Article:
Agryzkov, Taras, Tortosa, Leandro, Vicent, Jose et al. (1 more author) (2017) A Centrality Measure for Urban Networks Based on the Eigenvector Centrality Concept. Environment and Planning B: Urban Analytics and City Science. ISSN 2399-8083

https://doi.org/10.1177/2399808317724444

Journal Title
XX(X):1–15

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

A Centrality Measure for Urban Networks Based on the Eigenvector Centrality Concept

Taras Agryzkov[1] Leandro Tortosa[1] José F. Vicent[1] and Richard Wilson[2]

Abstract

A massive amount of information as geo-referenced data is now emerging from the digitization of contemporary cities. Urban streets networks are characterized by a fairly uniform degree distribution and a low degree range. Therefore, the analysis of the graph constructed from the topology of the urban layout does not provide significant information when studying topology–based centrality. On the other hand, we have collected geo-located data about the use of various buildings and facilities within the city. This does provide a rich source of information about the importance of various areas. Despite this, we still need to consider the influence of topology, as this determines the interaction between different areas. In this paper, we propose a new model of centrality for urban networks based on the concept of Eigenvector Centrality for urban street networks which incorporates information from both topology and data residing on the nodes. So, the centrality proposed is able to measure the influence of two factors, the topology of the network and the geo-referenced data extracted from the network and associated to the nodes. We detail how to compute the centrality measure and provide the rational behind it. Some numerical examples with small networks are performed to analyse the characteristics of the model. Finally, a detailed example of a real urban street network is discussed, taking a real set of data obtained from a fieldwork, regarding the commercial activity developed in the city.

Keywords

Network graphs, street networks, spatial analysis, network centrality, eigenvector centrality

Introduction

We live in a time when the data constitute a source around which emerge new business models and new forms of exploitation. The cities, in general, are great sources of data [Behhisch and Ultsch(2007)]. Many private and public enterprises manage large volumes of data generated in urban environments where the positional component, that is, the ability to geo-locate data, is crucial to obtain valuable information for strategic decision making ([Fischer and Wang(2011), Haining(2003), Oyana and Margai(2015)]). Therefore, spatial data become important geo-marketing tools aimed at enhancing land management processes and business through the integration and exploitation of the geographical position of some human activity [Bradlow et al(2005), Gliquet(2002)].

Networks can be represented by graphs and their structure can be analysed using different concepts; one of the most important is centrality.The centrality indices [Freeman(1977)] are measures of the varying importance of the nodes in a network according to a specific geometrical or topological criterion [Crucitti et al(2006a)]. As [Porta et al(2006)] states, a set of centralities measures proposed in [Freeman(1977), Freeman(1979)] can be grouped into two different families. The first one considers a central graph entity in terms of being near to others ([Freeman(1977), Freeman(1979), Nieminen(1974), Sabidussi(1966)]), where the closeness centrality is the most representative.

The second family may be viewed as centralities in terms of being the intermediary of others [Freeman(1977), Freeman(1979), Freeman et al(1991), Brandes(2001), Newman(2003)], where the betweenness centrality is the most representative of this category. There are another families of centralities which are

[1]University of Alicante, Alicante, Spain.
[2]University of York, York, UK

Corresponding author:
Leandro Tortosa, University of Alicante. Ap.Correos 99, E-03080, Alicante, Spain.
Email: tortosa@ua.es

relevant for the study of basic network properties. One of them is a family of centralities which deal with the global efficiency of the networks, it includes such measures as efficiency, straightness and information centralities [Latora and Marchiori(2004)].

Another family of centralities measures the relative influence of the nodes in the network, including eigenvector centrality [Bonacich(1987), Bonacich(1991)], Katz centrality [Katz(1953)], and Page Rank based centralities [Page et al(1999), Berkhin(2005), Langville and Mayer(2005), Agryzkov et al(2012)] among others.

Urban spatial networks belong to a particular type of complex networks. They are most similar to the geometric network, given that spatial close nodes are more likely to be linked. They do not exhibit the small–world properties as they are planar and there are large topological distances between some nodes. In addition, urban networks are characterized by a fairly uniform degree distribution, with the usual degree variation between 1 and 5. Therefore the use of the graph constructed from the topology of the urban layout does not provide much interest studying the centrality based exclusively from the network geometry or topology, as is the case of the standard eigenvector centrality.

In this work a new model of analysis based on the calculation of a type of eigenvector centrality in urban spatial networks is proposed, which includes as a component the geo-referenced data. The aim of proposed centrality is to identify more and less influence areas of the city by taking into account its geo-located data. For that task the eigenvector centrality is the most appropriate. Besides, endow the proposed measure with a mechanism that allows us to weight the contribution of the network topology to the final node score.

The association of geo-referenced data to the graph breaks the uniform distribution of values, also it provides the additional information in addition to the topology for the analysis. Both elements are key to correctly identifying important areas in the network. The relationships of the data can only be properly understood in presence of the network topology. Consequently, the proposed centrality measure is an adaptation of the eigenvector centrality for the spatial networks, which in addition to preserving the characteristics of the original centrality, also includes in the computation process the geo-referenced data.

Related work

There is a large volume of published studies describing the role of complex networks or spatial analysis when we try to understand the current cities. In this section we briefly summarize the work developed by some researchers or working groups, related to the study of cities as complex networks, which we consider relevant.

Introducing the science of cities, Batty [Batty(2013)] suggests that to understand cities we must view them not simple as places in space but as systems of networks or flows. Last decades have seen the development of diverse analytic techniques for describing spatial layouts and their properties. One of them is space syntax, a set of techniques for the analysis of the spatial form developed by Bill Hillier and his colleagues at University College London (see [Hillier and Hanson(1984), Hillier(1999)]).

Space Syntax Analysis threat the urban space as dual graph where nodes are represented as straight lines of unobstructed pedestrian movement, and graph edges are defined as its intersections. The basic component of this dual graph is what they call axial space a straight sight-line and possible path. Axial lines are defined in the model as the fewest and longest lines of sight that can be drawn through the open street spaces of a study area [Hillier and Hanson(1984)]. The centrality measures proposed by Space Syntax depart from its global vision of urban space as the social space of human interactions and a physical space of building environment [Hillier and Hanson(1984), Hillier(1989), Hillier and Vaughan(2007)]. These centrality measures are mostly based on topological distances measured in terms of steps.

The applications of Space Syntax Analysis is numerous and cover a variety of research areas and applications in architecture, urban design and planning, transport, information technology, and many others. Some great specialists in the field of spatial analysis or urban modelling, as Bin Jiang or Michael Batty, combined space syntax with traditional transport networks models, using intersections as nodes and constructing visibility graphs to link them. In [Jiang and Claramunt(2002)] the authors present a data modelling process based on a combination of complex system theory and the object-oriented paradigm, producing an object-oriented spatio-temporal data model.

Some authors and research groups, such as S. Porta and City Form Lab, provide a different technique, using the direct approach to the spatial system by using the primal graph, where edges represent an urban streets and nodes are its intersections [Chambers(1988), Porta et al(2006), Sevtsuk(2012)]. This approach to the urban space preserve its topology and offers an opportunity to work with Euclidean distances, which are an important property for almost all spatial systems.

