

A Variant of the Current Flow Betweenness Centrality and its Application in Urban Networks

Taras Agryzkov^a, Leandro Tortosa^a, Jose F. Vicent^{a,*}

^a*Departamento de Ciencia de la Computación e Inteligencia Artificial, Universidad de Alicante, Campus de San Vicente, Ap. Correos 99, E-03080, Alicante, Spain*

Abstract

The current flow betweenness centrality is a useful tool to estimate traffic status in spatial networks and, in general, to measure the intermediation of nodes in networks where the transition between them takes place in a random way. The main drawback of this centrality is its high computational cost, especially for very large networks, as it is the case of urban networks. In this paper, a new approach to the current flow betweenness centrality for its practical application in urban networks with data is presented and discussed. The new centrality measure allows the estimation of pedestrian flow developed in urban networks, taking into account both the network topology and its associated data. In addition, its computational cost makes it suitable for application in networks with a large number of nodes. Some examples are studied in order to better understand the characteristics and behaviour of the proposed centrality in the context of the city.

Keywords: urban centrality measures, random-walk betweenness, current flow betweenness, spatial networks, urban networks.

1. Introduction

Centrality is a topic of great interest in the field of complex networks science [1]. The findings of some important nodes with high centralities to characterize the properties on the networks have significant uses in many fields. For instance, to identify the influential spreaders in complex networks [2] or to quantify the nodal spreader capability and differentiate the influence of various nodes [3].

Throughout the years, numerous researchers designed different types of centralities [4, 5]. One of them is the shortest-path betweenness centrality (SPB), introduced by Freeman in 1977 [6]. This measure quantifies the frequency or the number of times a node acts as a bridge along the shortest path between two other nodes. The information on intermediary nodes is of great interest for many practical applications. For example, in telecommunication networks, SPB centrality measures the control exerted by a given node over an information traffic [7]. However, in transportation networks, this centrality measures the traffic flow through the nodes [8], while in biological networks it allows to identify critical genes [9]. Finally, in social networks, it measures the intermediation capacity of network users [10].

But, there are many networks in which the traffic does not flow following geodesic paths [11, 12]. For instance, we have pedestrian flows in urban networks, news, rumours and messages throughout social networks that are usually not transmitted following geodesic paths, but rather spread throughout the network to reach their destination.

Freeman [11] suggested a new centrality known as flow betweenness, whose main contribution is the idea of including non-geodetic routes to the final score of a node. This measure was based on the well-known

[✉]Partially supported by the Spanish Government, Ministerio de Economía y Competividad, grant number TIN2017-84821-P.

*Corresponding author.

Email addresses: taras.agryzkov@ua.es (Taras Agryzkov), tortosa@ua.es (Leandro Tortosa), jvicent@ua.es (Jose F. Vicent)

problem of maximum flow in networks [13, 14], where the maximum flow between the origin and destination node involves several types of paths, not only geodesics. In general, this centrality can be defined as the proportion of the quantity of flow that passes through the node, when the flow reaches the maximum value in its transmission between the source and the target, respectively. Nevertheless, the paths followed by the flow must be optimized to achieve its maximum value. This approach is valid and very useful for some types of networks, but does not solve many real situations where information moves in a random way. In order to solve this question, Newman proposes another measure of betweenness centrality based on the concept of random walks [15], showing the close relationship of this new measure with the flow of the electrical charge in the circuits, thus introducing current-flow betweenness centrality (CFB).

The CFB centrality is useful to estimate traffic status in spatial networks and, in general, to measure the intermediation of nodes in networks where the transition between them takes place in a random way. The main weakness of this centrality is its high computational cost, especially for very large networks, as in the case of urban networks. Attempting to improve the computational cost, several approaches have been developed, focusing on the speed or accuracy of the results [16, 17, 18]. All developed approaches to CFB centrality are not easily adaptable to urban networks, where the traffic between nodes is conditioned by the topology and the location of data describing different types of services present in the city.

The main objective of this work is twofold: firstly, to design and implement an alternative approach to approximate betweenness centrality based on the concept of current flow; secondly, to adapt the proposed centrality for its practical application in urban networks with geolocated data.

The paper is organized as follows: in section 2, an adaptation of betweenness centrality in urban networks is discussed. The basic characteristics of the generalized CFB centrality are introduced in Section 3, while Section 4 analyses with detail the new approach to approximate the generalized CFB centrality and its adapted version for urban networks with data. In Section 5 some examples are shown where the new proposed centrality is evaluated. Finally, conclusions are summarized in Section 6.

2. Betweenness centrality and its application in urban networks

Before analysing centrality measures based on betweenness centrality concept, it is appropriate to make a reflection on the usefulness and convenience of this centrality in urban networks.

The measurement of pedestrian and transport flows has always aroused great interest in the disciplines related to the study of the city. Previous studies (see [19, 20, 21]) show that information on places more and less transited by people allows to measure the quality of urban places, study the trend of local commerce and estimate business opportunities. Likewise, the quantitative analysis of traffic flow allows forecasting of traffic jams in transportation networks [22, 23].

The original idea of the shortest-path betweenness SPB centrality can be related to the flow in spatial networks (or transited networks) [24, 23]. However, several studies show that the direct application of the SPB does not provide satisfactory results for spatial networks, due to its traffic does not strictly follow the pattern of the shortest paths [25, 24, 23].

There exists different adaptations of the SPB centrality according to the type of traffic and the type of spatial networks considered for the analysis. For the analysis of traffic flow in transportation networks, betweenness centralities are usually used. The reason is that, in addition to considering topological or euclidean distances, they also consider the qualitative characteristics of the roadway and the reasons for displacement [25, 24, 26]. On the other hand, the pedestrian traffic in spatial networks is less deliberate than in transportation networks, as well as it presents a more random nature (see [25, 27]). Furthermore, in urban networks, pedestrian flows depend not only on the topology of the network, but also on the data describing different types of services or facilities present in the city [21, 28]. These facts leads us to consider more appropriate, in these cases, the use of betweenness centrality based on the concept of random walks or current flow. Therefore, if our objective is the design and application of centrality measures, such as current flow betweenness, to urban networks, we must include in their computation the contribution or influence of data from the urban context.

3. Betweenness centrality based on the concept of current flow

In [15], Newman introduces the betweenness centrality based on the concept of current flow (CFB), in order to treat the intermediation of nodes in networks where data traffic does not follow strictly along geodesic paths. Later, Brandes and Fleischer [16] propose a generalized version of this centrality that can be applied to weighted and unweighted graphs. In the following, we briefly describe the generalized version of CFB, since it is the centrality on which the model proposed in this work is based.

Consider the example with the undirected graph $G = (V(G), E(G))$ representing an electrical circuit (see Fig. 1). The resistance $r(e_{ij})$ and the current $x(e_{ij})$ are associated to each edge $e_{ij} \in E(G)$ between the nodes i and j . Let x_s be an external current introduced to the source node s and x_t the current extracted from the target node t . Let us calculate the generalized CFB centrality for every node in G .

