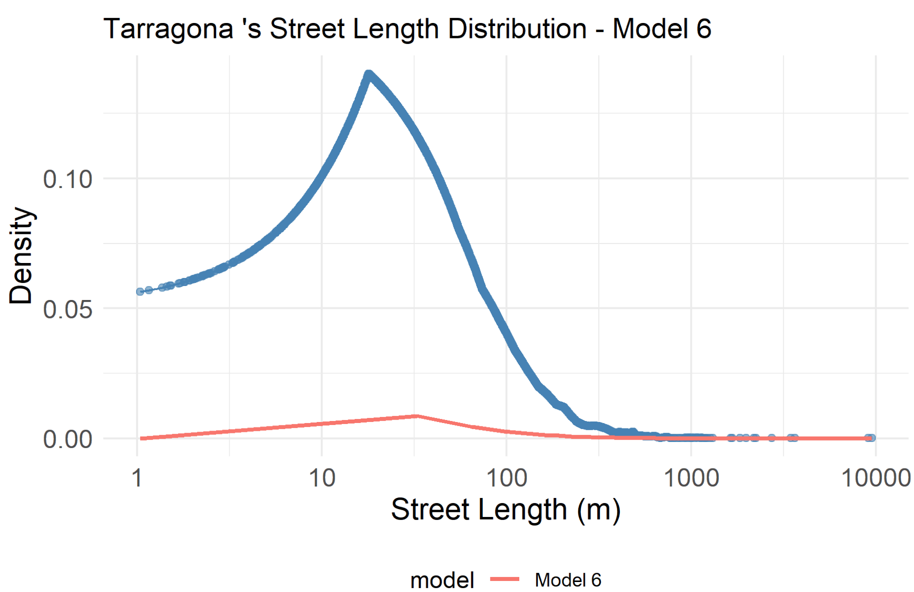

Fig. 20. Street length distribution in Tarragona (blue dots) with the best model fit (red line).

4 Discussion

4.1 Measures of Street Network Structure

Measures For each metric, we can find some observations that we want to mention:

- Number of nodes (**N**): Barcelona has the largest amount of nodes (8885) because it is the largest city in Catalonia; on the opposite side, Prats de Lluçanès (153) and el Pont de Suert (189) have significantly fewer nodes, reflecting simpler street networks which mean that are the smallest ones with around 2.500 inhabitants.
- Average Node Degree (**D**): Most cities have an average node degree of around 4, indicating that, on average, each node (intersection) connects to 4 other nodes. Higher values like in Prats de Lluçanès (5.12) suggest more interconnected street networks.
- Total Length (m) (**L**): Barcelona again leads with the highest total street length (1,392,727.88 meters), consistent with its large size. Smaller towns have much shorter total lengths, e.g., Prats de Lluçanès (20,213.62 meters).
- Average Street Length (m) (**S**): The average street length varies, with el Pont de Suert having the longest average street length (312.20 meters), indicating longer streets due to its intricate situation in the valleys of a mountain. Cities like Sabadell (79.37 meters) have shorter average street lengths, suggesting a denser street network.
- Average Betweenness Centrality (Nodes) ($\mathbf{B_N}$): Values are generally low, with Barcelona having the lowest (0.0050), indicating that no single node is overly critical for connectivity. This is due to the Eixample's grid shape. Higher values in smaller towns like Sort (0.0728) suggest specific nodes are more critical for connectivity. Sort is a town built in a valley, and 2 main roads meet in the middle of the town, being this a node with a very high Betweenness Centrality.
- Average Betweenness Centrality (Edges) ($\mathbf{B_E}$): Similar to node betweenness, values are low overall, with Barcelona having the lowest (0.0028). Higher values in smaller towns like Sort (0.0338) indicate certain edges are more critical for connectivity, the two roads mentioned above.
- Normalized Measure of Orientation-Order (ϕ): Values vary, with Granollers having the highest value (0.2737), indicating a more ordered street orientation. The most unordered town is Manresa (0.0062).
- Proportion of Dead-ends ($\mathbf{P_D}$): Varies significantly, with el Pont de Suert having the highest proportion (0.2487), indicating many dead-ends. Cities like Sabadell (0.0407) have fewer dead-ends, suggesting a more interconnected network.
- Proportion of k=4 Intersections ($\mathbf{P_4}$): Barcelona has the highest proportion (0.2929), indicating many 4-way intersections. Smaller towns like Sort (0.0591) have fewer 4-way intersections.
- Detour Index (**I**): Values are generally close to 1, indicating efficient street networks. Higher values like in el Pont de Suert (1.3704) suggest less direct routes from node to node.

