

## Complex network modeling of a river basin: an application to the Guadalquivir River in Southern Spain

R. Rodríguez-Alarcón<sup>a</sup> and S. Lozano <sup>b,\*</sup>

<sup>a</sup> Confederación Hidrográfica del Guadalquivir, Avda. del Brillante, 57, Córdoba 14012, Spain

<sup>b</sup> Escuela Superior de Ingenieros, University of Seville, Isla de la Cartuja, Seville 41092, Spain

\*Corresponding author. E-mail: slozano@us.es

 SL, 0000-0002-0401-3584

### ABSTRACT

In this paper, complex network analysis is used to analyze the structure of a whole river basin. A classification of nodes and edges that allow the river basin to be represented as a directed network is given. Both global and local characterization metrics are presented. Thus, centralization indexes, average path length, diameter, network efficiency, edge length, degree and strength distributions can be computed and interpreted. The closeness and betweenness centrality as well as the PageRank index of the different nodes can also be calculated and provide information on the relative position of each node in the network in terms of their average distance to upstream and downstream nodes, their reachability from other nodes, etc. In addition, a new centrality index based on the environmental impact of a potential spill or contaminant release on each node of the network has been applied. This type of centrality index is particularly useful in hydrologic networks and helps to identify key nodes from a contamination propagation point of view. The application of the proposed approach to the Guadalquivir River basin, in Southern Spain, is presented. Apart from validating the environmental impact centrality index, the most important finding is the scale-free character of the network and its consequences.

**Key words:** complex network analysis, environmental impact centrality, river basin, scale-free, small world

### HIGHLIGHTS

- A weighted directed network of a river basin with different types of nodes and edges.
- Global and local CNA characterization measures can be computed.
- A new environmental impact centrality index is proposed.
- An application to the Guadalquivir River basin has found power-law degree distribution and scale-free character.
- Key nodes for contaminant propagation, flood vigilance and prevention have been identified.

## 1. INTRODUCTION

As shown in Sivakumar (2015) and Sivakumar *et al.* (2018), complex network analysis (CNA) is a natural paradigm for studying hydrologic systems, including the hydrologic cycle, rainfall and streamflow monitoring networks, global climate model outputs and river networks. Thus, there are some CNA approaches that deal with water distribution networks (WDNs). Many of those papers use community detection algorithms to segment the network (e.g., Diao *et al.* 2014; Giustolisi & Ridolfi 2014a, 2014b, Giustolisi *et al.* 2015; Campbell *et al.* 2016), while other papers focus on the vulnerability of the networks (e.g., Yazdani *et al.* 2011; Shuang *et al.* 2015) or on the role of topological features like edge betweenness (Simone *et al.* 2020). The results of these studies are encouraging and show the usefulness of CNA for analyzing, designing and managing WDN.

Other papers (e.g., Jha *et al.* 2015; Sivakumar & Woldemeskel 2015; Jha & Sivakumar 2017; Naufan *et al.* 2018; Kim *et al.* 2019) consider precipitation networks built using different ways of measuring the correlation among the rainfall time series of a set of rain gauge stations scattered in a region or from satellite-based rainfall data (e.g., Marwan & Kurths 2015). CNA approaches that study river runoff time series (Lange *et al.* 2018) also belong to this category. The main purpose of these papers is generally community detection (e.g., Agarwal *et al.* 2018).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Finally, a large number of CNA approaches correspond to studying streamflow dynamics using stream gauging station data (e.g., Tang *et al.* 2010; Sivakumar & Woldemeskel 2014; Braga *et al.* 2016; Serinaldi & Kilsby 2016; Han *et al.* 2018; Yasmin & Sivakumar 2018). The purpose of these studies is catchment classification based on the modules (i.e., the node communities) extracted from the network (Fang *et al.* 2017; Tumiran & Sivakumar 2021) as well as assessing the importance of the individual stations (e.g., Joo *et al.* 2021).