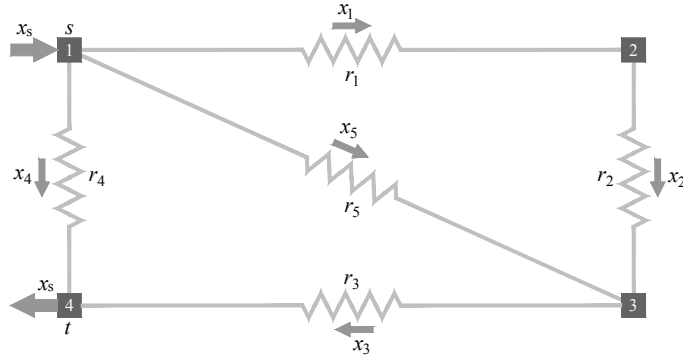


Figure 1: The example of electrical circuit

First, we obtain its voltage. According to Kirchhoff's first law, the algebraic sum of currents entering and leaving the node is equal to zero. Therefore,

$$\sum_{j=1}^{k_i} x(e_{ij}) = \sum_j \frac{u_i - u_j}{r(e_{ij})} = 0, \quad (1)$$

where k_i is the degree of the node i , and $u_i - u_j$ is the difference of electrical potential between i and j . Applying Eq. (1) to each node in G we obtain

$$Q\mathbf{u} = \mathbf{c}. \quad (2)$$

In Eq. (2), $Q \in \mathbb{R}^{n \times n}$ is the matrix of conductances $q(e_{ij}) = 1/r(e_{ij})$ associated to the edges e_{ij} of G , $\mathbf{u} \in \mathbb{R}^{n \times 1}$ is the voltage vector and $\mathbf{c} \in \mathbb{R}^{n \times 1}$ is the external current vector defined by

$$c_i = \begin{cases} +x_s & \text{for } i = s, \\ -x_t & \text{for } i = t, \\ 0 & \text{other case.} \end{cases} \quad (3)$$

Remark that Q is the Laplacian matrix $\mathcal{L}_w = (l_{ij}) \in \mathbb{R}^{n \times n}$ for weighted graphs whose elements are given by

$$l_{ij} = \begin{cases} \sum_j^{k_i} q(e_{ij}) & \text{if } i = j, \\ -q(e_{ij}) & \text{if } e_{ij} \in E, \\ 0 & \text{other case.} \end{cases}$$

Since Q is the Laplacian matrix of weighted graphs, it can be calculated directly by the expression

$$Q = B\Omega B^T, \quad (4)$$

where $B \in \mathbb{R}^{n \times m}$ is the incidence matrix and $\Omega = (\omega_{ij}) \in \mathbb{R}^{m \times m}$ is the diagonal matrix with conductances assigned to the edges of the graph.

$$\omega_{ij} = \begin{cases} q(e_{ij}) & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

The next step is to compute the voltages vector \mathbf{u} of the expression Eq. (2), which we use to obtain the currents $x(e_{ij})$.

As Q is singular, the vector \mathbf{u} can not be calculated directly by using the inverse of Q . This drawback may be solved by transforming any node of the circuit into a *reference node* r with a known voltage $u_r = 0$. Therefore, the row and the column corresponding to the node r can be deleted from Q without compromising the results of the analysis. By removing the r -th row and column of Q , we obtain a new matrix denoted by $Q_r \in \mathbb{R}^{(n-1) \times (n-1)}$. Since Q_r is non-singular, we calculate its inverse and later, we insert the r -th row and column of zeros into the resulting matrix. The outcome matrix is denoted by $M = (m_{ij}) \in \mathbb{R}^{(n) \times (n)}$. For instance, by taking the n -th node as reference we obtain

$$M = \begin{bmatrix} Q_r^{-1} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix}.$$

Now, we use M to compute the voltages vector \mathbf{u} , where

$$\mathbf{u} = M\mathbf{c}. \quad (5)$$

Taking into account the characteristics of the vector \mathbf{c} from Eq. (3) and the construction of the matrix M by columns, it turns out that

$$\mathbf{u} = x_s \mathbf{m}_s - x_t \mathbf{m}_t,$$

where \mathbf{m}_s and \mathbf{m}_t are s -th and t -th columns of the matrix M . Since $x_s = x_t$,

$$\mathbf{u} = x_s (\mathbf{m}_s - \mathbf{m}_t).$$

Thus, the voltage u_i of the n nodes of the graph is given by the expression

$$u_i = x_s (m_{is} - m_{it}). \quad (6)$$

Once the voltages of all the nodes of the graph are known, the current flow $I_i^{(st)}$ from the source s to the target t , through the node i , is calculated as

$$I_i^{(st)} = \frac{1}{2} \sum_j^{k_i} \frac{|u_i - u_j|}{r(e_{ij})}, \quad \text{for } i \neq s, t. \quad (7)$$

Replacing Eq. (6) in Eq. (7) and taking into account that $1/r(e_{ij})$ is the conductance $q(e_{ij})$ of the edge e_{ij} , then

$$I_i^{(st)} = \frac{x_s}{2} \sum_j^n q(e_{ij}) |m_{is} - m_{it} - m_{js} + m_{jt}|, \quad \text{for } i \neq s, t. \quad (8)$$

Obviously, this expression is not valid for the source s and the target t , where a constant flow of a x_s is specified, so that

$$I_s^{(st)} = x_s = I_t^{(st)}. \quad (9)$$

It should be noted that we can use an alternative criterion to avoid consider the extreme nodes s and t for the current routes, in which case

$$I_s^{(st)} = I_t^{(st)} = 0. \quad (10)$$

Finally, the generalized CFB centrality of the node i can be expressed as the average of the current flow $I_i^{(st)}$ accumulated by i when all the pairs of source and target nodes of the graph are considered, that is,

$$b_i = \frac{1}{N_b} \sum_{s < t} I_i^{(st)}. \quad (11)$$

In Eq. (11), N_b is the *normalization constant* that represents the number of source and target pairs considered for the calculation. For instance, if all the possible source and target pairs of nodes are used,

$$N_b = \frac{1}{2}n(n-1).$$

The calculation process of the generalized current flow betweenness centrality (GCFB) using Eq. (10) and, for $x_s = 1$, is summarized by the pseudo-code shown in Algorithm 1.

In the networks representing circuits of continuous electrical current this centrality measure the node importance in terms of its frequency of participations in the routes that follows the current between a source and a target.

Input: $\mathcal{L} \in \mathbb{R}^{n \times n}$, the graph Laplacian matrix of $G = (V(G), E(G))$
 $E = \{e_{ij} \mid i, j \in V(G)\}$, the set of edges
Output: $\mathbf{b} \in \mathbb{R}^n$, the vector of CFB centralities for n nodes

```

begin
   $\mathbf{b} \leftarrow \mathbf{0}$ 
   $M \in \mathbb{R}^{n \times n} \leftarrow \mathcal{L}^{-1} \in \mathbb{R}^{(n-1) \times (n-1)}$ 
  for  $t \leftarrow n$  to  $t > 1$  do
    for  $s \leftarrow 1$  to  $s < t$  do
      foreach  $e_{ij} \in E$  do
        if  $i \neq s, t$  then
           $b[i] \leftarrow b[i] + (1/2)|M[i, s] - M[i, t] - M[j, s] + M[j, t]|$ 
        end
      end
    end
  end
   $\mathbf{b} \leftarrow 2\mathbf{b}/(n(n-1))$ 
end

```

Algorithm 1: Generalized CFB centrality.

An important aspect that must be taken into account when computing this centrality measure is its computational cost, which is of order $O(I(n-1) + mn^2)$. Remark that $O(I(n-1))$ is the computational cost corresponding to the calculation of the inverse matrix and $O(mn^2)$ is the computational cost of all currents for n nodes using $n(n-1)/2$ pairs of nodes. The computational cost is acceptable if small networks are considered. However, for larger networks, approximate methods are usually used. Among the approximate methods we can highlight the probabilistic approach [16], the distributed setting by defining a vertex-centric, gossip-based algorithm [18], or the method which approximate the generalized inverse of the graph Laplacian matrix [17]. All of them offer an important improvement in the calculation of current flows; however, they alter the accuracy of the results or present an excessively laborious implementation.