We can see that larger cities like Barcelona have more complex and interconnected street networks, while smaller towns have simpler networks with more critical nodes and edges. Cities with higher average node degrees and lower proportions of dead-ends tend to have more interconnected street networks. Orientation entropy and normalized measure of orientation-order provide insights into the randomness and order of street orientations, with larger cities showing more order.

If we look at table 3, we see that for both methods of normalizing, the Orientation Entropy is higher or around 0.9. The ideal value should be around 0.45-0.55, meaning that for all our cities, the values are way higher than the desired value (see [7], [6], and [1]).

Betweenness Centrality of Vic Figure 1 shows that the streets that are most likely to be crowded are the ones that surround the city. If we look in detail, "Ronda de Camprodon" street and "Plaça del Mil·lenari" Square are some of the streets with the highest betweenness centrality values and are also some of the most transited ones (see [10]), confirming the relation between the street metrics and the real-life city fluxes.

Polar histograms If we look at Figure 4, we can clearly see the relation between the polar orientation of the streets and the histogram as Granollers is a city mainly built almost from North-South and Est-West streets. If we look at Figure 2 we can see the correlation between the entropy of the streets and their polar histograms. The less entropy they have, the more defined the orientations of the streets are: meaning that there are fewer bins with a high accumulated value (only four bins with values would be a perfect grid) and the more entropy we have, the more divided are the values amongst all the bins tending to form a circular shape.

Correlations

- Strong Positive Correlations: The Number of Nodes is highly correlated with the Total Length (r=0.99) and moderately with the Proportion of k=4 Intersections (r=0.57). This suggests that larger networks tend to have longer total street lengths and a greater proportion of four-way intersections. Average Betweenness Centrality (Nodes) and Average Betweenness Centrality (Edges) (r=0.99) show a near-perfect correlation, indicating that these centrality measures tend to align in their evaluation of network connectivity.
- Negative Correlations: The Average Node Degree negatively correlates with the Number of Nodes (r=0.52), suggesting that as networks grow in size, the average degree per node decreases. Normalized Measure of Orientation-Order has a perfect negative correlation with Orientation Entropy (r=1.00), reflecting that higher spatial order (grid-like patterns) directly corresponds to lower entropy due to its definition. Average Betweenness Centrality is also

negatively correlated with the number of nodes, suggesting that larger cities tend to have less centrality.

- Interesting Trends: The Proportion of Dead-ends is positively correlated with the Detour Index (r=0.87), indicating that networks with more dead-ends tend to have more circuitous routes. This happens because dead-ends create difficulty for the node to communicate with all the other nodes.

  Average Street Length correlates positively with the Proportion of Dead-ends (r=0.64) and the Detour Index (r=0.69), suggesting that longer streets are associated with less connected networks. This could be the case of more mountainous cities where they tend to have a main street that follows the valley and a lot of dead ends connected to it, leading to less connectivity shown in the Detour index relation.

  Some metrics, such as Orientation Entropy and Number of Nodes (r=0.05), show little to no correlation, indicating that the size of the network does not strongly affect its spatial disorder.

