Graph entropy as tool for understanding complex urban networks. The case of Ensenada city, Mexico

Antonio Aguilera*

El Colegio de San Luis, A.C.,
Parque de Macul 155, Fracc. Colinas del Parque,
C.P. 78299, San Luis Potosi, San Luis Potosi, Mexico
Fax: +52-444-8110101
E-mail: aaguilera@colsan.edu.mx
*Corresponding author

Edgardo Ugalde

Instituto de Física,
Universidad Autonoma de San Luis Potosi,
Av. Manuel Nava 6 Zona Universitaria,
C.P. 78290, San Luis Potosi, San Luis Potosi, Mexico
E-mail: ugalde@ifisica.uaslp.mx

Abstract: Spatial structure of cities is the substratum over which human urban life develops. Understanding the relevance of the streets that form such substratum is fundamental to understand the movements of people in the city. Our purpose in this study is to propose and explore methods to find relevant streets sets that conforms the basis of complex urban networks. We used two methods based on previous work by Volchenkov and Blanchard (2008) and Shetty and Adibi (2005), and a method developed by us. To illustrate the use of these tools, we performed an analysis of the main streets network of the coastal city of Ensenada, Mexico.

Keywords: graph entropy; urban structure; complex urban networks.

Reference to this paper should be made as follows: Aguilera, A. and Ugalde, E. (2014) 'Graph entropy as tool for understanding complex urban networks. The case of Ensenada city, Mexico', *Int. J. Society Systems Science*, Vol. 6, No. 1, pp.87–99.

Biographical notes: Antonio Aguilera is a Professor in El Colegio de San Luis, a Mexican Public Research Center in Social Sciences. He is a National Researcher Level I. His research involves the development of computer models of spatial socio-economic phenomena and spatial analysis methods. He is a member of the National Research System of Mexico and the European Social Simulation Association.

Edgardo Ugalde is a Professor in the Physics Institute of the Universidad Autonoma de San Luis Potosi, Mexico. He is a National Researcher Level II. His research interests are dynamic systems, discrete mathematics and modelling and data analysis. His is a member of the National Research System of Mexico, and the Mexican Academy of Sciences. In 2011, he received the prize 'Jorge Lomnitz Adler' from the Institute of Physics of the Universidad Nacional Autonoma de Mexico (UNAM).

1 Introduction

Spatial structure of a city is a regular and permanent substratum on which human functions are developed. Indeed, sets of streets, intersections and buildings form the basic systems on which human urban life is developed. One relevant factor of urban structure is connectivity (e.g., Filin and Vered, 2009; Krüger, 1979, 1980; Marshall, 2005; Sampson, 2004) which is related to several social and economic phenomena since crime to community health.

In geographical and urban studies, entropy has been used broadly to study spatial structure and population concentration (Batty, 1972, 1974, 1976), spatial interaction (Honma et al., 2010; Nijkamp and Reggiani, 1988; O'Kelly, 2010; Wilson, 1970) spatial urban segregation (Massey and Denton, 1988; White, 1983, 1986; Wong, 2002; Yeh and Li, 2001) and urban sprawl (Rashevsky, 1955; Tsai, 2005). As a tool for spatial analysis, graph entropy has been recently suggested to find relevant streets related with their connectivity and centrality (Shetty and Adibi, 2005; Volchenkov and Blanchard, 2008).

Given that the urban structure can be abstracted using the techniques of space syntax (Hiller, 1996; Marshall, 2005) and following works of Shetty and Adibi (2005) and Volchenkov and Blanchard (2008), we developed a new technique to find relevant streets based on a graph-theoretical notion of entropy which incorporates public and private services on streets. To illustrate the theory, we performed a study of the main streets network of the Mexican city of Ensenada.

The structure of this paper is follows: in the next section the mathematical fundamentals on graph theory and graph entropy are developed. Then, in Section 3, Ensenada city main streets network is analysed, and results are showed. The final section is devoted to the concluding remarks.

2 Graph theory and graph entropy

A network can be mathematically represented as a graph $G(V, E)$ using two sets. First one, a non-empty and finite set of vertices or nodes V, and second one a non-empty and finite set of links E. Then each element of E relates one element of V with another element of V. Formally, each $e \in E$ is associated to a couple $(v_i, v_j) \in V \times V$. In our study, we are interested only in undirected graphs without loops, then the following restrictions must be made to definition: if (v_i, v_j) is associated to e, then (v_j, v_i) is also associated with e, and no couple of the kind (v_i, v_i) is associated to an element of E.