On the other hand, we want to highlight the work done by Andres Sevtsuk and his colleagues at the City Form Lab in Cambridge, MA. Among other projects they developed a toolbox called Urban Network Analysis (UNA) for ArcGIS, which is able to compute five types of graph analysis measures on spatial networks (Reach, Gravity, Betweenness, Closeness, and Straightness), as well as some other indices.

An original contribution of the urban analysis network tools implemented is that they include a third network element: buildings. The unit of analysis thus becomes a building, enabling the different graph indexes to be computed separately for each building. They are used as the spatial units of analysis for all measurements and, what may be very useful, they can be weighted according to some characteristics, (see [Sevtsuk(2012)]).

A number of authors has shown that network analysis measures can be useful predictors of some urban phenomena, as for example in the distribution of retail and service establishments in urban environments [Porta et al(2009), Porta et al(2012), Sevtsuk(2014)]. It is interesting the work developed by Sevtsuk [Sevtsuk(2010)], analysing retail location patterns in urban settings and investigating whether and how the distribution of retailers is affected by the spatial configuration of the built environment. In our case, we are not considering as a priority the built environment, since we are focusing the work in the developing of a centrality measure that acts directly over the data.

In [Porta et al(2009), Porta et al(2012)] the authors examine the relationship between street centrality and densities of commercial and services activities in the cities of Bologna and Barcelona. The centrality measures they use are classical ones as closeness, betweenness and straightness. They show a high correlation between areas with high centrality values and high commercial density. We analyse a similar question but with a quite different point of view. Their starting point is the network topology, while our starting point is a set of retail and services data in the city, applying the eigenvector centrality proposed to this set of data.

Urban Street Networks and Geo–located Data

In an urban network, we have a great amount of information; much of it has a geographic location, allowing us to perform their representation in the urban fabric itself. However, we must be able to quantify and distribute this information in the graph, which is the geometric representation of the urban network. This is the basic issue that is discussed in this section. Thus, the issue we ask ourselves is how to perform the assignment and quantification of the city information in the graph constructed from the urban fabric.

Data Organization in Urban Networks

When working with urban networks, there are a number of different ways to represent the topology of the city [Crucitti et al(2006b), Crucitti et al(2006c)]. In this case, we opted for a classic representation of cities through the concept of *primal graph* [Crucitti et al(2006b)]. The graph is a pair $G = (V, E)$ with nodes $v \in V$ representing street intersections and edges $e = (u, v)$ representing connectivity between intersections u and v. Additionally, the nodes have geometric information associated with them via a spatial position $\mathbf{x}(v)$ for each intersection. This graphical representation of cities generated a particular type of graph, an undirected plane graph. This type of network has a number of similarities to the random geometric graph (RRG), although the degrees are much more uniform. The structure of the street network means that the degrees are typically between 1 and 5.

The topology of the primal graph represents the layout of the city itself, but the geometric positions allow us to associate position data about the city with the graph. Information can be extracted from many different sources, many of which also provide geolocation information. This data may be distributed over the primal graph. For example, consider a data item with position $\mathbf{y}$ and value d.

Consider the case where we have a number of different data categories, indexed by $j = 1 \ldots l$. Each sample from the source is associated to a node. Then we have a *data matrix* $\mathbf{D}$ with entries d_{ij} representing the value of category j associated with node i. If, as an example, we want to analyse the number and location of the different restaurants in a city, we can establish the characteristic j as *restaurants*. Then, the column $\mathbf{d}_j$ reflects the number of restaurants associated to each of the nodes of the network. The element $d_{ij} \in D$ represents the number of restaurants that are located in the proximity of the node n_i.

A real set of data extracted from a city

Since we take the city of Murcia as an example of a real urban network, we expose some characteristics of it to understand the urban environment under study. The city of

Murcia is located in the south eastern area of Spain and the urban centre of the city originates from the IX century. Since then, the city has had important stages of territorial expansion.

We create the network (primal graph) from a connected graph where the streets become undirected edges. Nodes usually represent the intersections of the streets, but we can also assign nodes to some points of interest in long streets. The primal graph allows us either to represent the topology of an urban fabric as well as to organize the geo-located data. The network is composed of 1196 nodes and 1867 edges (see Figure 1 for a graphical display of the topology of the urban fabric and the urban network constructed).

For this example, we will work only with a part of the city, the historical centre and the neighbourhoods that are placed around it. The reason that motivates this limitation lies, on the one hand, in reducing the amount of data to work with and, on the other hand, because the historical centre is the most active area of the city and where most activity takes place.

We should note that the example described in the previous section constitutes a small portion of the network that we are now studying as it is the city of Murcia.

The data collection used for this example starts with a fieldwork that consists of collecting the data or information from visual inspection or pictures. These data were assigned to the nodes of the network so that each node has a set of numerical values associated with the information that is being studied. We collected data about existing facilities and commercial activity. In the analysis we perform, we distinguish the following types of facilities:

- Type I: Bars, restaurants, coffee, snack bar, ...
- Type II: Shops with an area less than 300 square meters.
- Type III: Sales offices and bank offices.
- Type IV: Big shops (department stores, shopping centres, ...) with an area greater than 300 square meters.

The number of tertiary facilities that have been collected through fieldwork can be summarized in the following, taking into account the established categories.

- Type I: 552 venues.
- Type II: 2216 venues.
- Type III: 285 venues.
- Type IV: 33 venues.

Figure 2 displays the map of the city where we have geo-located the tertiary facilities collected from the fieldwork.

Summarizing, we can say that we have, approximately $2,760$ data associated to the nodes of the network and the maximum value of data associated to a node is 55. With this set of data, we can construct a data matrix $\mathbf{D}$ that is able to summarize all the information of the data we have in an organized form. So, we can define $\mathbf{D}$ as

$$\mathbf{D} = \begin{array}{cc} & \overbrace{\begin{array}{cccc} d_1 & d_2 & d_3 & d_4 \end{array}} \\ \begin{array}{c} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \\ \vdots \\ n_i \\ \vdots \\ n_{1196} \end{array} & \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} \\ d_{11} & d_{22} & d_{23} & d_{24} \\ d_{31} & d_{32} & d_{33} & d_{34} \\ d_{41} & d_{42} & d_{43} & d_{44} \\ d_{51} & d_{52} & d_{53} & d_{54} \\ \vdots & & & \\ d_{i1} & d_{i2} & d_{i3} & d_{i4} \\ \vdots & & & \\ d_{1196\ 1} & d_{1196\ 2} & d_{1196\ 3} & d_{1196\ 4} \end{bmatrix} \end{array}. \tag{1}$$

Thus, the matrix $\mathbf{D}$ given by (1) has 1196 rows, corresponding to the 1196 nodes of the urban network studied, and has 4 columns, each corresponding to the four different types of data that were obtained. This means that in the column d_1 we have all the Type I data (food-service sector) associated to each of the nodes. In the column d_2 we have all the Type II data (shops) associated to the different nodes. And so on with the columns d_3 and d_4 for the Type III and Type IV data, respectively.

The arrangement of the data in a matrix form like this one has the advantage that separate and organize the data, according to its category, so we can analyse them together or separately, as we shall see in the numerical results.

As already mentioned in the introduction, the classic centrality measures do not allow us, in a simple way, to work with the data associated with a network. Therefore, it becomes necessary to have centrality measures which take account of two factors, first, the network topology and, moreover, the importance of existing data.