4. An adapted model of current flow betweenness centrality for urban networks

because of the large size of urban networks, a direct computation of the GCFB centrality is not practical. Consequently, an approximate model should be implemented. In addition, the approximate model should also be compatible with the intrinsic characteristics of urban networks with data.

The centrality measure proposed in this section is based on the generalized current flow betweenness (GCFB) previously described. Our objective is twofold: on the one hand, to present an alternative approach

to centrality GCFB with an improved computational cost and, on the other hand, to adapt the proposed centrality for its practical application in urban networks with data.

4.1. The resized approximation of current flow betweenness centrality

The main idea of the proposed centrality, that we denote as *resized approximation of current flow betweenness centrality* (RCFB), consists on an approximation of the GCFB centrality by using only n source nodes, instead of $n(n-1)/2$ pairs of source and target nodes.

Considering the undirected graph representing an electrical circuit of Fig. 1. The first task is to add an additional node to the graph G and link it to the rest of the nodes. As it can be seen in Fig. 2, we obtain an extended graph G' with $n+1$ nodes and $m+n$ edges. The following condition is required: the additional

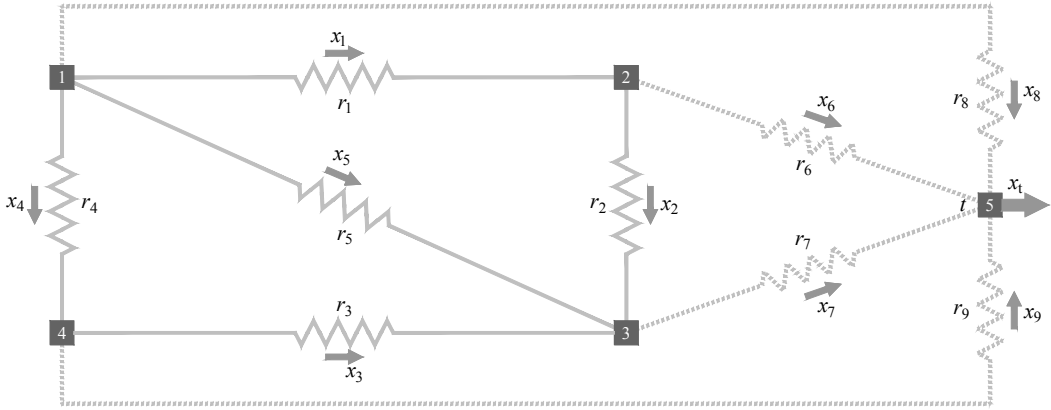


Figure 2: The example of electrical circuit

node t must be the target, from which we extract the current x_t , while the remaining n nodes of G are used as the sources s .

The next step is to make that all the currents $x(e_{it})$ entering the node t reach values as similar as possible. Thus, we can estimate the approximate value of $x(e_{it})$ and handle the amount of the residual current that leaves from each node towards the target and that does not participate in the computation of the centrality of the nodes of the original graph G .

For this purpose, we assign to the edges e_{it} the resistance $r(e_{it})$ with a very high value. As an example, we can consider a resistance with a value greater than the algebraic sum of all the resistances of the circuit, that is,

$$r(e_{it}) > \sum_{e_{ij} \in E} r(e_{ij}), \quad \text{for } i, j \neq t.$$

With this arrangement, $r(e_{it})$ is greater than the resistance of any circuit path composed of n nodes. Therefore, the greater proportion of current will be conducted preferably by any path with the nodes $i = 1, 2, \dots, n$, before the edges e_{it} . Besides, the resistance $r(e_{it})$ affects minimally to the local distribution of the currents $x(e_{ij})$ in the i node, with $i \neq t$. Thus, all the currents that flow into the target node have a value $x(e_{it}) \approx x_t/n$, which can be subtracted in the final computation of the totality of the current that passes through the node i .

Remark that, with high resistances in the edges, we do not intend to force the situation so that all residual currents are equal. If so, this would imply that all the voltages of the circuit nodes would be equal and, consequently, there would be no current in the original circuit. Conversely, if the resistance of the edges is not large enough the final result will be less precise with respect to the original GCFB centrality.

Next, we proceed with the definition of the extended conductance matrix, denoted by $Q_r \in \mathbb{R}^{(n+1) \times (n+1)}$. Following a similar reasoning as used for obtaining Eq. (4),

$$Q_r = B_r \Omega_r B_r^T, \tag{12}$$

where $B_r \in \mathbb{R}^{(n+1) \times (m+n)}$ is the extended incidence matrix and $\Omega_r = (\omega_{ij}) \in \mathbb{R}^{(m+n) \times (m+n)}$ is the diagonal matrix of conductances extended with n new edges. The elements of Ω_r are given by

$$\omega_{ij} = \begin{cases} q(e_{ij}) & \text{for } i = j, \text{ with } j \leq m, \\ \varepsilon & \text{for } i = j, \text{ with } j > m, \\ 0 & \text{for } i \neq j, \end{cases} \quad (13)$$

where $q(e_{ij}) = 1/r(e_{ij})$ is the conductance of the edge e_{ij} and $\varepsilon < 1/r(e_{it})$ is the conductance of the edges e_{it} , that is

$$\varepsilon < \frac{1}{\sum_{e_{ij} \in E} r(e_{ij})}, \quad \text{for } i, j \neq t. \quad (14)$$

Considering the target node t as the reference node, the voltages of the nodes can be calculated analogously to the GCFB centrality. However, if this consideration is assumed, the whole process from the beginning can be simplified.

Recall that once the GCFB centrality matrix Q_t^{-1} was calculated, we inserted the row and the column of zeros corresponding to the node t , obtaining in this way another matrix M . Later, in the computation of $I_i^{(st)}$, the t -th row of M was used to compute the currents of the incident edges in the t node, while the t column of zeros was not used. With the proposed centrality, it is not necessary to insert the t -th row in matrix Q_r^{-1} . This is because the currents corresponding to the incident edges in the t node are the residual currents with known value, $x(e_{it}) \approx x_s/n$. Neither the t -th column of zeros has no interest. Thus, we can dispense with t components of Q_r^{-1} . This implies that, when considering the additional node as the target and reference node at the same time from the beginning, the auxiliary node can be dispensed with and consequently it is not necessary to use the extended graph.

Hence, we build the conductance matrix $Q = (q_{ij}) \in \mathbb{R}^{n \times n}$ directly using the expression

$$Q = \varepsilon I + \mathcal{L}_w, \quad (15)$$

where

$$\mathcal{L}_w = B\Omega B^T = \Delta_w - A_w, \quad (16)$$

is the Laplacian $n \times n$ matrix of weighted graph G and ε is a constant previously defined. The matrix $\Delta_w \in \mathbb{R}^{n \times n}$ of Eq. (16) is the degree matrix and $A_w \in \mathbb{R}^{n \times n}$ is the adjacency matrix, both associated to the weighted graph G .