Given a graph with a vertex set V, the degree of a vertex $v \in V$ is the number of edges associated a couple in $V \times V$ containing v, and it is symbolised by $\deg(v)$. Another graph measure is betweenness. It was defined by Freeman (1977) as follows. Given a vertex $v_k \in V$ in a graph $G(V, E)$ and an couple of vertices (v_i, v_j), where $i \neq j$, the partial betweenness $b_{ij}(v_k)$ is the number of shortest paths connecting v_i and v_j and passing through v_k, divided by the total number of shortest paths connecting v_i and v_j. Hence, if $g_{ij}(v_k)$ denotes the number of shortest paths connecting v_i and v_j and passing through v_k, and g_{ij} denotes the total number of shortest paths connecting v_i and v_j, then

$$b_{ij}(v_k) = g_{ij}(v_k) / g_{ij} \tag{1}$$

can be understood as the probability that vertex v_k belongs on a randomly selected geodesic connecting v_i and v_j. The overall centrality of a vertex v_k is defined as the sum of its partial betweenness for all unordered pairs of vertex,

$$C_B(v_k) = \sum_{i<j}^{N} \sum_{j}^{N} b_{ij}(v_k) \tag{2}$$

where $i \neq j \neq k$, and N is the total number of vertex in the graph. To understand graph entropy is necessary introduce first the concept of entropy. It is a parameter used to describe the uncertainty of a system. It is based on the measure of the probability distribution of the states reachable by a system. In 1948, Shannon proposed a general entropy formulation:

$$S(\Omega, P) = -K \sum_{i=1}^{\Omega} P_i \log P_i \tag{3}$$

where Ω is the number of possible states in the system, P_i in the probability of the i^{th} state, and K is a constant. Generally, $K = \log_2 10$ is suitable. Using a change on the base of the logarithm, entropy can be rewritten as:

$$S(\Omega, P) = -\sum_{i=1}^{\Omega} P_i \log_2 P_i \tag{4}$$

The concept of entropy for the study of graphs was introduced by Rashevsky (1955) for the study of graph topology. Köhner (cited by Dehmer and Mowshowitz, 2011) introduced a graph entropy measure to determine the performance of a best possible encoding of messages emitted by an information source where the symbols belong to a finite vertex set.

In the context of our work, graph entropy is defined by Volchenkov and Blanchard (2008) and Shetty and Adibi, (2005). In the work of Volchenkov and Blanchard (2008), there are two formulations for the specification of probabilities associated with each vertex $v \in V$. The first one uses the concept of local connectivity of each vertex. Using a probability distribution defined as:

$$\pi_i = \deg(v_i) / 2M, \tag{5}$$

where $M = | E |$ is the total number of links in the graph. Volchenkov and Blanchard (2008) use another probability distribution based on the concept of betweeness. Using equation (2), this probability is defined as:

$$p_i = C_B(v_i) / C_B, \tag{6}$$

where $C_B = \sum_{i=1}^{N} C_B(v_i)$.

Using either kind of probability distributions, the graph entropy, which it is understood as the structural information of a graph G, it is defined as:

$$H(G, P) = \sum_{i=1}^{N} p_i \log_2 (1 / p_i) \tag{7}$$

Following Volchenkov and Blanchard (2008), if P is the probability distribution that is function of node's degree, then entropy is called connectivity entropy. If P is the probability distribution related to the betweeness of a node, the entropy is named centrality entropy.

The entropy $H(G, P)$ has the property that it is additive independently of how network is being divided into parts, therefore it can be calculated as the sum of the partial entropies of its single nodes. Volchenkov and Blanchard (2008) defined the relevance of a node in terms of its contribution to the entropy. It could be estimated by the entropy participation ratio (EPR), which is defined as

$$h_i = \left(p_i / H(G, P) \right) \log_2 \left(1 / p_i \right), \tag{8}$$

where $H(G, P)$ can be either connectivity entropy or centrality entropy.

Shetty and Adibi (2005) suggested an alternative method to discover relevant nodes in a graph. The main idea is that the result of adding or removing edges in a graph affects its overall connectivity. Then the interpretation of relevant nodes of a graph is that there are those who have the most effect in the graph entropy when they are removed from the graph. The procedure suggests is as follows:

1 entropy of the whole graph must be calculated

2 an arbitrary node is removed from the graph and the graph entropy of the remaining graph is calculated.