The classical eigenvector centrality measure

Eigenvector centrality, denoted by c_E, was proposed by Bonacich [Bonacich(1982)] to measure the influence of a node in a network from the importance of its connections.

Degree centrality gives an idea about the number of connections a vector has. However, not all the connections or links are equally important. Therefore, somehow we should weight the importance of each node connections. If we assume the idea that a node is more central if it is in relation with nodes that are themselves central, we can argue that the

Figure 1. Map and network of the city area studied.

Figure 2. Map of the geo-located tertiary venues collected in the city area studied.

centrality of the nodes of a graph does not only depend on the quantity of its adjacent nodes, but also on their value of centrality.

If we denote the centrality of node n_i by x_i, then we can take into account the importance of each node links by making x_i proportional to the average of the centralities of i's network neighbours:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j, \tag{2}$$

where λ is a constant.

Defining the vector of centralities $\vec{x} = (x_1, x_2, \ldots)$, we can rewrite (2) in matrix form as

$$\mathbf{A} \cdot \vec{x} = \lambda \vec{x}. \tag{3}$$

It is clear from the expression (3) that $\vec{x}$ is an eigenvector of the adjacency matrix $\mathbf{A}$ associated to the eigenvalue λ. As $\mathbf{A}$ is the adjacency matrix of an undirected graph and A is non negative, it can be shown (using the Perron-Frobenius theorem) that there exists an eigenvector of the maximal eigenvalue (we denote it by λ_1)with only non negative (positive) entries. This eigenvector constitutes a classification of the nodes in the graph.

The data centrality for urban networks

The eigenvector centrality in its classical form described in Section 3 it only takes into account the topology of the network and the importance of the neighbouring nodes. It does not incorporate any notion of the spatial information which may be present on a urban network. Although the eigenvector centrality can straightforwardly accommodate weighted edges (and hence edge information) by using a weighted adjacency matrix, care must be taken in how to encode the urban spatial data into a centrality index.

Consequently, the main objective in this section is to construct a centrality map for the data collected in matrix $\mathbf{D}$ based on the concept of eigenvector centrality.

The relative importance of different data categories may vary according to the problem under study and if we employ edge weights, we can only incorporate one value on each edge. We therefore combine the data categories into a single measurement for each node. The construction of a data matrix as above allows us to control, in a simple manner, the importance we assign to each feature $\mathbf{d}_j$ being measured. It is enough just to define a vector $\vec{v}_0$ of size $l \times 1$, that we are going to call *weight vector*, which values are in the range $[0, 1]$. The function of this vector is to establish the importance that we consider to each of the characteristics measured in the matrix D. For example, a vector with all components equal to 1 would mean that we consider all equally important features. Then the *data vector* is

$$\vec{v} = \mathbf{D}\vec{v}.$$

Our goal is combine this data vector with the topological information provided by the city network. The obvious choice is to weight the edges of the network according to the data on the nodes. If we are to combine the information from the data vector and the edges, then they must be of the same order otherwise one type of information will swamp the effect of the other. Since we want edge weights of $O(1)$, we normalize the data vector as follows

$$\tilde{\vec{v}} = \frac{1}{\max_i v_i} \vec{v},$$

so the largest component is 1. According with the idea of eigenvector centrality, a node n_i is important if its neighbours are important. If we want to link this idea with the influence of data, we should say that node n_i is important if node n_j is important and they are linked by a street that is supported by a large amount of data, that is, large values of $\vec{v}$ at each node incident on the edge. Following this reasoning, we can establish the importance of an edge in the graph from the data associated with the two end points. So, if and edge e has n_i and n_j as the two end points, and the data associated to these nodes are $\tilde{v}(i)$ and $\tilde{v}(j)$, respectively, we can establish the *importance* of the edge e as

$$w_{ij} = \tilde{v}(i) + \tilde{v}(j). \tag{4}$$

This provides us with the weight matrix $\mathbf{W}$ for the data.

Looking at the primal graph with data shown in Figure 2, draws our attention a very common fact when working with urban networks, where data from many different sources and characteristics are collected. Looking at the graph it is noted that the value of the data associated with some nodes may be zero. In other words, in some areas no facilities are located and the corresponding entry in the weight matrix is zero.

We can solve this drawback by introducing what we can call as a basic level of importance associated to all the edges. All edges have a small level of importance , even if no facilities have been identified. Intuitively, this means that areas are still linked, even if no particular facility exists on a street. We denote this basic level of importance as α and defined it as the smallest non-zero level of importance assigned by the data:

$$\alpha = \min\left(w_{ij}\right)_{w_{ij}\neq 0}. \qquad (5)$$

The idea of introducing a basic minimum level of importance associated with the edges is also in agreement with the idea of the own centrality, where the importance of a node is given by the importance of the neighboring nodes. A node with no data is always influenced, albeit minimally, by the data of the nearest nodes, even if they are not directly connected to it.

Finally, traditional eigenvector centrality tends to lead to solutions with rapid variations between nodes. We do not view this as natural for urban networks. Rather, we are looking for more diffuse centres of activity. This is consistent with the idea that existing geographic information linked to a node results influential both the node itself and its neighbouring nodes.

When working with urban networks of medium or large size, it is appropriate to introduce some mathematical technique that smooth out the solutions, so that sudden changes in the components of the vector solution are minimized. In our case, the objective is to smooth the solutions that are obtained by the dominant eigenvector of the matrix A^* in the centrality algorithm. A common technique in optimization problems that acts as a regularize of the solution consists in the introduction of a matrix of very small values $\epsilon\mathbf{J}$ (where $\mathbf{J}$ is the matrix of all-ones). This matrix is a quantity that represents, a small influence of between all nodes regardless of the network connections.

We define ϵ as a new parameter which is based in the parameter α given by (5). If α represents the basic level of importance assigned to every node by the data, we can establish that ϵ is a small percentage of the value computed for α. Various experimental tests with different networks of various sizes have led us to the conclusion that the parameter ϵ can be established as

$$\epsilon < \frac{1}{10}\alpha. \qquad (6)$$

There may be other ways to define the parameter ϵ; however, numerical tests with different amounts of data in real urban networks have given us expected and consistent results when ϵ is defined as (6). In the numerical examples we will see the effect on the calculations when we take $\epsilon = 0$. The fact of not smoothing the solutions produces that they are much more discontinuous, which allows us to determine zones or areas in the network that act like a kind of hubs or attractors with respect to the data involved. This is evidenced when analyzing data on the commercial activity, where we clearly see the difference between $\epsilon = 0$ and $\epsilon \neq 0$.

With these ingredients, we can now construct the data adjacency matrix $\mathbf{A}^*$ from the graph adjacency matrix $\mathbf{A}$:

$$\mathbf{A}^* = \mathbf{A} \circ (\mathbf{W} + \alpha\mathbf{J}) + \epsilon\mathbf{J} \qquad (7)$$

where $\mathbf{J}$ is the matrix of all ones and $\circ$ is the element-wise multiplication operation. Note that this final term is reminiscent of the teleportation matrix of the PageRank algorithm, except that we determine ϵ from the data. In effect, this term is a regularization to avoid localized solutions.

Our final step is to compute the centrality measure of the nodes as

$$\vec{c} = \frac{1}{\lambda_1}\left[\mathbf{A}\vec{x_1}(j) + \vec{x_1}\right], \qquad (8)$$

where $\vec{c}$ constitutes the centrality values for the nodes of the graph, and $\mathbf{A}$ is the adjacency of the original urban network.