By introducing εI with $\varepsilon > 0$, Q is strictly diagonally dominant by rows

$$|q_{ii}| > \sum_{j \neq i}^n |q_{ij}|, \quad \forall i,$$

as well as by columns

$$|q_{jj}| > \sum_{i \neq j}^n |q_{ij}|, \quad \forall j.$$

Thus, based on the Levy-Desplanques theorem [29], we can say that Q is a non-singular matrix, so that the voltages of the nodes can be calculated directly, that is,

$$\mathbf{u} = Q^{-1} \mathbf{c},$$

where \mathbf{c} is the vector with all its components equal to zero, except the s -th component that has the value of x_s . Consequently, the voltage of the node i is given by

$$u_i = x_s [Q^{-1}]_{is}.$$

Once the voltages u_i are obtained, we proceed with the calculation of the totality of the current, denoted by $I_i^{(s)}$, from the source s through the node i . Let us remember that we are interested only in the part of

the current flowing through the original circuit with n nodes. Therefore, the residual current $x(e_{ij})$ must be subtracted from the totality of the current $I_i^{(s)}$, that is,

$$I_i^{(s)} = \frac{x_s}{2} \left(-\frac{1}{n} + \sum_{e_{ij} \in E} q(e_{ij}) \left| [Q^{-1}]_{is} - [Q^{-1}]_{js} \right| \right), \quad \text{for } i \neq s. \quad (17)$$

Consequently, the total current that enter the node i is equivalent to the total current leaving it, except for the residual current $x(e_{ij})$ taken as $\frac{x_s}{n}$. So,

$$I_i^{(s)} = x_s \sum_{e_{ij} \in E} q(e_{ij}) \left([Q^{-1}]_{is} - [Q^{-1}]_{js} \right), \quad (18)$$

for $i \neq s$ and $\left([Q^{-1}]_{is} - [Q^{-1}]_{js} \right) > 0$. For the node s , the totality of the current can be taken as 0, or it can be calculated as

$$I_s = x_s - \frac{x_s}{n}. \quad (19)$$

Finally, we calculate the value of the RCFB centrality in a similar way as for the centrality GCFB. Considering that we only take n pairs of source and target nodes, the expression of the RCFB centrality b_i of the node i is given by

$$b_i = \frac{1}{n} \sum_s^n I_i^{(s)}. \quad (20)$$

The process to compute the RCFB centrality with the criterion $I_s = 0$ and for $x_s = 1$ is summarized by the pseudo-code of [Algorithm 2](#).

Input: $\mathcal{L} \in \mathbb{R}^{n \times n}$ The graph $G = (V(G), E(G))$ Laplacian matrix
 $E = \{e_{ij} \mid i, j \in V(G)\}$ The set of edges
 $\epsilon \in \mathbb{R} > 0$ The minimal conductance

Output: $\mathbf{b} \in \mathbb{R}^n$ The vector of RCFB centralities for n nodes

```

begin
   $\mathbf{b} \leftarrow \mathbf{0}$ 
   $M \in \mathbb{R}^{n \times n} \leftarrow (\epsilon I_n + \mathcal{L})^{-1}$ 
  for  $s \leftarrow 1$  to  $n$  do
    foreach  $e_{ij} \in E \mid i < j$  do
       $x \in \mathbb{R} \leftarrow M[i, s] - M[j, s]$ 
      if  $x > 0$  then  $b[i] \leftarrow b[i] + x$ 
      else  $b[j] \leftarrow b[j] + (-x)$ 
    end
  end
   $\mathbf{b} \leftarrow \mathbf{b}/n$ 
end

```

Algorithm 2: RCFB centrality.

The results obtained by the proposed centrality are very similar to the results obtained by the GCFB centrality, when we use all possible pairs of source and target nodes. In order to confirm the similarity between the two centralities, we show an example where four different networks composed by 75 nodes generated in a random way are evaluated.

We show the results obtained in the plots of [Fig. 3](#) and [Fig. 4](#); on the horizontal axis we sort the nodes based on the values of the GCFB centrality, while on the vertical axis we represent the b_i values of the GCFB and RCFB centralities corresponding to the ordered nodes.

The gap between the results of both centralities is due to the residual current not computed for the RCFB centrality. Despite the similarity of the results between the GCFB and RCFB centralities, it is important to highlight some differences.

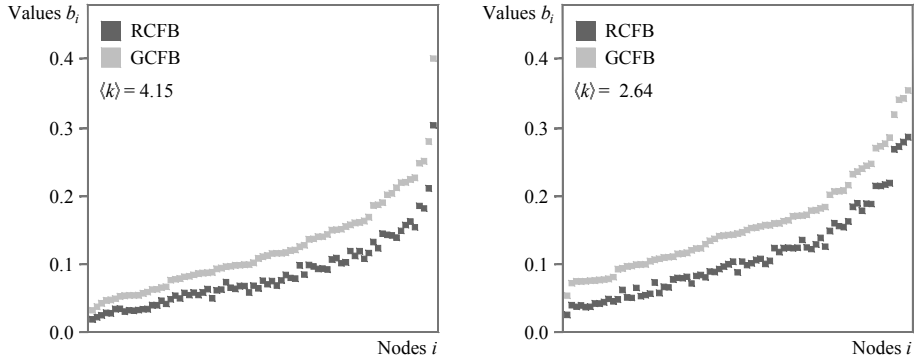


Figure 3: Comparison of GCFB and RCFB centralities

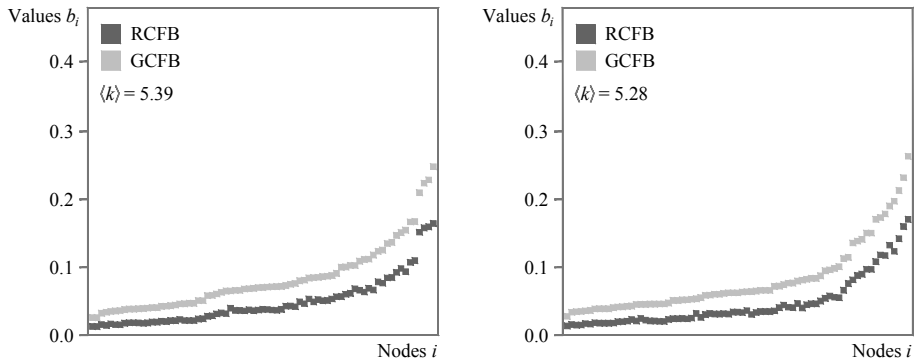


Figure 4: Comparison of GCFB and RCFB centralities

The first difference is the computational cost corresponding to the calculation of the current totalities for n nodes. With the RCFB centrality being the most efficient, since it uses only n source nodes in its calculation, instead of $n(n-1)/2$ pairs of source and target nodes used in the case of the GCFB centrality. The computational cost of the RCFB is $O(nm/2)$, which is of an order less than for the GCFB centrality. In addition, when calculating the inverse of Q directly, we avoid the elimination and insertion of the t -th matrix components, which represents a certain improvement for the execution time, especially for very large matrices.

The other difference between both centralities lies in the concept and interpretation of the results. We associate the GCFB centrality with the flows of the current that depend on the target node, while we relate the RCFB centrality with the flows of the current distributed along the circuit without taking into account the target node.

An interesting aspect of the RCFB centrality is that as the number of nodes in the circuit approaches infinity, the residual current, that is, the unaccounted current, approaches zero and consequently we count a greater quantity of the current for the centrality.

Up to now, we have developed the first objective of the section consisting on the design of the RCFB centrality. In the next subsection, we proceed with the second objective, which is to adapt the proposed centrality to estimate pedestrian flows in urban networks with data.

4.2. RCFB centrality for urban networks

In the case of the electric circuits treated throughout previous sections, the current (or electric charge flow) depends on the topology and the resistance of the circuit. In the case of urban networks, we assume

that the pedestrian flow depends on the topology of the network and the data context. Because of the RCFB centrality is based on the principles of electric current, it is necessary to modify certain considerations for its application in urban networks.