This step should be repeated until all nodes are removed. In order to measure the effect of removing the node, the following quantity must be calculated:

$$H_{effect}\left(v_i \right) = H_{\neq v_i} / \log_2 \left(H_{\neq v_i} / H \right). \tag{9}$$

In our problem, we have not only the topology of the graph, but also socio-economical data must be incorporated to understand the relevance of streets with respect to public and private infrastructure as hospitals, schools, cultural centres and so on. In order to incorporate this kind of data, we propose a new probability distribution on the node that incorporates the 'density' of the social phenomena in which we can be interested. The new probability distribution is defined as follows:

$$p\left(v_i \right) = \left(\delta\left(v_i \right) - \sum\nolimits_{k \in N\left(v_i \right)} \left| g\left(v_i \right) - g\left(v_k \right) \right| \right) / \sum\nolimits_i \left(\delta\left(v_i \right) - \sum\nolimits_{k \in N\left(v_i \right)} \left| g\left(v_i \right) - g\left(v_k \right) \right| \right), \tag{10}$$

where $\delta(v_i)$ is the degree of node v_i, $g(v_i)$ quantifies the density of the socio-economical data at node v_i, $g(v_k)$ is the corresponding quantity at node v_k which is adjacent to node v_i, and $N(v_i)$ is the size of the neighbourhood of node v_i.

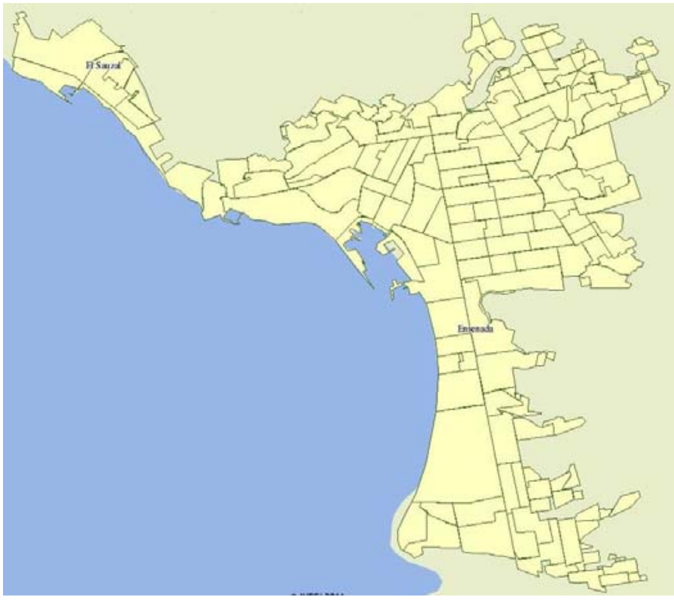
3 Study of Ensenada's main streets network

Ensenada is a coastal city located in the Mexican Pacific Ocean shore [see Figure 1(a)]. It is the capital of Mexican State of Baja California. Ensenada has a population of 466,814 inhabitants. The commercial relevance of the city is based in its closeness to the

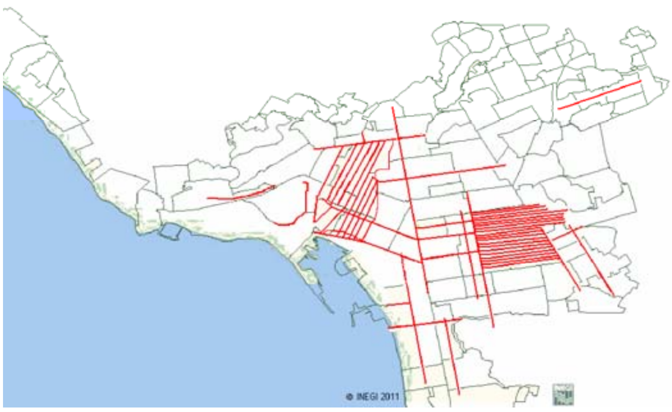
American port of San Diego and it is the Mexican offshore port most near to Asia. The city is home of the Center of Scientific Research and Higher Education of Ensenada (CICESE for its acronym in Spanish) which is the more important oceanographic research centre in Mexico. Also, Ensenada is a strategic place for Mexican homeland security and the city is the base of important Mexican Navy and Mexican Air Force facilities.

Ensenada's main street network is formed by 44 streets with 141 intersections between them. The main streets pattern is showed in Figure 1(b).

Figure 1 (a) Ensenada City (b) Ensenada's main streets (see online version for colours)



(a)



(b)

Table 1 Results of computed structural measures for the Ensenada's streets network