Density functions If we look at Figures 7, 8, 9 and 10 we can see that most of the distributions are bimodal, suggesting two different types of cities in the distribution. By looking at the *vegueries* we can see that more cosmopolitan regions have less betweenness centrality due to their necessity to have fewer hot spots that may lead to high-density traffic. This is the case of Barcelona Metropolitan Area, Tarragona, one modality of the Lleida vegueria (being the city of Lleida) and one modality of Girona (also being the city of Girona). On the other side, we can see that smaller cities and more mountainous places have a higher betweenness centrality: Alt Pirineu i Aran is the most obvious case, but if we look closely, we can see that one modality of Catalunya central, one of Camp de Tarragona and one of Girona also have higher centrality values, corresponding to their mountainous regions. In the middle are the smaller cities in the plain, like Lleida, Terres de l'Ebre, and the majority of Catalunya Central. If we look at the betweenness Centrality of the Provinces we can reach a similar conclusion, but as the densities are more general, it is more difficult to understand them.

The Orientation measure grouped by vagueries does not differ much among all the distributions, corroborating that it isn't correlated with almost any other measure.

Clustering By looking at Figure 13 we see a clear relation between dimension 1 (TSNE1) and the size of the cities (Number of nodes). The second dimension (TSNE2) is mainly defined by the entropy of the streets having the least entropic city at the top (Granollers) and the most entropic one at the bottom (Manresa). The clusters are formed amongst these two relations.

In Figure 16, we left out the size factor, and the results are slightly different and less obvious. In the top left corner, we have the more mountainous capitals. The first dimension (TSNE1) is influenced by the orientation indicator. Cluster 5 is influenced by the average node degree containing cities with the highest values like Prats de Lluçanès, Gandesa or Moià. Clusters 1 and 2, as they are the mountainous regions, have the highest Detour Index values.

4.2 Street Segment Length Analysis

Tables 12 and 13 summarise our results. We can see that the best model for Barcelona is model 4 with parameters $\alpha = 5.776$, $\beta = 0.01149$. The closest competitor is model 6, with a Δ of 1.96. Nevertheless, we see that model 6 has parameters $\alpha = 5.751$, $\beta = 0.01148$, and $\gamma = 0$, so it is essentially the same model. In the case of Girona, the best model is model 6 with parameters $\alpha = 7.29$, $\beta = 0.01006$, and $\gamma = 0.3452$. The closest competitor is model 4, which has a Δ of 31.99. For Tarragona, model 6 with parameters $\alpha = 10.43$, $\beta = 0.003323$, and $\gamma = 1.037$ is the best. All the other models are substantially far from it in terms of Δ. Finally, the best model for the city of Lleida is model 3 with parameters $\alpha = 28.02$, and $\gamma = 1.927$. With a Δ of 3.59, model 6 is the closest competitor. Actually, in terms of goodness of fit, model 6 is better. It has a value of 22 380.24 for the negative log-likelihood versus 22 699.48 for model 3, but it is penalised for its higher complexity.

We observe that the fit of the distribution from the null model (model 1) differs significantly from that of the others. It achieves the highest Δ value for all cities, which in all cases is at least one order of magnitude larger than the others (see Table13).

Figure 17 presents the empirical street length distribution for Barcelona with the best model fit overlaid. Overall, the model provides a reasonably good approximation. However, it underestimates the distribution for street lengths between approximately 1 and 3 meters, then overestimates it until about 10 meters. Additionally, while the model suggests a unimodal distribution, the empirical data display a trimodal pattern.

Figure 18 displays the empirical street length distribution for Girona alongside the best model fit. The model generally aligns well with the data but underestimates the distribution for very short street segments and then overestimates it until around 32 meters. Importantly, it captures the unimodal nature of the empirical distribution.

Figure 19 shows the empirical street length distribution for Lleida with the model fit superposed. The fit is broadly acceptable. However, the model underestimates the distribution for the shortest lengths and overestimates it up to 32 meters. Near 100 meters, the model again underestimates the empirical distribution. Nonetheless, it successfully captures the unimodal shape of the observed data.

In contrast, Figure 20 illustrates a terrible fit for Tarragona's street length distribution. The model substantially underestimates the distribution across almost the entire range of street lengths, indicating that the models tried are inadequate for capturing the distribution of Tarragona's street segment length.

The findings reported in [13] for London do not straightforwardly extend to the four additional cities examined in this study. In their work, the analogue of our model 6 emerged as the best fit for London, with estimated parameters $\alpha = 145$, $\beta = \frac{1}{2000}$, and $\gamma = 3.36$.