To begin with, we must establish an analogy between the flow of electric charge and the pedestrian flow. In this sense, the difference in electric potential between two nodes of a circuit can be interpreted in terms of the need (or will) that the pedestrian has to move from one node to another within the urban network. The resistance of the edge of an electrical circuit can be related to the amount of data associated with the edge of the urban network.

Note that the resistance and the current are inversely proportional, that is, the more resistance the electric conductor (edge of the circuit) has, the smaller amount of the current is transmitted through it. However, in urban networks, data and pedestrian flow are directly proportional, ie, the more data the transition path has (edge of the urban network), the greater number of pedestrians will transit through it. Thus, the pedestrian flow between two nodes i and j connected by an edge e_{ij} , denoted by $x(e_{ij})$, is expressed as

$$x(e_{ij}) = d(e_{ij}) (u_i - u_j),$$

where $d(e_{ij})$ is the amount of data associated to the edge e_{ij} .

Once the analogy between the electric current and the pedestrian flow is established, we proceed with the definition of parameters involved in the calculation of the centrality.

We can say that the unique difference between the RCFB centrality previously exposed and its adapted version for urban networks is found in the construction of the conductance matrix Q . In this case, Q is calculated as

$$Q = \epsilon I + \Delta_w - A_w. \quad (21)$$

In Eq. (21) the parameter ϵ is the constant which is defined following the same reasoning used for electric circuits, while $\Delta_w = (\delta_{ij}) \in \mathbb{R}^{n \times n}$ with

$$\delta_{ij} = \begin{cases} \sum_{j=1}^{k_i} d(e_{ij}) & \text{for } i = j, \\ 0 & \text{for } i \neq j, \end{cases}$$

and $A_w = (a_{ij}) \in \mathbb{R}^{n \times n}$ is the adjacency matrix associated with a weighted graph with data

$$a_{ij} = \begin{cases} d(e_{ij}) & \text{if } e_{ij} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

The data of the urban networks can be associated with both edges and nodes. In this case we associate the data to the edges using the following expression

$$d(e_{ij}) = \alpha (v_i^* + v_j^*),$$

where the parameter α represents a scale factor for the importance of the data, v_i^* and v_j^* are elements of \mathbf{v}^* representing the data associated with the nodes i and j , respectively. Following this reasoning, the matrix A_w can be calculated as

$$A_w = A + \alpha (A \text{diag}(\mathbf{v}^*) + \text{diag}(\mathbf{v}^*) A),$$

where A is the adjacency matrix representing the topology of the network and $\text{diag}(\mathbf{v}^*) \in \mathbb{R}^{n \times n}$ is the diagonal matrix obtained from the data vector \mathbf{v}^* .

We calculate the data vector \mathbf{v}^* from the data matrix $D \in \mathbb{R}^{n \times m}$ and the weighting vector $\mathbf{v}_0 \in \mathbb{R}^{m \times 1}$,

$$\mathbf{v}^* = D\mathbf{v}_0.$$

The n rows of D represent the amount of data associated with the n nodes, while its m columns correspond to the m types of data. The weighting vector \mathbf{v}_0 assigns the importance to the different types of data according to the case study. More information about the data matrix, weighting vector and, in general, the entire process of data assignment can be consulted in [30, 31].

Once the components of the matrix Q are defined, the pedestrian flows can be computed using Eq. (18). Note that, instead of the conductances $q(e_{ij})$, in this case we use the data $d(e_{ij})$. Finally, the centrality of n nodes is calculated using Eq. (20).

5. Numerical examples of application of RCFB centrality in urban networks

In this section, we develop several examples where we apply the concepts and procedures of RCFB centrality previously exposed.

Before applying the proposed centrality in urban networks with data, we make an example applying the RCFB and GRFB centralities in an urban network without data. This example helps us to briefly compare both centralities in an urban context and show the similarity in their results. For this purpose, we analyse a network composed of 186 nodes that represents a portion of the urban network of Murcia (Spain).

We calculate the RCFB centrality using n source nodes and the GCFB centrality using $n(n-1)/2$ pairs of source and target nodes. In this case, for the calculation of the RCFB centrality we do not use the data of the network, aspect that allows us to compare it with the centrality GCFB in a direct way.

Once the two centralities are computed, we show the results obtained for the RCFB and GCFB centralities in Figures 5 (a) y 5 (b), respectively. On comparing both figures we observe, analysing the results of the chromatic range, that the differences between both centralities are minimal. In the detailed evaluation of the results, it should be noted that the small differences observed in the two centralities are due to the fact that the residual current used in the computation of the RCFB centrality is not the same for all the nodes, which is the condition assumed in the proposed model.

In order to insist on the similarity of the results of GCFB and RCFB centralities, we perform another analysis that consists on the comparison of the two centralities using a single source node s . Thereby, in Fig. 5 (c) and (d) we show the results of both centralities using a single source node located in the upper right part of the network (indicated by the arrow). As in the previous case, we do not appreciate significant or discordant differences.

It is worth noting the aspect related to the numerical difference of the maximum and minimum values of the results obtained for the RCFB centrality. Recall that in the calculation of the RCFB centrality, the only current that is computed was the leaving current of the node, different from the residual current.

Therefore, the node that presents a minimum value of the RCFB centrality is the one with only residual current leaving it. As a consequence, we do not take it into account for the node centrality. A similar reasoning is valid for the node that has the maximum value (0.995) of RCFB centrality, corresponding to the source node s . In this case, from the value 1.000 of the node s the residual current of 0.005 is subtracted, as indicated in Eq. (17).

In the following example we show the influence of data on the results of the RCFB centrality adapted for urban networks. For this example we consider two urban network typologies with different degree distributions: the urban network represented by a primal graph of the previous example (with homogeneous degree distribution) and the network represented by its dual graph following the linearity criterion [32] (with the characteristic degree distribution of scale-free networks [32, 33]). In both cases we consider three data units and a single source node s indicated with the arrow in Fig. 6(a,b,c,d) and Fig. 7(a,b,c,d). Next, we perform the calculations of the RCFB centrality with the scale factor $\alpha = 1000$. Once the calculations are performed, the obtained results are shown in the graphics of Fig. 6(a,b,c,d) and Fig. 7(a,b,c,d). Each graphic represents the results of the RCFB centrality with a different data location (indicated by a rectangle and the letter d). From these graphics we can observe a remarkable variation of the results when varying the data location. As we can see, in both networks with different degree distribution, starting from the source node, the pedestrian flow is concentrated in the vicinity of the nodes with data (or places of interest from the citizen point of view).

Moreover, these results depend on the scale factor α that weighs the influence of data. Regarding this parameter, it is clear that α affects the results only within a specific range that depends on the size of the network. After performing several tests with urban networks of different sizes and types, the more suitable range for α is $0 \leq \alpha \leq 10000$. For $\alpha > 10000$ there is no variation in the results, even for very large networks.

Once we have studied examples of a small network, next we apply the adapted model of the RCFB centrality to a real urban network with a real dataset. We use the urban network of the centre of Murcia city and the data set describing its commercial activity. In Fig. 8 the network and the dataset of the studied area are shown. The purpose of this example is the estimation of pedestrian flows through the RCFB centrality which takes into account the data of the urban context and the topology of the network.

Following the steps described in the previous section, we construct the data matrix D with 1196 rows representing the nodes of the network and 4 columns for the four types of data. We assign the same importance to the four categories of data, that is, the weighting vector is $\mathbf{v}_0 = [1, 1, 1, 1]^T$. Next, we construct the matrix $Q \in \mathbb{R}^{1196 \times 1196}$ by using Eq. (21). Once the inverse of Q is calculated and Eq. (20) is applied, we obtain the result of the RCFB centrality shown in Fig. 9.