Street name	*Node*	*Grade*	*EPR connectivity*	*EPR centrality*	*Shetty's entropy*
Ambar	1	8	0.0497	0.0828	2.9031
Ryerson	2	2	0.0172	0	2.9264
Alvaro Obregon	3	2	0.0172	0.0012	2.9338
Ruiz	4	3	0.0237	0.0057	2.9245
Gastelum	5	1	0.0098	0	2.9357
Miramar	6	3	0.0237	0.005	2.9335
Riveroll	7	4	0.0297	0.0111	2.9364
Mar	8	7	0.0451	0.0338	2.9379
Reforma	9	11	0.0622	0.1495	2.9316
Bronce	10	2	0.0172	0	2.9322
Alvarado	11	4	0.0297	0.0056	2.9317
Blancarte	12	4	0.0237	0.0047	2.9363
Castillo	13	4	0.0297	0.0056	2.9317
Benito Juarez	14	8	0.054	0.0577	2.9292
Adolfo Lopez Mateos	15	8	0.0497	0.0888	2.9325
Lazaro Cardenas	16	7	0.0451	0.0182	2.9184
Doctor Pedro Loyola	17	3	0.0237	0.0122	2.9291
Estancia	18	3	0.0237	0.0402	2.9085
General Juan Zertuche	19	1	0.0098	0	2.9262
Ignacio Manuel Altamirano	20	1	0.0098	0	2.9324
Manuel Ponce	21	1	0.0098	0	2.9324
De Las Aguilas	22	3	0.0237	0.0642	2.9254
Gregorio Torres Quintero	23	1	0.0098	0	2.9324
Libertad	24	1	0.0098	0	2.9324
Eusebio Francisco Kino	25	1	0.0098	0	2.9324
Cortez	26	3	0.0237	0.0642	2.9254
Primer Ayuntamiento	27	1	0.0098	0	2.9324
Del Magisterio	28	1	0.0098	0	2.9371
Del Puerto	29	1	0.0098	0	2.9371
Benito Juarez 1	30	1	0.0098	0	2.9371
16 de Septiembre	31	2	0.0172	0.0055	2.9299
Diamante	32	6	0.0403	0.1289	2.9059
18 de Marzo	33	1	0.0098	0	2.9371
Niños Heroes	34	1	0.0098	0	2.9371
Francisco Gonzalez Bocanegra	35	1	0.0098	0	2.9371
Cuahutemoc	36	1	0.0098	0	2.9371
Jaime Nuno	37	1	0.0098	0	2.9371
Pipila	38	4	0.0297	0.0178	2.9109

Table 1 Results of computed structural measures for the Ensenada's streets network (continued)

Street name	*Node*	*Grade*	*EPR connectivity*	*EPR centrality*	*Shetty's entropy*
Real Del Castillo	39	1	0.0098	0	2.9339
Aseguradores	40	1	0.0098	0	2.9339
General Lazaro Cardenas1	41	2	0.0172	0.0376	2.9465
Pablo Orta	42	1	0.0098	0	2.9465
Mexico	43	18	0.0863	0.1597	2.6945
Zertuche	44	1	0.0098	0	2.9465

Figure 2 EPR for connectivity of the main streets of Ensenada (see online version for colours)

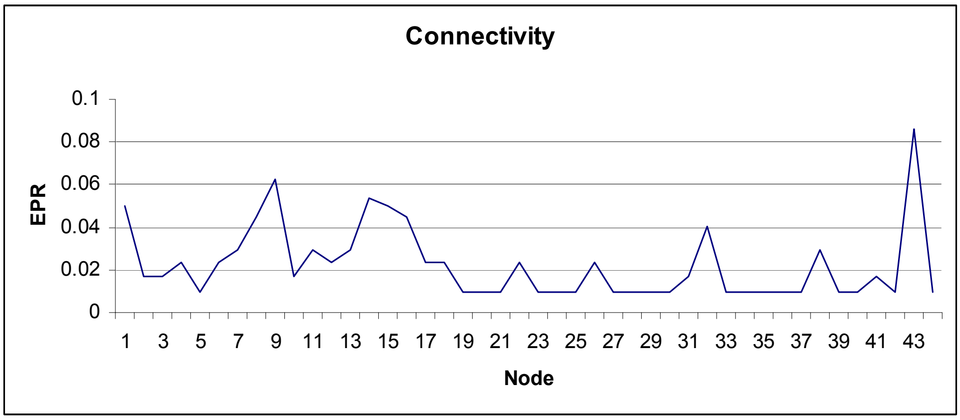
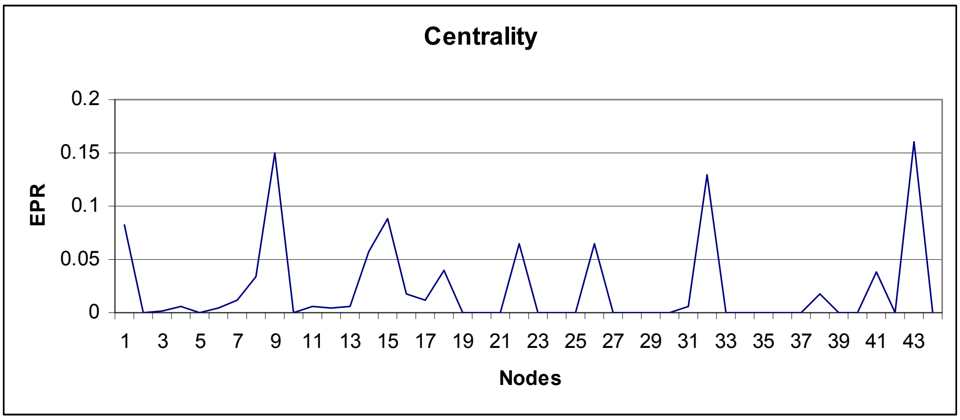