Note that in the expression (2) $\vec{x_1}$ is the principal eigenvector of $\mathbf{A}^*$ and the components represent the traditional eigenvector centrality of eigenvector centrality of $\mathbf{A}^*$. To this we add $A\vec{x_1}$ which spreads the importance of neighbouring nodes in the network. This results in a smoother and more diffuse solution.

The following algorithm summarizes the steps we can follow to obtain a new eigenvector centrality measure from the data network.

Algorithm 1. *Eigendata centrality.*
Input: $\mathbf{A}, \mathbf{D}, \vec{v}_0$.
Output: $\vec{c}$
1: Construct the data vector $\vec{v} = D \cdot \vec{v}_0$.
2. Normalization of $\vec{v}$.

$$\tilde{\vec{v}} = \frac{1}{\max_i v_i}\vec{v}.$$

3. Construct the weight matrix $\mathbf{W}$ *as*

$$w_{ij} = \tilde{v}(i) + \tilde{v}(j)$$

4. Compute α *using the expression (5).*
5. Take ϵ*, according to the expression (6).*
6. From $\mathbf{A}$, $\mathbf{W}$, *and* α, ϵ *construct* A^* *as*

$$\mathbf{A}^* = \mathbf{A} \circ (\mathbf{W} + \alpha\mathbf{J}) + \epsilon\mathbf{J}.$$

7. Compute the dominant eigenpair of $\mathbf{A}^*$, $(\lambda_1, \vec{x_1})$.
8. From A *and* $\vec{x_1}$ *compute*

$$\mathbf{c} = \frac{1}{\lambda_1}\left[\mathbf{A}\vec{x_1} + \vec{x_1}\right].$$

Numerical results

In this section we will carry out some numerical experiments with networks of different sizes and using several data sets.

An example with a small network

We started the numerical section with an example of a very small graph composed only of 10 nodes. The idea is to check on a small graph the influence of data, no matter how small the network is, when the Algorithm 1 runs.

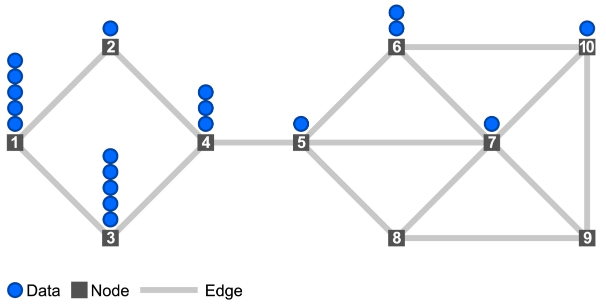

Figure 3. A small graph with 10 nodes.

Figure 3 shows an example network for the data vector $v = [5, 1, 5, 3, 1, 2, 1, 0, 0, 1]^T$.

Running all the steps described in Algorithm 1, we arrive to obtain a centrality value for each node. This vector is

$$\mathbf{c} = [0.2780, 0.2925, 0.2925, 0.3257, 0.1754, 0.0553, 0.0499, 0.0256, 0.0101, 0.0231]^T.$$

It is clear that the most important node, which has a larger value of its centrality (0.3257) is the node n_4, while nodes n_2 and n_3 occupy the second place in the ranking. It is observed as the presence of more data on the left side of the graph makes the centrality values of these nodes are much greater than the others. Remark that the graph presents a higher connectivity in its right part; it is clear that if we did not consider the influence of the data, the centrality of the nodes in the right connected component would be much higher. We will compare these results with those provided by the classical eigenvector centrality, in order to see the importance of introducing the data in the measurement.

We try some computations with different data vectors, as it is reflected in Table 1. The first column shows the ID of each node; the second column shows the degree of each node, while in the third column we have the values of the eigenvector centrality eig for each of the nodes, which does not depend on the data associated to the nodes of the graph. This centrality index is independent of the data vector we choose to perform the numerical experiments, so they are fixed. In the rest of the columns we show the eigenvector centrality $eigdata$ for different data vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$.

We have represented in Figure 4 the values of the classic eigenvector centrality c_E and the $eigdata$ centrality obtained according to Algorithm 1, for the four data vectors, $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ that we have in Table 1. Typological characteristics of this graph means that there is a difference of connectivity between their two distinct components formed by the nodes n_1 to n_4 and the rest. When establishing the data vectors have considered appropriate to consider cases in which most data focus on the most disadvantaged component, from the point of view of connectivity. Therefore, in several data vectors we have established the large volume of data between nodes n_1 to n_4.

Analysing the results shown in Table 1 and Figure 4, some points may be highlighted.

- The values of the eigenvector centrality are quite expected, since the most central node is n_7, which is the one who has greater connectivity. The same also applies to the second node in importance (n_5), that has connectivity degree 4. Really, the topology of the network is the crucial factor in the measurement. Note in this case the rapid changes occurring in the value of the centrality between a node with its immediate neighbours. Furthermore, there is a difference in values between the nodes who form the two components of the graph; obviously the lack of connectivity of the nodes n_1 to n_3 is reflected in very low values of its centrality.
- We pay attention to the case of vector v_3. We clearly observe in the graphic values of the centrality absolutely uniform for all the nodes and very low compared to other data vectors. This case is quite strange since it has the characteristic that all data values of the nodes are equal and balanced ($v_3(i) = 2$, for all i). This makes the α value obtained is not the most appropriate for this measure, so there is little variation between the centrality of the different nodes, if we compare with other vectors. This case produces virtually never in large complex urban networks, where differences in the values of the data between different nodes are usually much larger and there is never this feature of equality and uniformity.
- In the case of vector v_2 it is absolutely clear the effects of the concentration of the volume of data at nodes n_1 and n_3. The values of the centrality of these nodes are clearly well above the rest, causing a sudden drop in the nodes as from n_5. Note that the vector v_1 also produced this feature that the initial nodes (component

Table 1. Numerical values of the data eigenvector centrality $eigdata$, for different data vectors.

Node	degree	eig	v_1 data	$eigdata$	v_2 data	$eigdata$	v_3 data	$eigdata$	v_4 data	$eigdata$
1	2	0.0278	5	0.2780	15	0.3500	2	0.0069	10	0.2408
2	2	0.0487	1	0.2925	1	0.3578	2	0.0122	10	0.2390
3	2	0.0487	5	0.2925	15	0.3578	2	0.0122	10	0.2390
4	3	0.1427	3	0.3257	3	0.3662	2	0.0357	2	0.2644
5	4	0.4028	1	0.1754	1	0.1319	2	0.1007	2	0.1213
6	3	0.3689	2	0.0692	2	0.0179	2	0.0922	2	0.0363
7	5	0.5306	1	0.0813	1	0.0192	2	0.1327	2	0.0454
8	3	0.3689	0	0.0654	0	0.0177	2	0.0922	2	0.0363
9	3	0.3593	0	0.0290	0	0.0026	2	0.0898	2	0.0198
10	3	0.3593	1	0.0340	1	0.0033	2	0.0898	2	0.0198