In view of the results of Fig. 9, let us highlight several theoretically most transited places in the network: the central artery and the four central bridges that connect the two parts of the city. These results seem to be expected, considering the fact that the largest volume of data is located in the surroundings close to the central artery and that bridges usually concentrate pedestrian flow between both parts of the city.

6. Conclusions

In this paper, we present a variant of the betweenness centrality based on the concept of the current flow. Besides, we adapt the proposed centrality measure for its practical application in urban networks with data. The proposed centrality allows us to estimate, from a theoretical point of view, the most transited paths by the people taking into account the topology of an urban network and the data associated to its nodes. Through the presented examples we test the similarity of the proposed centrality with the original one. Moreover, we demonstrate its utility in the urban context where the pedestrian flows are influenced by different types of services present in the city.

- [1] Stephen P Borgatti and Martin G Everett. A graph-theoretic perspective on centrality. *Social networks*, 28(4):466–484, 2006.
- [2] Chao Li, Li Wang, Shiwen Sun, and Chengyi Xia. Identification of influential spreaders based on classified neighbors in real-world complex networks. *Applied Mathematics and Computation*, 320:512–523, 2018.
- [3] Juan Wang, Chao Li, and Chengyi Xia. Improved centrality indicators to characterize the nodal spreading capability in complex networks. *Applied Mathematics and Computation*, 334:388–400, 2018.
- [4] Ulrik Brandes. *Network analysis: Methodological foundations*, volume 3418. Springer Science & Business Media, 2005.
- [5] Charu C Aggarwal. *An introduction to social network data analytics*. Springer, 2011.
- [6] Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.
- [7] EC Juan Segovia and P Vila. New applications of the betweenness centrality concept to reliability-driven routing. In *VIII Workshop in G/MPLS Networks*, 2009.
- [8] Yihui Ren. *Betweenness centrality and its applications from modeling traffic flows to network community detection*. University of Notre Dame, 2015.
- [9] Chintagunta Ambedkar, Kiran Kumar Reddi, Naresh Babu Muppalaneni, and Duggineni Kalyani. Application of centrality measures in the identification of critical genes in diabetes mellitus. *Bioinformatics*, 11(2):90, 2015.
- [10] Linton C Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1978.
- [11] Linton C Freeman, Stephen P Borgatti, and Douglas R White. Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks*, 13(2):141–154, 1991.
- [12] Karen Stephenson and Marvin Zelen. Rethinking centrality: Methods and examples. *Social Networks*, 11(1):1–37, 1989.
- [13] Stephen P Borgatti. Centrality and network flow. *Social networks*, 27(1):55–71, 2005.
- [14] Lester R Ford. Network flow theory. Technical report, Rand Corp Santa Monica Ca, 1956.
- [15] Mark EJ Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27(1):39–54, 2005.
- [16] Ulrik Brandes and Daniel Fleischer. Centrality measures based on current flow. In *STACS*, volume 3404, pages 533–544. Springer, 2005.
- [17] Enrico Bozzo and Massimo Franceschet. Approximations of the generalized inverse of the graph laplacian matrix. *Internet mathematics*, 8(4):456–481, 2012.
- [18] Alessandro Lulli, Laura Ricci, Emanuele Carlini, and Patrizio Dazzi. Distributed current flow betweenness centrality. In *Self-Adaptive and Self-Organizing Systems (SASO), 2015 IEEE 9th International Conference on*, pages 71–80. IEEE, 2015.
- [19] Jason Corburn. *Toward the healthy city: People, places, and the politics of urban planning*. Mit Press, 2009.
- [20] Jan Gehl and Birgitte Svarre. *How to study public life*. Island Press, 2013.
- [21] Sergio Porta, Vito Latora, Fahui Wang, Salvador Rueda, Emanuele Strano, Salvatore Scellato, Alessio Cardillo, Eugenio Belli, Francisco Cardenas, Berta Cormenzana, et al. Street centrality and the location of economic activities in barcelona. *Urban Studies*, 49(7):1471–1488, 2012.
- [22] Amila Jayasinghe, Kazushi Sano, and Hiroaki Nishiuchi. Explaining traffic flow patterns using centrality measures. *International Journal for Traffic and Transport Engineering*, 5(2):134–149, 2015.
- [23] PX Zhao and SM Zhao. Understanding urban traffic flow characteristics from the network centrality perspective at different granularities. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41, 2016.

- [24] Aisan Kazerani and Stephan Winter. Can betweenness centrality explain traffic flow. In *12th AGILE International Conference on Geographic Information Science*, pages 1–9, 2009.
- [25] Song Gao, Yaoli Wang, Yong Gao, and Yu Liu. Understanding urban traffic-flow characteristics: a rethinking of betweenness centrality. *Environment and Planning B: Planning and Design*, 40(1):135–153, 2013.
- [26] Rami Puzis, Yaniv Altshuler, Yuval Elovici, Shlomo Bekhor, Yoram Shiftan, and Alex Pentland. Augmented betweenness centrality for environmentally aware traffic monitoring in transportation networks. *Journal of Intelligent Transportation Systems*, 17(1):91–105, 2013.
- [27] Bin Jiang and Tao Jia. Agent-based simulation of human movement shaped by the underlying street structure. *International Journal of Geographical Information Science*, 25(1):51–64, 2011.
- [28] Andres Sevtsuk. Location and agglomeration: The distribution of retail and food businesses in dense urban environments. *Journal of Planning Education and Research*, 34(4):374–393, 2014.
- [29] Olga Taussky. A recurring theorem on determinants. *The American Mathematical Monthly*, 56(10):672–676, 1949.
- [30] Taras Agryzkov, José L Oliver, Leandro Tortosa, and José F Vicent. An algorithm for ranking the nodes of an urban network based on the concept of pagerank vector. *Applied Mathematics and Computation*, 219(4):2186–2193, 2012.
- [31] Taras Agryzkov, Leandro Tortosa, José F Vicent, and Richard Wilson. A centrality measure for urban networks based on the eigenvector centrality concept. *Environment and Planning B: Urban Analytics and City Science*, pages 239–, 2017.
- [32] Sergio Porta, Paolo Crucitti, and Vito Latora. The network analysis of urban streets: a dual approach. *Physica A: Statistical Mechanics and its Applications*, 369(2):853–866, 2006.
- [33] Bin Jiang. A topological pattern of urban street networks: Universality and peculiarity. *Physica A: Statistical Mechanics and its Applications*, 25(1):51–64, 2011.