Figure 3 EPR for centrality of the main streets of Ensenada (see online version for colours)



This network was coded as a dual graph where streets are nodes and intersections are links. We build the 44 × 44 adjacency matrix for the graph and we computed both connectivity entropy and centrality entropy. The value for connectivity entropy is 2.9360 and for centrality entropy is 3.1740. It must be noted that the value of both entropies for the fully connected graph on 44 nodes is $\log_2 N = \log_2(44) = 5.4594$. We use equation (8) to calculate the EPR for both connectivity and centrality. Also, with equation (9) we computed the relevant nodes according to Shetty and Adibi's method, which for simplicity we named as Shetty's entropy. Results can be observed in Table 1 and in Figures 2 and 3.

Using result for connectivity EPR and for centrality EPR, we create a street's hierarchy showed in Figures 4 and 5.

Figure 4 Street's hierarchy for connectivity EPR (see online version for colours)

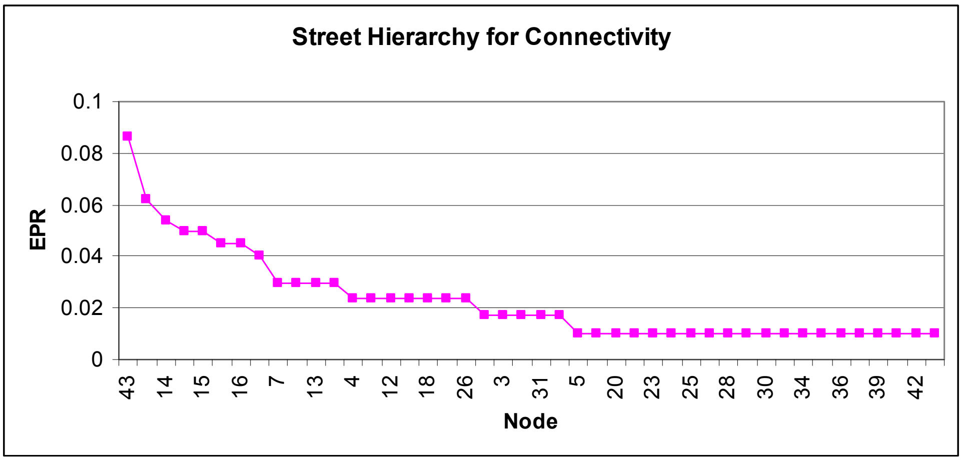
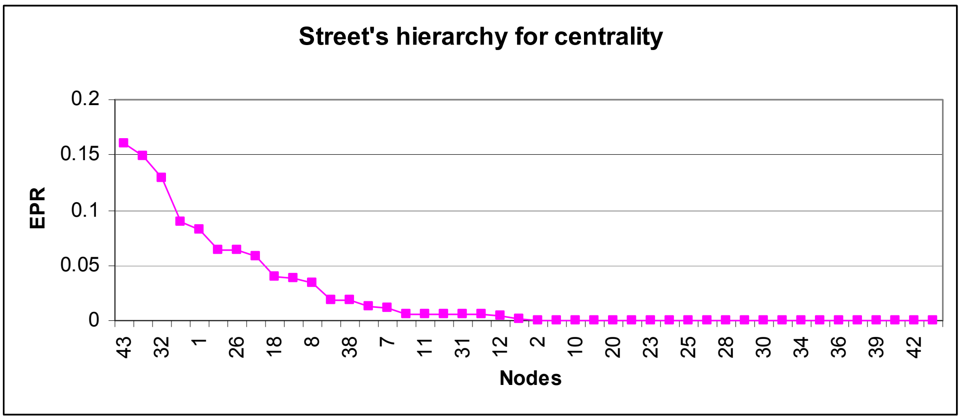
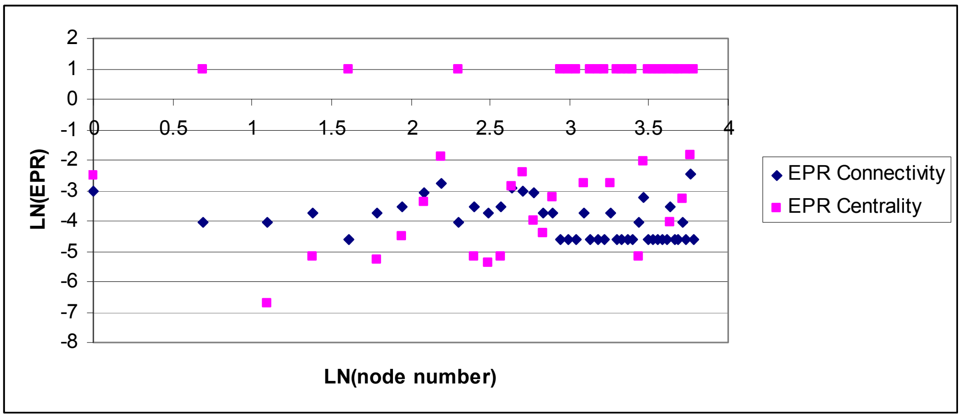