River basins, the basic unit of analysis in hydrology studies of rivers and streams, are complex systems, with multiple, interconnected, heterogeneous components, such as reservoirs, irrigated areas (IAs), drinking water treatment plants, wastewater treatments plants and stream gauging stations. Studying the pattern and nature of these connections and the interactions between the different components is a challenging task and requires a holistic perspective. CNA provides a methodology for representing and analyzing those interactions (e.g., Porse & Lund 2015, 2016). The density, average path length and diameter of the network provide information about the overall degree of connectivity and the geographical span of the river basin. The distribution of the degree and strength of the nodes provides information about the physical connectivity as well as the flow characteristics of each type of node. CNA can also compute a measure of the network efficiency based on the geodesic distances between the nodes. This and other CNA measures (like transitivity and centralization indexes) provide basic topological features of the river basin that define a sort of profile or signature of the river basin under study. These characteristics can thus be compared with those of other river basins. In addition, CNA can be used to compute centrality indexes for each node thus characterizing the individual nodes of the network. This can be done by looking at the node in- and out-closeness centrality, its betweenness centrality or its PageRank. Although these common centrality indexes are generally useful, we believe that computing a specific centrality index that takes into account the environmental impact of a potential contaminant spill occurring on a given node would be more interesting. The proposed environmental impact centrality index uses a simplified propagation model of a contaminant spill from any given source node. This contaminant propagation model is static and does not use differential equations, having similarities with those used to measure the spread of animal diseases along a network of cattle movements (e.g., Büttner *et al.* 2013; Payen *et al.* 2019), and also resembles the simulation of water quality parameters approach in Zuluaga *et al.* (2020). As will be detailed below, however, this centrality index differs from those in the literature as it has been designed specifically for a river context and from an environmental impact perspective.

To validate the proposed CNA approach, in this work it is applied to a specific river basin. This extends the study in Rodríguez-Alarcón & Lozano (2019), which used CNA to analyze the whole river system of Spain, including inter-basin water transfer infrastructure, but using a coarser level of analysis in which each river basin corresponded to a node of the network. In this paper, a zoom-in is performed on one such node so that this river basin gives rise to a whole complex network by itself.

The structure of the paper is as follows. Section 2 presents the methodology used, while Section 3 discusses the application of the proposed approach to the Guadalquivir River basin (GRB). Finally, Section 4 summarizes and concludes.

## 2. METHODOLOGY

The first step is to build the network that represents the river basin considered. This requires identifying all the nodes and links that form it. The nodes can be reservoirs, IAs, stream gaging stations, drinking water treatment plants (with their corresponding urban water use), wastewater treatment plants and run-of-river dams. A node representing the river mouth (RM) is also considered. About the edges, which are directed, data regarding their length and annual flow rate are required. The length of the edges is used when actual geodesic distances are computed. Unweighted geodesic distances can also be computed in which case each edge is assigned a notional unit length. The flow rates are used to compute the in- and out-strength of the nodes, i.e. the total in- and out-flows of each node. Because they may have associated demands (e.g., urban water demand), the total out-flow rate may be different from the total in-flow rate at some nodes.

Before computing any CNA measure, it may be interesting to analyze the data (e.g., the distribution of the edge lengths, the edge lengths for each type of node, the distribution of the in-flow rate of the nodes and the in-flow rate for each type of nodes). This provides useful information about the characteristics of the different nodes and edges that form the network.

As regards CNA measures, we should distinguish between global (i.e., network-level) and local (i.e., node-level) characterization measures. The global measures involve the density (i.e., the fraction of all possible edges are present in the network), the transitivity (i.e., the fraction of times that the existence of edges  $i \rightarrow j$  and  $j \rightarrow h$  is accompanied also by the existence of edge  $i \rightarrow h$ ), the average path length (i.e., the average geodesic distance between connected nodes), the diameter (i.e., the largest geodesic distance between connected nodes), the network efficiency (i.e., the inverse of the harmony mean of the geodesic distance

between all pairs of nodes), in- and out-degree as well as the in- and out-strength distributions and the centralization index of the in- and out-degrees (i.e., the extent to which the degree distribution corresponds to that of an in- or and out-star), among others. All these global measures are useful to characterize the network, identifying its main features, i.e., if it is sparse, the average and maximum distances between the nodes, if the network is a small world in which connected nodes are generally separated by a small number of edges and if the network is scale-free with a non-homogeneous distribution (i.e., a power-law distribution) so that most nodes have a small in- or out-degree but a few nodes (called hubs) have a large number of connections.

A second group of CNA measures corresponds to diverse centrality indexes that help identify the role and position of each node in the network. The centrality indexes most commonly used are those based on the in- and out-closeness of the node (i.e., the average number of edges required to reach a node from upstream or the average number of edges required to go from a node to any other node downstream), the node betweenness (i.e., the fraction of geodesic paths that go through a specific node) or the PageRank of a node (i.e., the ease with which a node can be reached by traversing the edges of the network randomly). Edge centrality indexes, like the edge betweenness centrality (i.e., the fraction of geodesic paths that make use of a specific edge), also provide information about which edges occupy a more central position in the network.