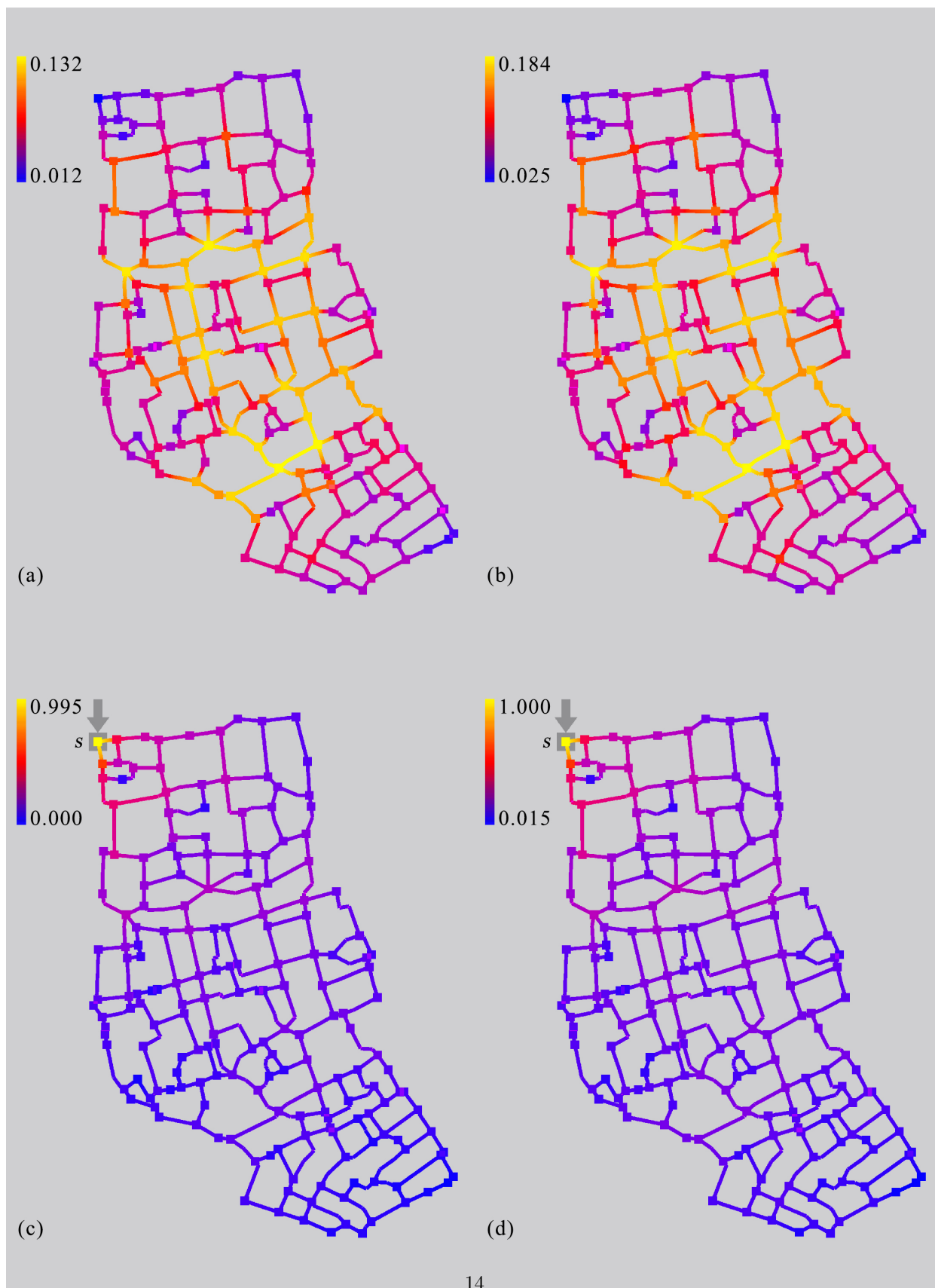


Figure 5: Comparison of GCFB and RCFB centralities applied to urban network.

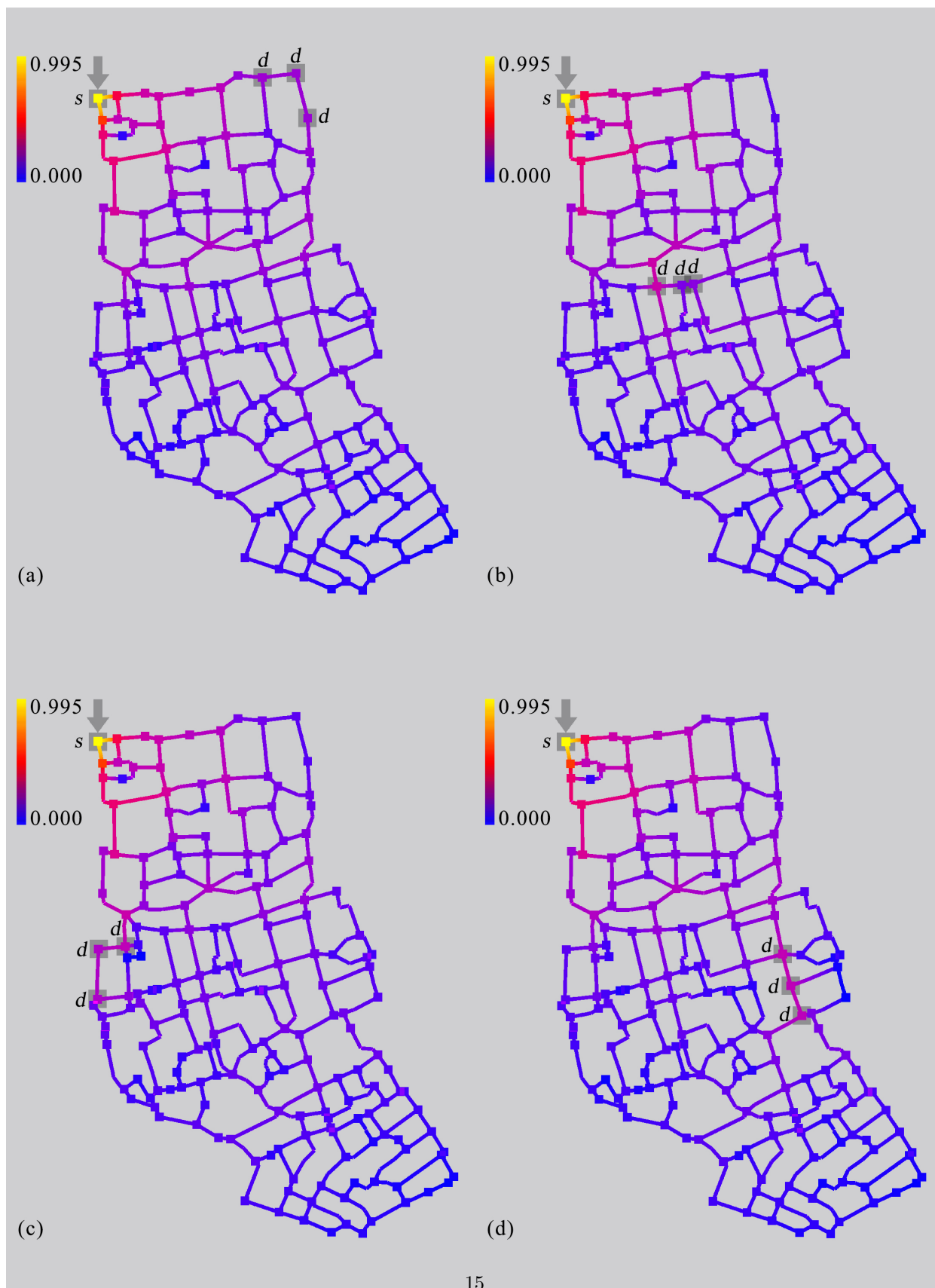


Figure 6: The RCFB centrality applied to urban network where the variation of data location is performed.