Figure 5 Street's hierarchy for centrality EPR (see online version for colours)



The interpretation of the results for EPR's entropies is as follows. Connectivity entropy is a local measure of the structure of the network, whereas centrality entropy is a global measure of the structure of the network. In spatial syntax, the intelligibility of a network is defined as the degree of correlations between connectivity and centrality (Hiller, 1996). In our case, connectivity EPR and centrality EPR have a Pearson correlation index of –0.684130, whereas a correlated set of data must have a Pearson correlation index close to 1. Since in our case connectivity EPR and centrality EPR are not correlated, we can conclude that Ensenada's street network has a poor intelligibility. This can be seen in a graphical way in Figure 6.

Figure 6 The rank-EPR plot (in LN-LN scale) of the Ensenada's streets network (see online version for colours)



Relevant streets for connectivity, in order of relevance, are presented in Table 2.

Table 2 Relevant streets under connectivity EPR criteria

Node number	*Street name*	*Grade*
43	Mexico	18
9	Reforma	11
14	Benito Juarez	8
1	Ambar	8
15	Adolfo Lopez Mateos	8
8	Mar	7
16	Lazaro Cardenas	7
32	Diamante	6

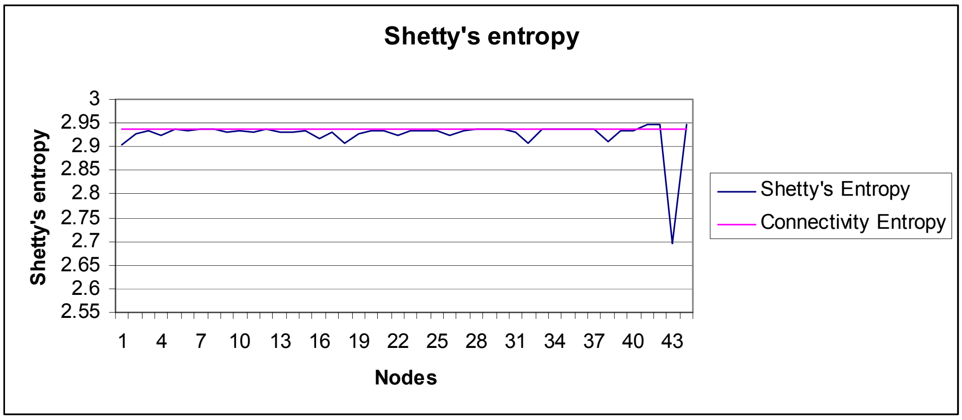
Obviously, nodes with greater grade are more relevant for connectivity. It is worth noticing that the sub-network formed by the relevant streets it is composed by two clearly distinguishable clusters, one around Reforma Avenue and the other around Mexico Avenue. These clusters are connected only by the Diamante Avenue, which is therefore a critical node.

For the case of centrality EPR, there are only three relevant nodes, 43, 9 and 32, which correspond to Mexico Avenue, Reforma Avenue and Diamante Avenue. This result reinforces our conclusion of the EPF connectivity analysis. Then we can asseverate that the sub-network of Ensenada's main streets is formed only by three nodes that connect Reforma Avenue and Mexico Avenue through Diamante Avenue. Mexico and Reforma avenues are North-South Streets and Diamante Avenue is a East-West street.

Now, we interpret results from Shetty's entropy. Connectivity entropy for the whole networks is the constant line in Figure 7, whereas the shaped line corresponds to the connectivity entropy obtained by removing the corresponding node. It is possible to observe that there is only one relevant node that increase entropy connectivity, namely node 42 (Pablo Orta Avenue), whereas the removal of node 43 (Mexico Avenue) decreases substantially the connectivity of the whole network.

It is important to remark that method suggested by Volchenkov and Blanchard (2008) and method proposed by Shetty and Adibi (2005), both are complementary and give us different information about the structure of the whole network.

Figure 7 Shetty's entropy for Ensenada's street network (see online version for colours)



Now, we show the results obtained using our proposed joint entropy [equation (10)]. For this, we first consider the presence of public and private hospitals as socio-economical data. There are 24 hospitals and health clinics in Ensenada city. Seventeen are public hospitals and the rest are private. Only five of them are on main streets (see Table 3).