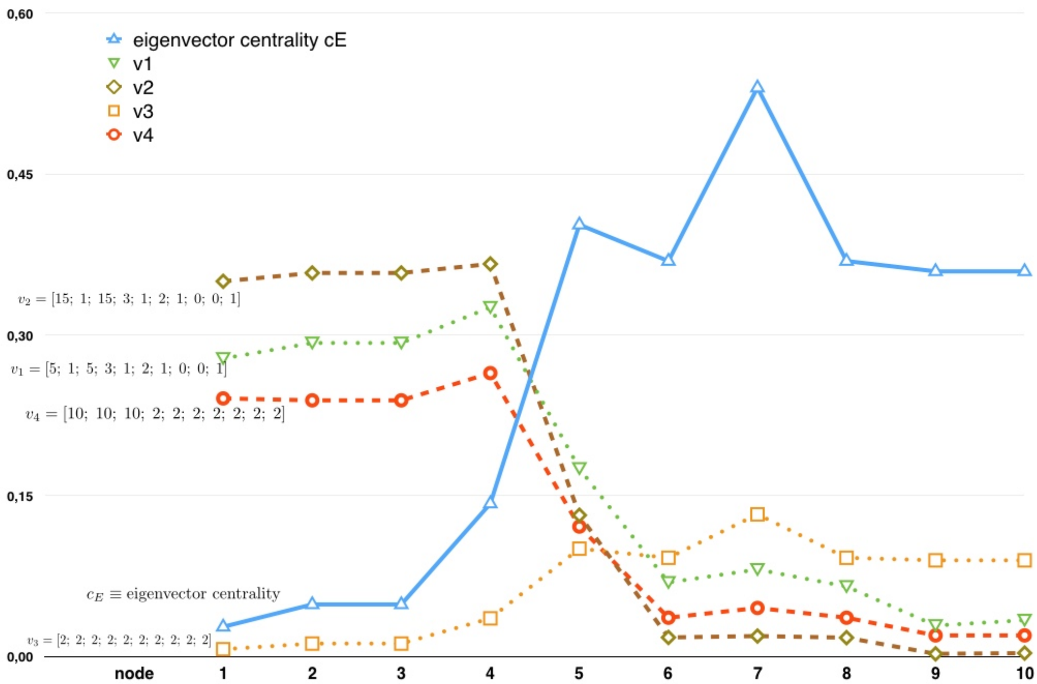

Figure 4. Graphical representation of the centrality $eigdata$, for different values of the data vector, $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$.

on the left side of the graph) have more data than the rest. The difference is that the data volume is not large enough to its influence on the centrality is decisive.

An example with a bigger network

In this section we present an example of a bigger network than before, with the aim to perform a comparison between the original eigenvector centrality and the proposed centrality, to understand the differences between them. Both measures will be applied to a real spatial network and dataset extracted from the city of Murcia (Spain).

We take a part of the city of Murcia where we have used the dataset described in Section . In figure 1 we have marked in red the specific area of the city that comprises the urban network that we study in this section. The data are related with the quantity of commercial facilities (retails, food-service, leisure venues, big shops) detected in the urban layout. The urban network shown in the images has 267 nodes, 361 edges and 775 geo-located commercial facilities.

In Figure 5, we have represented the network studied in this analysis, along with the total number of retail and services establishments associated to each of the nodes. These allocations are related to commercial activity in the city that were collected from the field work carried out in 2013.

Figure 6(a) shows the node degree distribution of the selected network, which is the main factor that affects the result of original measure of eigenvalue centrality. Figure 6(b) shows the density of data distribution in the

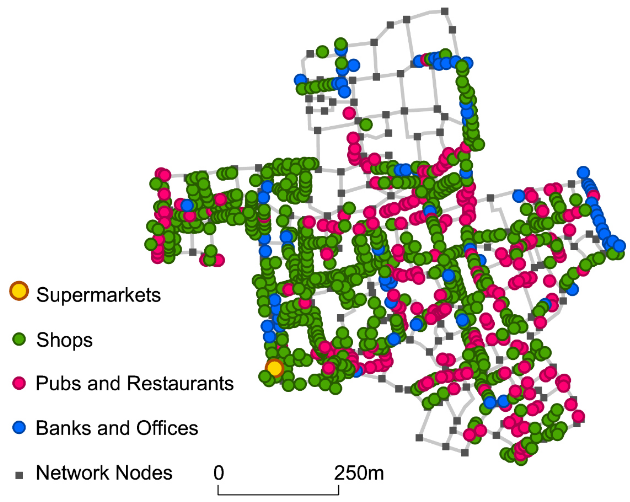

Figure 5. Number of retail and services data registered in the part of the city of Murcia studied.

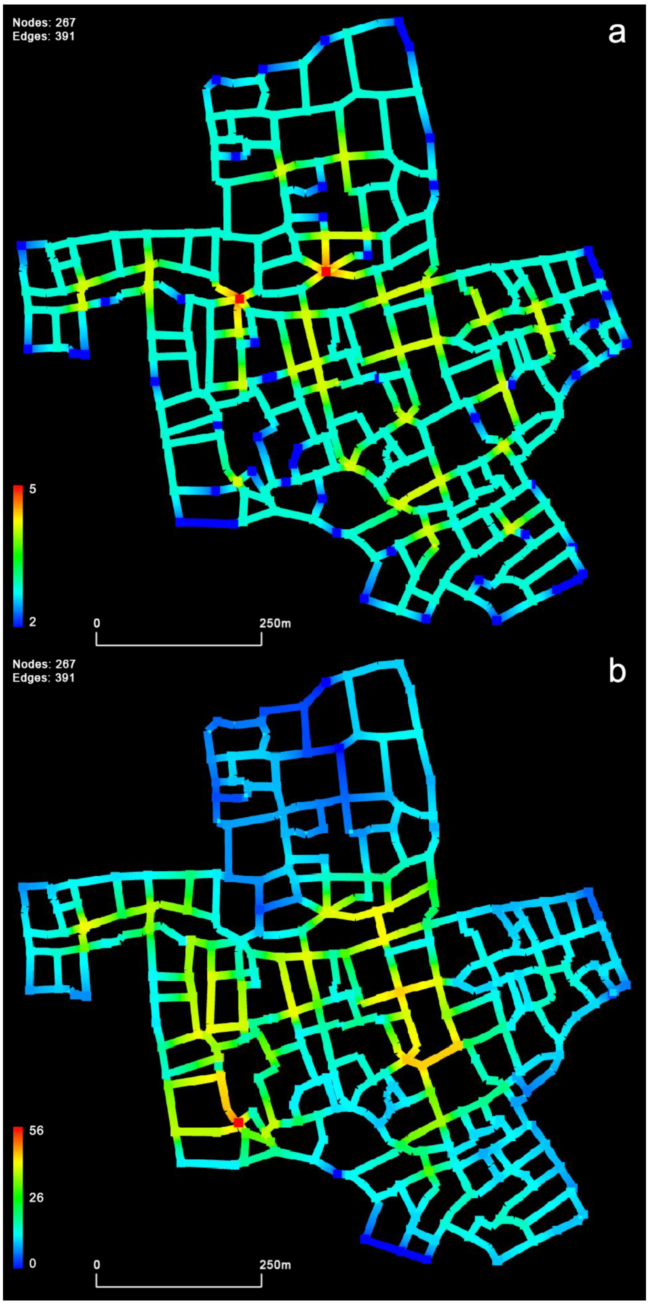

Figure 6. The data and degree distribution.

selected area of the network, where we can observe a high data concentration in the lower left corner, the place where an important urban square is located.