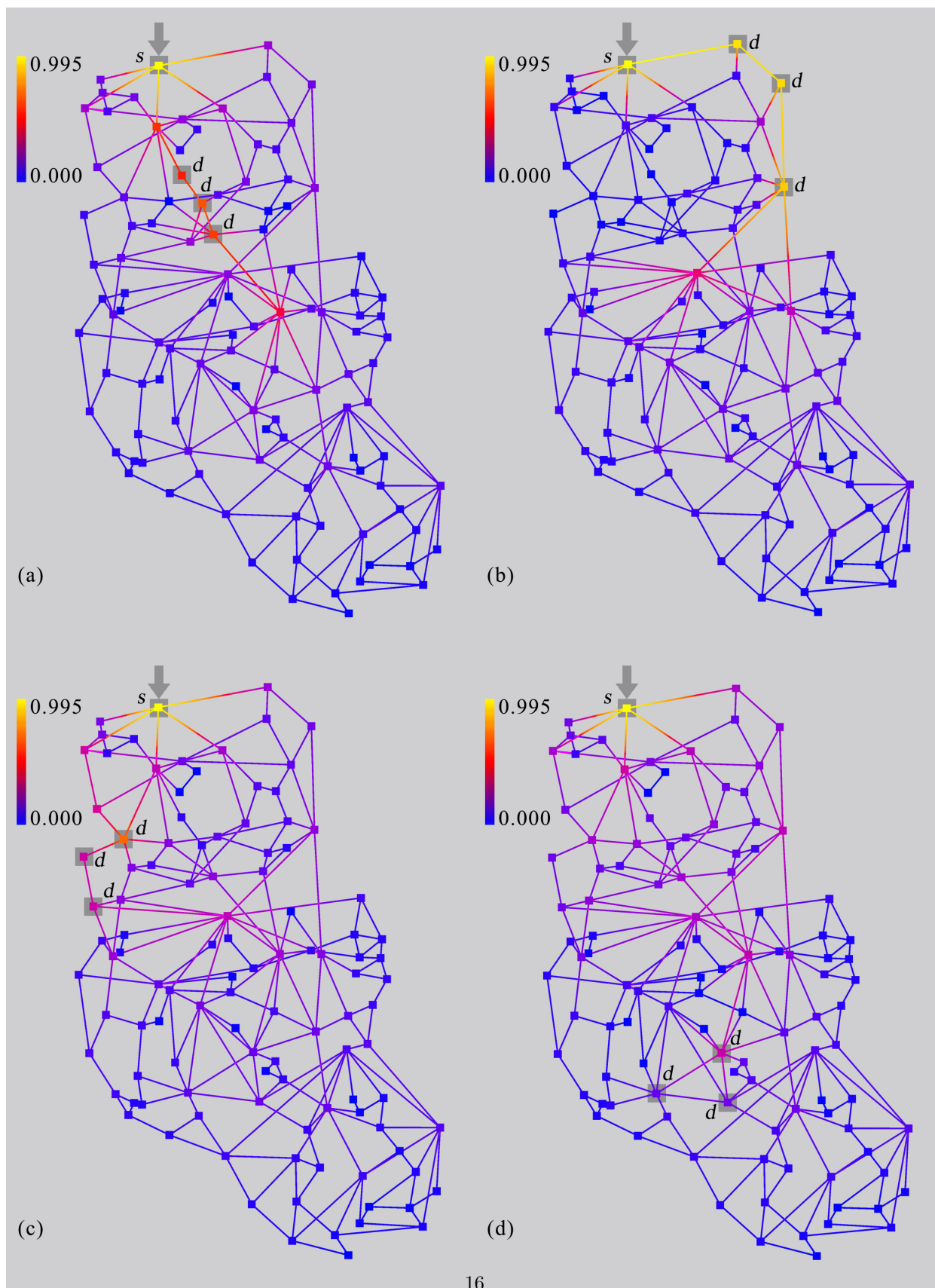


Figure 7: The RCFB centrality applied to urban network where the variation of data location is performed.

- Bancos y oficinas
- Comercio menor
- Comercio mayor
- Cafeterias y restaurantes



Figure 8: The urban network and dataset of Murcia city.

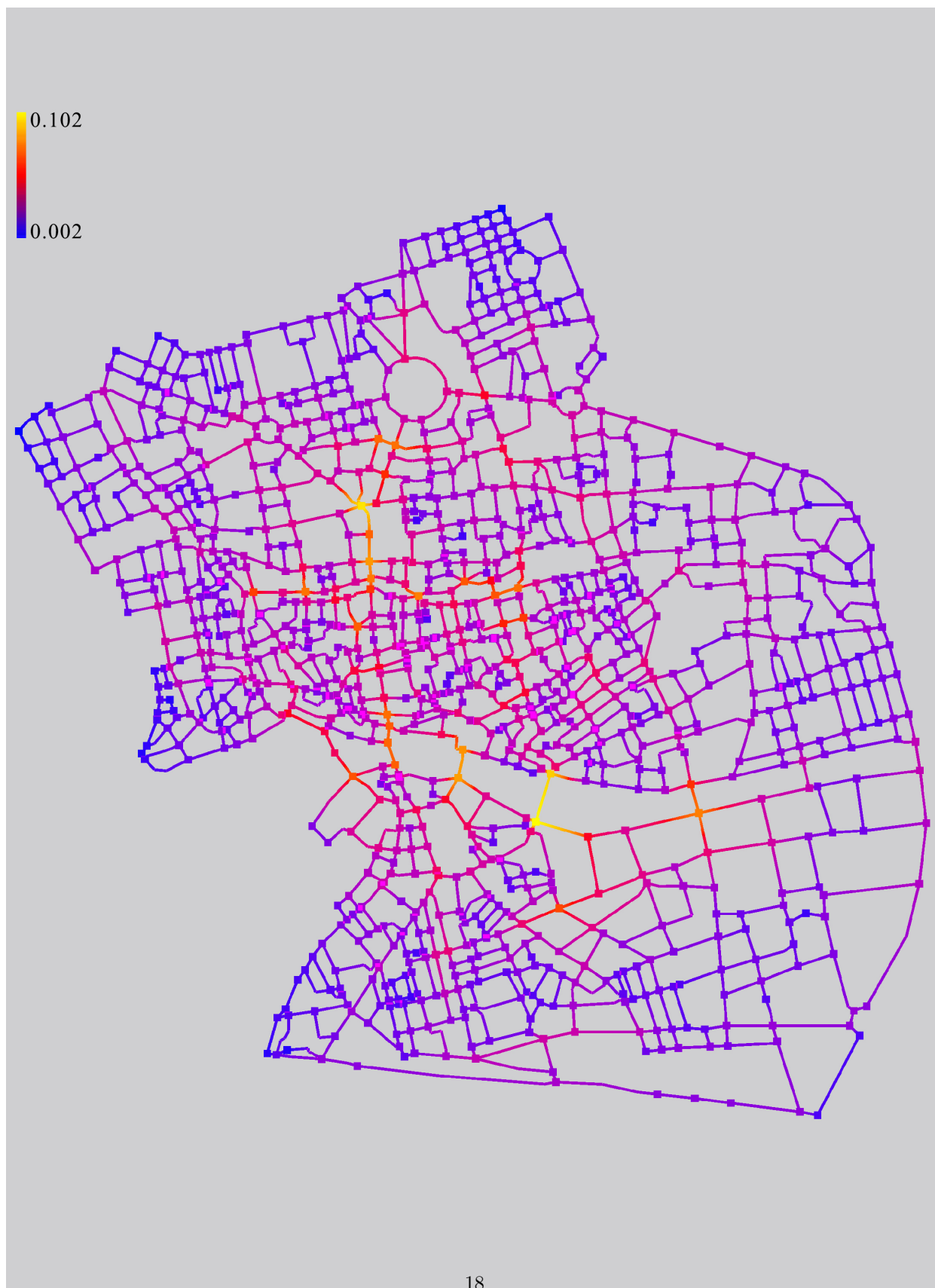


Figure 9: The RCFB centrality applied to urban network of Murcia city.