Table 3 Hospital's location in Ensenada's main street network

Hospital name	*Main street*	*Node number*
Unidad de Especialidades Medicas	Gastelum	5
Hospital General de Zona No. 8	Reforma	9
Hospital General Clinica 8 Urgencias	Reforma	9
Hospital San Jose	General Lazaro Cardenas	41
CARDIOMED	Alvaro Obregon	3
Sanatorio Naval de Ensenada	Lazaro Cardenas	16

To find relevant nodes related to the presence of hospitals, we used the previous defined concept of EPR. As there are very few hospitals on the main streets, this means that the joint entropy related to the hospitals must be similar to connectivity entropy. This can be seen in the Figure 8.

Now, if our socio-economical data is the presence of economic spatial units such as retail stores, shopping centres, drugstores, hardware stores, cloth stores and so on, we would expect to see a difference between the joint EPR and the connectivity EPR. However, we can see in Figure 9 that the behaviour of both EPRs is very similar.

Figure 8 Joint EPR and connectivity EPR for the Ensenada streets network with hospitals (see online version for colours)

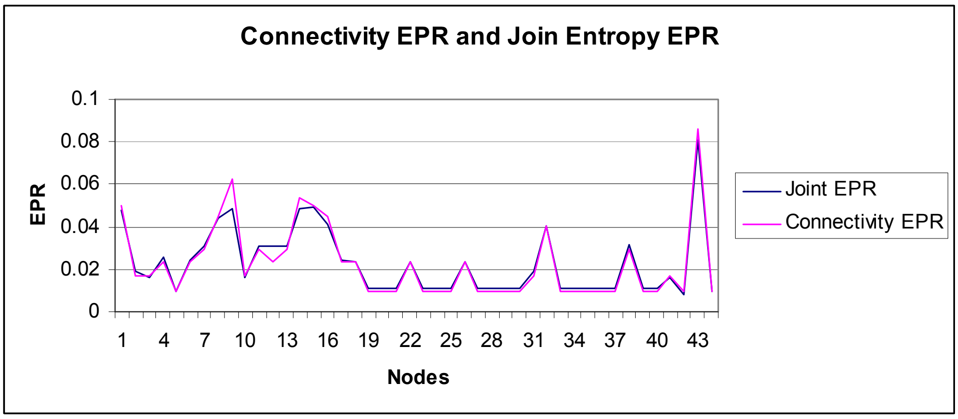
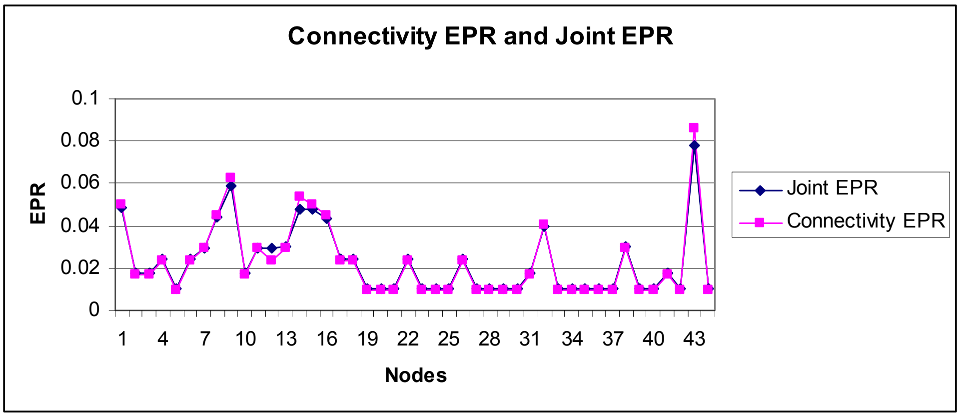


Figure 9 Joint EPR and connectivity EPR for the Ensenada streets network with economic spatial units (see online version for colours)



It seems that for our proposed joint entropy the number of economic units by street is of no relevance. Then we can interpret this as follows: the spatial structure is the relevant phenomena and the human activities and services rest over that structure having no real play with respect to the behaviour of city.

4 Conclusions

In this work, we have described three methods to discovering relevant streets in urban networks. Relevance of streets is function of their connectivity and their centrality. The first method called EPR allowed us to measure the intelligibility of the whole main street network of Ensenada City. We found a poor intelligibility in this case. We also found that the sub-network of relevant nodes is formed by two clusters liked by a single critical node. Using the second method, we found that there is only one relevant node that increases entropy connectivity, whereas the removal of only one node decreases substantially the connectivity of the whole network. This is a consistent result because the removed node corresponds to a street with high connectivity and centrality measures. The third method showed, at least for the socio-economical data we considered, that the density of economic spatial units is completely determined by the topology of the network.