Figure 7(a) shows the fairly expected result of the original eigenvector centrality. From this image we can appreciate that the node with the maximum ranking is, as expected, a central node with the maximum degree value. Figure 7(b) shows the result of the proposed eigenvector centrality measure, in which both the topology and the data density are involved. We can observe that the node with the maximum ranking is located in the area of the network which is characterized by high density of data and degree distribution.

An example with a real urban network

In this example we will use the data described in Section 2.2 concerning the city of Murcia regarding the commercial activity described in detail in this section. We are going to work in this example with data of Type I and Type II, that is, the most numerous, since we want to study the city from the point of view of the commercial activity related to the shopping and food-service sectors. The total amount of data is $2,760$ while the maximum value of data associated to a node is 55.

Therefore, when we proceed to normalize the data vector of this network. We have that

$$\vec{v}^N = \left(\frac{v(j)}{55}\right)_{j=1}^{1196}.$$

For this example, we construct the matrix W, and compute α, that is

$$\alpha = \min\left(w_{ij}\right)_{w_{ij}\neq 0} = \frac{1}{55}.$$

Now, we compute the parameter ϵ. Following expression 6, we have that

$$\epsilon < \frac{1}{10}\alpha = 0.0018.$$

Having computed the parameters α and ϵ, we are ready to run Algorithm 1. The result of the centrality map obtained for this urban network is shown in Figure 8. In this figure, it is clearly revealed a main central area corresponding to an arterial axis of the city, running from north to south and containing the most central nodes in the whole network. This arterial axis is a famous commercial street where it is placed on the north (close to the most central node in red colour) the most important department store in the city.

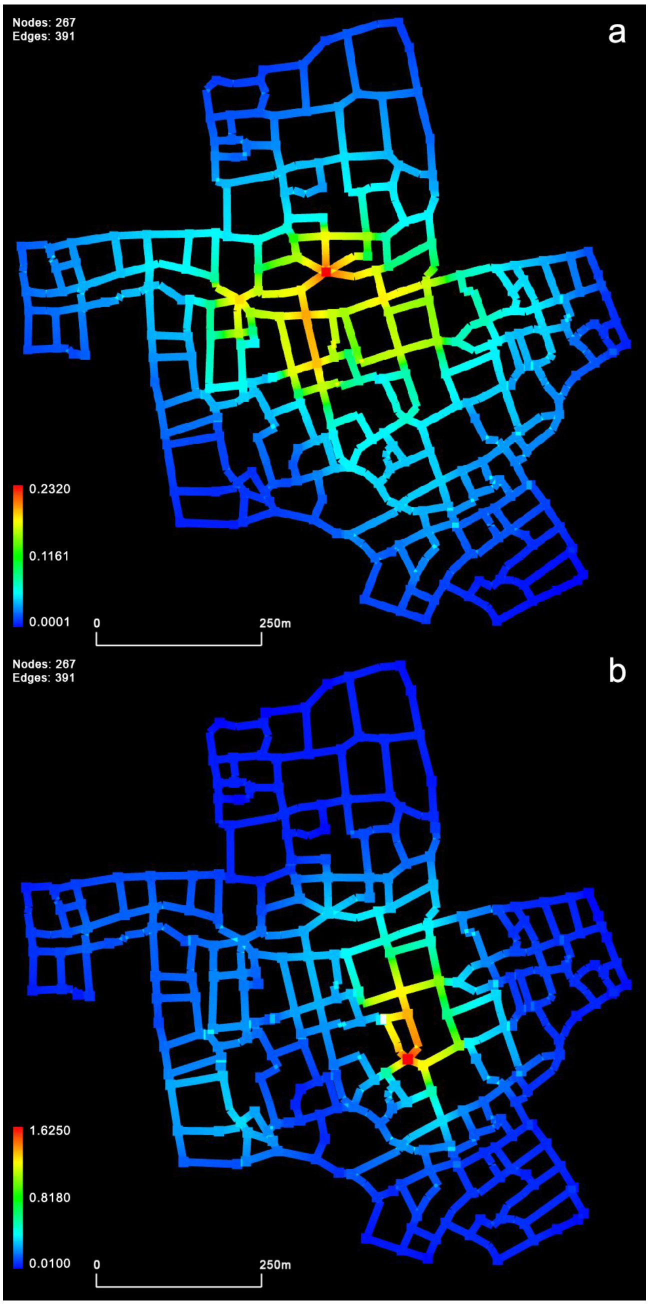

Figure 7. Eigenvector centrality and Algorithm 1 applied over the selected area of the network.

We can understand the parameter ϵ as a parameter that helps us in addition to take account of existing global data in the city, to smooth solutions and get a better visualization of the results. Notice what happens when the parameter ϵ is zero (see Figure 9).

When comparing the images, we can not say that the analysis that displays a picture invalidate the analysis offered by the other picture. The difference is in the display. Since we do interpolation is linear in the graph, most of the values are displayed on a bluish hue, except those with the greatest centrality. When we introduce the parameter and get smooth solutions, the network is best viewed, since the difference in centralities does not cause a sudden change in colour.

In Figure 9, where we do not take into account the whole information present in the network, we see from the information analysed that the entire urban network is reduced to a small hub where nodes with high centrality are concentrated. If we wish to reduce all the information analysed at a certain point or reduced area, this measure of centrality is the most appropriate for their characteristics. However, as shown in Figure 8, we introduce the parameter ϵ, the centrality map that is exposed is much more faithful to the reality of the urban network. The values of the centrality have been smoothed.

We will briefly comment on the usefulness of using such measures in urban networks analysis. The application of this measure of centrality in this city allows us to determine those areas of the network that present a greater commercial activity compared to more disadvantaged ones, as can be clearly seen in the images. If we think that centrality may be considered in some sense as a fairly good indicator of *power* in the network, we can see in Figure 8 where is located the most influential commercial area. In this case, with the parameter $\epsilon = 0$, we determine a hub in the city, a *commercial activity power*, which corresponds to the main commercial artery of the city. It is important to highlight that the highest values of centrality are concentrated in the main artery of the city, called *Gran Vía*, which turns out to be interesting since this avenue is characterised by the greatest concentration of outstanding multinational business establishments related to the sale of textile products (especially in its southern part).

It is also noteworthy other aspect that characterise this main commercial street; in this area we find the highest land market values, as well as the highest rental prices. Then, centrality captures the essence of location advantage in this urban area, and its value is reflected in the intensity of land uses and densities of economic activities.

Therefore, we can say that the results of the performed analysis identify hubs of the very influential material goods located at the studied area of the city, by taking into account the topology of the urban layout and the geo-located data density.