By contrast, in our analysis, model 6 was the best fit for only two of the four cities: Girona and Tarragona. Moreover, the fit for Tarragona was notably poor.

It is important to note that in [13], they did not explicitly detail their methodology for selecting the best-fitting model for London's street segment lengths. Therefore, this complicates any direct numerical or procedural comparison between their results and those obtained in our study.

5 Conclusions

While this study provides valuable insights into the structure of street networks in the cities it examines, it also has several limitations that highlight areas for future improvement.

One limitation is the analysis's aggregation level, which considers the entire city a single unit. Such a coarse level of granularity may not provide sufficiently actionable insights. Finer levels of analysis, such as neighbourhoods or even census districts, could yield more targeted recommendations for urban planning and design.

Another consideration is that focusing solely on the street network captures only one facet of a city. A city is inherently a complex three-dimensional system comprising not only streets but also social, economic, and environmental networks. Although these interconnected systems are beyond the scope of the present study, they should be considered for comprehensive urban planning and design.

Furthermore, the analysis presented is primarily descriptive, identifying empirical patterns without prescribing specific interventions. In professional city-making disciplines, merely identifying empirical values is insufficient. A key future challenge would be to bridge the gap between the "what"—the descriptive measures—and the "how".

References

1. Analytics, A.U., Design: A vision for barcelona's future: City science illuminating urban and economic development (2024), https://www.aretian.com/visionforbarcelona, executive Summary Report
2. Barthelemy, M.: Spatial Networks: A Complete Introduction: From Graph Theory and Statistical Physics to Real-World Applications. Springer Nature, Heidelberg (2022). https://doi.org/10.1007/978-3-030-94106-2, https://doi.org/10.1007/978-3-030-94106-2
3. Barthelemy, M., Boeing, G.: A review of the structure of street networks. Findings (August 2024). https://doi.org/10.32866/001c.122117, https://doi.org/10.32866/001c.122117
4. Boeing, G.: Urban spatial order: Street network orientation, configuration, and entropy. Applied Network Science **4**(1), 1–9 (2019). https://doi.org/10.1007/s41109-019-0189-1, https://doi.org/10.1007/s41109-019-0189-1

5. Boeing, G.: Modeling and analyzing urban networks and amenities with osmnx (2024), https://geoffboeing.com/publications/osmnx-paper/, working paper
6. Burke, J., Gras Alomà, R.: City Science: Performance Follows Form. Actar (2023)
7. Burke, J., Alomà, R.G., Yu, F., Kruger, J.: Geospatial analysis framework for evaluating urban design typologies in relation with the 15-minute city standards. Journal of Business Research **151**, 651–667 (2022). https://doi.org/10.1016/j.jbusres.2022.06.024, https://www.sciencedirect.com/science/article/pii/S014829632200563X
8. Everitt, B.S., Landau, S., Leese, M., Stahl, D.: Cluster Analysis. Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd, 5th edn. (2011). https://doi.org/10.1002/9780470977811, https://doi.org/10.1002/9780470977811
9. Institut Cartogràfic i Geològic de Catalunya: Publicacions cartogràfiques. https://www.icgc.cat/ca/Divulgacio/Publicacions/Publicacions-cartografiques (2024), accessed: 2025-01-15
10. Jaumira, A.: Els fums del carrer més transitat de vic. https://el9nou.cat/osona-ripolles/actualitat/societat/els-fums-del-carrer-mes-transitat-2/ (2019), online; accessed 2025-01-15
11. van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research **9**(86), 2579–2605 (2008), http://jmlr.org/papers/v9/vandermaaten08a.html
12. Marta Arias, R.F.i.C., Arratia, A.: Analysis of the degree distribution. https://www.cs.upc.edu/~CSN/lab.html (2024), complex and Social Networks (2024-2025), Master in Data Science (MDS)
13. Masucci, A., Smith, D., Crooks, A., Batty, M.: Random planar graphs and the london street network. Eur. Phys. J. B **71**(2), 259–271 (2009). https://doi.org/0.1140/epjb/e2009-00290-4