Our results show that graph entropy and specifically EPR method and Shetty's entropy, are suitable to perform spatial syntax analysis. In particular, both tools can be used by planning agencies to understand the underlying relevant structure of the urban patterns.

Acknowledgements

This work was financed by a grant of CONACYT-Mexico no. 101366.

References

Batty, M. (1972) 'Entropy and spatial geometry', *Area*, Vol. 4, No. 4, pp.230–236.

Batty, M. (1974) 'Spatial entropy', *Geographical Analysis*, Vol. 6, No 1, pp.1–31.

Batty, M. (1976) 'Entropy in spatial aggregation', *Geographical Analysis*, Vol. 8, No. 1, pp.1–21.

Dehmer M. and Mowshowitz, A. (2011) 'A history of graph entropy measures', *Information Sciences*, Vol. 181, No. 1, pp.57–78.

Filin, S. and Vered, Y. (2009) 'Effect of physical constraints on spatial connectivity in urban areas', *Nordic Journal of Surveying and Real Estate Research*, Special Series, Vol. 4, pp.69–83.

Freeman, L.C. (1977) 'A set of measures of centrality based on betweenness', *Sociometry*, Vol. 40, No. 1, pp.35–41.

Hiller, B. (1996) *Space is the Machine*, Cambridge University Press, Cambridge.

Honma, Y., Kurita, O. and Taguchi, A. (2010) 'Spatial interaction model for trip-changing behavior based on entropy maximizing method', *Journal of the Operations Research Society of Japan*, Vol. 53, No. 4, pp.245–254.

Krüger, M.J.T. (1979) 'An approach to built-form connectivity at an urban scale: system description and its representation', *Environment and Planning B: Planning and Design*, Vol. 6, No. 1, pp.67–88.

Krüger, M.J.T. (1980) 'An approach to built-form connectivity at an urban scale: relationships between built-form connectivity, adjacency measures, and urban spatial structure', *Environment and Planning B: Planning and Design*, Vol. 7, No. 2, pp.163–194.

Marshall, S. (2005) *Streets & Patterns*, Spon Press, New York.

Massey, D.S. and Denton, N.A. (1988) 'The dimensions of residential segregation', *Social Forces*, Vol. 67, No. 2, pp.281–315.

Nijkamp, P. and Reggiani, A. (1988) 'Entropy, spatial interaction models and discrete choice analysis: static and dynamic analogies', *European Journal of Operational Research*, Vol. 36, No. 2, pp.186–196.

O'Kelly, M.E. (2010) 'Entropy-based spatial interaction for trip distribution', *Geographical Analysis*, Vol. 42, No. 4, pp.427–487.

Rashevsky, N. (1955) 'Life information theory and topology', *Bulletin of Mathematical Biophysics*, Vol. 17, No. 3, pp.229–235.

Sampson, R.J. (2004) 'Networks and neighborhoods: the implications of connectivity for thinking about crime in the modern city', in McCarthy, H., Miller, P. and Skidmore P. (Eds.): *Network Logic: Who Governs in an Interconnected World?*, pp.157–166, Demos Press, London.

Shetty, J. and Adibi, J. (2005) 'Discovering important nodes through graph entropy the case of Enron email database', in *LinkKDD '05: Proceedings of the 3rd International Workshop on Link Discovery*, available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.92.2782&rep=rep1&type=pdf (accessed on 7 November 2010).

Tsai, Y-H. (2005) 'Quantifying urban form: compactness versus sprawl', *Urban Studies*, Vol. 42, No. 1, pp.141–161.

Volchenkov D. and Blanchard, P. (2008) 'Discovering important nodes through graph entropy encoded in urban space syntax', available at http://arxiv.org/PS_cache/arxiv/pdf/0709/0709.4415v1.pdf (accessed on 20 September 2010).

White, M.J. (1983) 'The measurement of spatial segregation', *American Journal of Sociology*, Vol. 88, No. 5, pp.1008–1018.

White, M.J. (1986) 'Segregation and diversity measures in population distribution', *Population Index*, Vol. 52, No. 2, pp.198–221.

Wilson, A.G. (1970) 'The use of the concept of entropy in system modeling', *Operational Research Quarterly*, Vol. 2, No. 2, pp.247–265.

Wong, D.W.S. (2002) 'Modeling local segregation: a spatial interaction approach', *Geographical and Environmental Modelling*, Vol. 6, No. 1, pp.81–97.

Yeh, A.G-O. and Li, X. (2001) 'Measurement and monitoring of urban sprawl in a rapidly growing region using entropy', *Photogrammetric Engineering and Remote Sensing*, Vol. 67, No. 1, pp.83–90.