The eigenvector centrality offers us certain aspect of centrality that is not captured by other measures. This conception of importance or centrality makes sense in different circumstances, for instance in urban environments, where the importance of a node is given, in a way, by the importance of its neighbouring nodes. If we talk about retail and services establishments (commercial sector), we see that

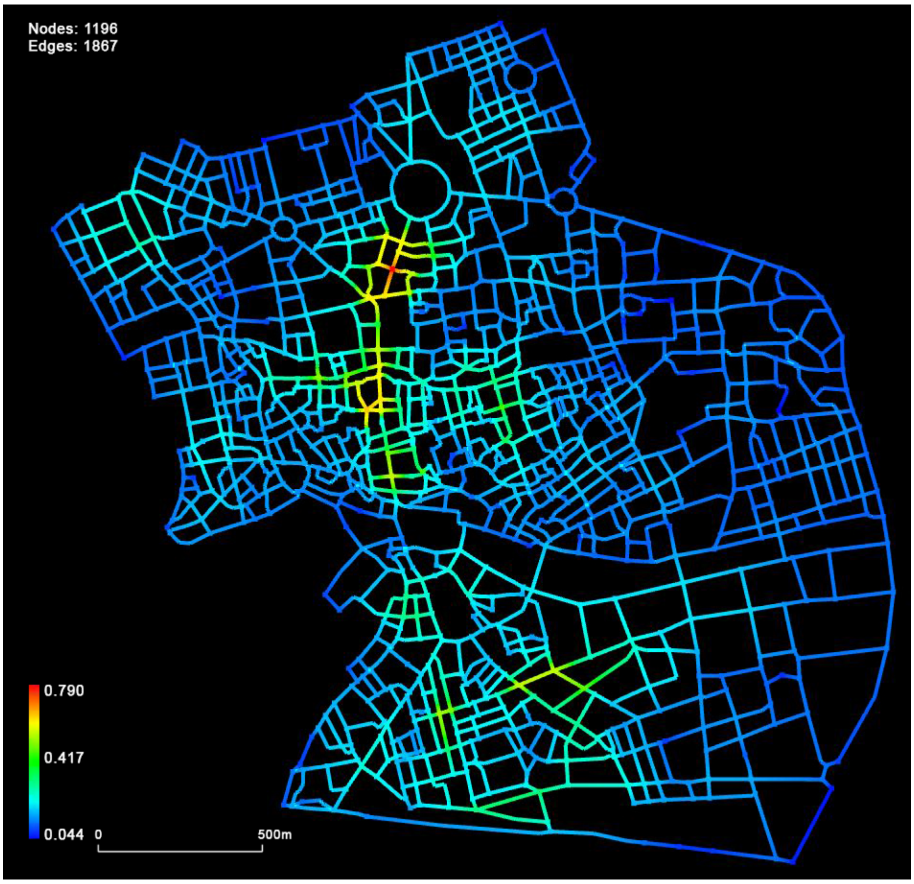

Figure 8. Centrality map of the city area studied, for $\epsilon = 0.0018$.

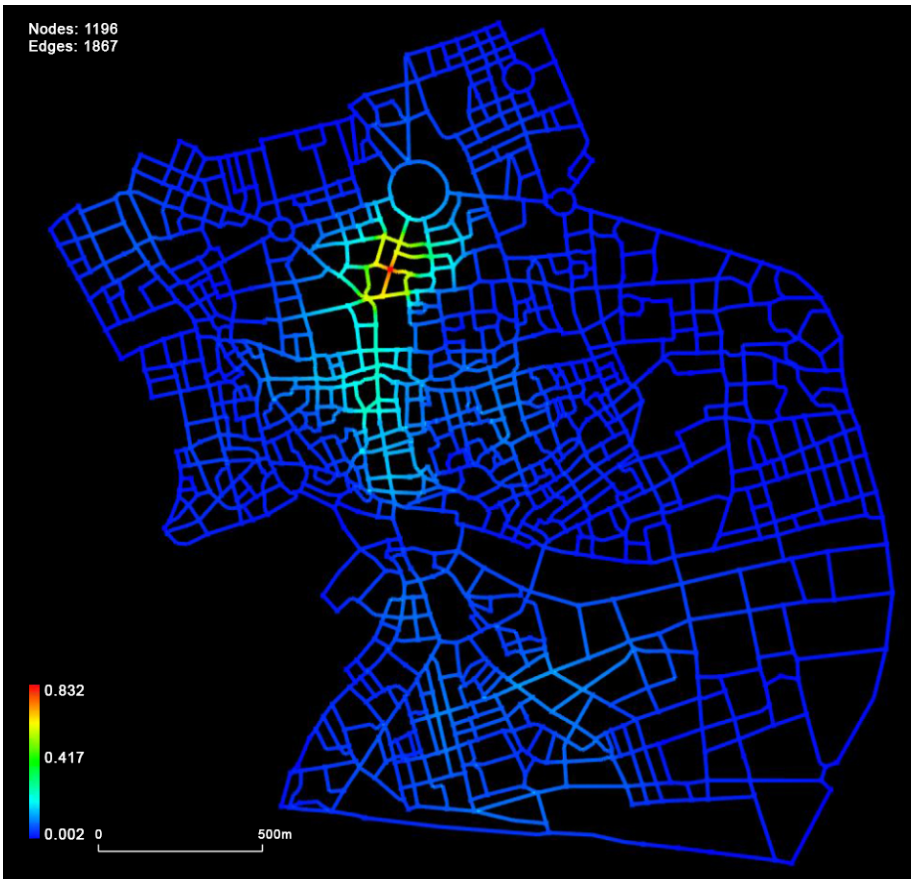

Figure 9. Centrality map of the city area studied for $\epsilon = 0$.

the value of a retail is given not only by the topological value of its location within the network, but also by the amount and importance of the retails and services establishments around it.

Regarding to the prediction capabilities of the model, we must say that we can easily modify the data associated to each node. Therefore, it is possible to simulate and visualize changes in the urban network where we can add or remove all the data that we consider appropriate. Moreover, we can simulate urban plans before they are developed in order to evaluate the impact that the extensions of the network cause on the neighbouring of it by means of the centrality measure. In other words, it is possible to introduce modifications in the commercial layout or modifications in the urban network topology, with the aim to evaluate the effect of such modifications over the whole network. Among other applications we can include those related to distribution of land uses, retail influence or rents.

Conclusions

In this paper we discuss the problem of how to locate the key areas of activity in the urban infrastructure of a city by using a centrality measure. We propose a new measure of centrality for these types of augmented urban networks, which is based on the concept of eigenvector centrality, and it is able to measure the influence of the topology of the network and the geo-referenced data extracted from the network and associated to the nodes. The motivation to analyse spatial data in the city has led us to develop this model of analysis based on the calculation of the eigenvector centrality in urban street networks, with the primary characteristic that it takes into account the component of geo-located data. The main contribution of the proposed model is the incorporation of the geo-located data factor to the computation structure for eigenvector centrality in the urban street networks. The data associated to the network provide additional features to the topological properties of the nodes, it allows to quantify and qualify the information located in their environments. In other words, the proposed centrality measure identifies the node topological importance within the urban street network according to their location and the amount of geo-referenced data associated. Through several examples shown in this paper it has been tested the centrality measure over networks with different sizes. In the example studied for a network of small size, it is evident the influence of the data on the calculation of the centrality measure for each node. For a big urban network we have seen how we can determine those areas or ”central” nodes of the urban network according to the commercial data that we are handling, taking into account that now the importance of a node is also related to the amount and importance of endowments and services of its own and its neighbours. In other words, the model proposed allows us to understand the distribution and relationships of retail and service establishments in this particular urban environment.

Funding

This work was partially supported by Spanish Govern, Ministerio de Economía y Competividad, the reference number of which is TIN2014-53855-P.

References

[Agryzkov et al(2012)] Agryzkov T, Oliver J L, Tortosa L, Vicent J, 2012, ”An algorithm for ranking the nodes of an urban network based on concept of PageRank vector” *Applied Mathematics and Computation* **219** 2186–2193

[Batty(2013)] Batty M, 2013, *The New Science of Cities* (Cambridge University Press, Cambridge, MA, USA)

[Behhisch and Ultsch(2007)] Behhisch M, Ultsch A, 2007, ”Urban Data Mining Using Emerging Self-Organizing Maps”, in *Data Analysis, Machine Learning and Applications* Eds C Preisach, H Burkhardt, L Schmidt-Thieme, R Decker (Springer, Heidelberg) pp 311–318

[Berkhin(2005)] Berkhin P, 2005, ”A survey on PageRank computing” *Internet Mathematics* **2**(1) 73–120

[Bonacich(1982)] Bonacich P, 1982, ”Factoring and weighting approaches to clique identification” *Journal of Matematical Sociology* **92**(1) 113–120

[Bonacich(1987)] Bonacich P, 1987, ”Power and centrality: a family of measures” *American Journal of Sociology* **92** 1170–1182

[Bonacich(1991)] Bonacich P, 1991 ”Simultaneous group and individual centrality” *Social Networks* **13** 155–168

[Borgatti,(2005)] Borgatti S P, 2005, ”Centrality and network flow” *Social Networks* **27** 55–71

[Bradlow et al(2005)] Bradlow E T, Bronnenberg B, Russell G J, Arora N, Bell D R, Duvuri S D, Hofstede F T, Sismeiro C, Thomadsen R, less S Y, 2005, ”Spatial Models in Marketing” *Marketing Letters* **16**(3) 267–278

[Brandes(2001)] Brandes U, 2001, ”A faster algorithm for betweenness centrality” *Journal of Mathematical Sociology* **25**(2) 163–177

[Chambers(1988)] Chambers R G, 1988, *Applied production analysis: a dual approach* (Cambridge University Press, Cambridge, UK)

[Crucitti et al(2006a)] Crucitti P, Latora V, Porta S, 2006, ”Centrality measures in spatial networks of urban streets” *Physical Review E* **73** 036125

[Crucitti et al(2006b)] Crucitti P, Latora V, Porta S, 2006, ”The network analysis of urban streets: a primal approach” *Environment and Planning B: Planning and Design* **33**(5) 705–725

[Crucitti et al(2006c)] Crucitti P, Latora V, Porta S, 2006, ”The network analysis of urban streets: a dual approach” *Physica A: Statistical Mechanics and its Applications* **369**(2) 853–866

[Fischer and Wang(2011)] Fischer M, Wang J, 2011, *Spatial Data Analysis. Models, Methods and Techniques* Springer Briefs in Regional Science (Springer, New York, USA)

[Freeman(1977)] Freeman L C, 1977, ”A set of measures of centrality based on betweenness” *Sociometry* **40**(1) 35–41

[Freeman(1979)] Freeman L C, 1979, ”Centrality in social networks: Conceptual clarification” *Social Networks* **1**(3) 215–239

[Freeman et al(1991)] Freeman L C, Borgatti S P, White D R, 1991, ”Centrality in valued graphs: a measure of betweenness based on network flow” *Social Networks* **13** 141–154

[Gliquet(2002)] Gliquet G, 2002, *Geomarketing. Methods and Strategies in Spatial Marketing* Geographical Information Series (ISTE Ltd, London)

[Haining(2003)] Haining R, 2003, *Spatial Data Analysis. Theory and Practice* (Cambridge Univertsity Press, Cambridge)

[Hillier and Hanson(1984)] Hillier B, and Hanson J, 1984, *The Social Logic of Space* (Cambridge University Press, Cambridge, MA, USA)

[Hillier(1989)] Hillier B, 1989, ”The Architecture of the urban object” *Ekistics* **334** 5–21.

[Hillier(1999)] Hillier B, and Hanson J, 1999, *Space is the Machine: A Configurational Theory of Architecture* (Cambridge University Press, Cambridge, MA, USA)

[Hillier and Vaughan(2007)] Hillier B, Vaughan, L, 2007, ”The city as one thing” *Prog. Plann.* **67**(3) 205–230.

[Jiang and Claramunt(2002)] Jiang B, Claramunt, C, 2002, ”Intergration of space syntax into GIS : new perspectives for urban morphology” *Transactions on GIS* **6**(3) 295–309.

[Katz(1953)] Katz L, 1953, ”A new index derived from sociometric data analysis” *Psychometrika* **18** 39–43

[Langville and Mayer(2005)] Langville A N, Mayer C D, 2005, ”Deeper inside PageRank” *Internet Mathematics* **1**(3) 335–380

[Latora and Marchiori(2004)] Latora V, Marchiori M, 2004, ”A measure of centrality based on network efficiency”, http://arxiv.org/abs/cond-mat/0402050

[Newman(2003)] Newman M E J, 2003, ”A measure of betweenness centrality based on random walks” Available via arXiv, http://arXiv:cond-mat/0309045

[Nieminen(1974)] Nieminen J, 1974, ”On centrality in a graph” *Scandinavian Journal of Psychology* **15**(322) 322–336.

[Offenhuber and Ratti(2014)] Offenhuber D, Ratti C, 2014, *Decoding the city. Urbanism in the age of Big Data* (Birkhäuser Verlag, Basel, Switzerland)

[Oyana and Margai(2015)] Oyana T J, Margai F, 2015, *Spatial Analysis. Statistics, Visualization, and Computational Methods* (CRC Press, New York)

[Page et al(1999)] Page L, Brin S, Motwani R, Winograd T, 1999, ”The pagerank citation ranking: Bringing order to the web” Technical report 1999-66 Stanford InfoLab (Stanford, CA, USA)

[Porta et al(2006)] Porta, S, Crucitti, P, Latora, V, 2006, ”The network analysis of urban streets: a primal approach” *Environment and Planning B: planning and design* **33**(5) 705–725

[Porta et al(2009)] Porta S, Strano E, Iacoviello, V, Messora, R, Latora V, Cardillo, A, Wang F, Scellato S, 2009, ”Street centrality and densities of retail and services in Bologna, Italy” *Environment and Planning B: planning and design* **36**(3) 450–465

[Porta et al(2012)] Porta S, Latora V, Wang F, Rueda S, Strano E, Scellato S, Cardillo A, Belli E, Cárdenas F, Cormenzana B, Latora L, 2012, ”Street centrality and the location of economic activities in Barcelona” *Urban Studies* **49**(7) 1471–1488

[Sabidussi(1966)] Sabidussi G, 1966, ”The centrality index of a graph” *Psychometrika* **31** 581–603

[Sevtsuk(2010)] Sevtsuk A, 2010, ”Path and Place: A Study of Urban Geometry and Retail Activity in Cambridge and Somerville, MA” PhD Dissertation, MIT Department of Urban Studies and Planning, Cambridge, MA, USA.

[Sevtsuk(2012)] Sevtsuk A, 2012, ”Urban Network Analysis Toolbox” *International Journal of Geomatics and Spatial Analysis* **22**(2) 287–305

[Sevtsuk(2014)] Sevtsuk A, 2014, ”Location and Agglomeration: the Distribution of Retail and Food Businesses in Dense Urban Environments” *Journal of Planning Education and Research* **34**(4) 374–393