Although the conventional centrality indexes can be useful to characterize the different nodes of the network and thus identify nodes (and edges) that may require special attention in terms of monitoring and hydrologic management, for this specific type of network, we propose a novel centrality index that measures the centrality of the nodes based on the environmental impact of a potential spill on that node. The details of the computation of this environmental impact centrality index, as a weighted average of the contaminant concentration on all network edges as a consequence of the propagation of a contaminant spill on a given source node, are given below. The idea is that, given the pollution source node, all the edges that leave that node get contaminated and propagate the contamination downstream. However, when a contaminated stream merges with an uncontaminated one, a dilution occurs. Hence, the pollutant concentration is generally reduced as we move away from the source node. The edges upstream of the source node, of course, are not polluted.

Let

$E$	Set of (directed) edges of the network
$Q_{jk}$	Flow rate along edge $(j, k) \in E$
$Q_s^{spill}$	Flow rate of contaminant spill originated in node $s$
$c_s^{spill}$	Contaminant concentration of the spill originated in node $s$
$c_{jk}^s$	Contaminant concentration along any edge $(j, k) \in E$ as a consequence of a spill in node $s$
$L_{jk}$	Length of edge $(j, k) \in E$
$L = \sum_{(j,k) \in E} L_{jk}$	Total river basin length, i.e., the sum of the lengths of all edges $(j, k) \in E$
$EI_s$	Environmental impact centrality of node $s$

The environmental impact centrality of a given node  $s$  is calculated as follows:

$$EI_s = \frac{\sum_{(j,k) \in E} L_{jk} c_{jk}^s}{\sum_{(j,k) \in E} L_{jk}} = \frac{1}{L} \cdot \sum_{(j,k) \in E} L_{jk} c_{jk}^s$$

The contaminant concentration of the different edges as a result of a spill in node  $s$  is computed recursively using the mass balance equation and assuming a steady-state regime, i.e., it is assumed that the spill goes on forever. This assumption greatly simplifies the calculations and is adequate for the intended purpose of identifying the nodes with a larger environmental impact on the whole river basin network. To understand the mass balance equation, it is better to consider first the edges that enter and leave any given node  $i$ .

The variables  $Q_i^{spill}$  and  $c_i^{spill}$  in Figure 1 are zero except for node  $i = s$ , i.e., we assume that the spill occurs at a single source node  $s$ . Assuming that the concentration along all the edges leaving node  $i$  is the same, i.e.,  $c_{ij}^s = c_i^s \quad \forall j: (i, j) \in E$ , the general mass balance equation in node  $i$  is calculated as follows:

$$\sum_{(h,i) \in E} Q_{hi} c_{hi}^s + Q_i^{spill} c_i^{spill} = \left( Q_i^{spill} + \sum_{(i,h) \in E} Q_{ih} \right) \cdot c_i^s$$

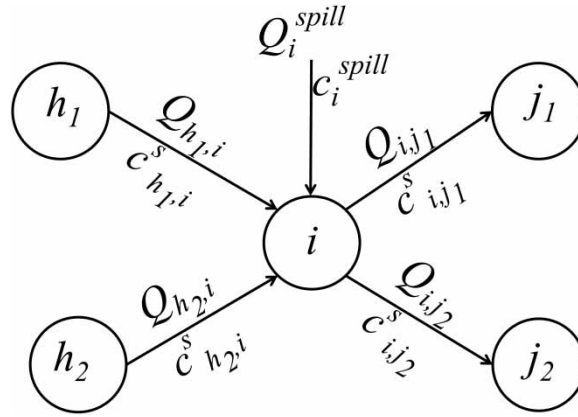


Figure 1 | Spill propagation at node  $i$ .

and therefore

$$c_{ij}^s = \frac{\sum_{(h,i) \in E} Q_{hi} c_{hi}^s + Q_i^{spill} c_i^{spill}}{Q_i^{spill} + \sum_{(i,h) \in E} Q_{ih}} \quad \forall j: (i, j) \in E$$

This equation is applied recursively starting with nodes with no incoming edges and following with the rest of the nodes downstream so that the concentration of the edges leaving a node can only be computed once the concentrations of all the edges entering that node have been computed.

To better understand how the above equation works, let us consider some of the possible cases:

- Node  $i$  has no incoming edges and  $i \neq s$ : In this case,  $Q_i^{spill} = c_i^{spill} = 0$ . Hence,

$$c_{ij}^s = 0 \quad \forall j: (i, j) \in E$$

- Node  $i$  is upstream of node  $s$ : In this case,  $c_{hi}^s = 0 \quad \forall h: (h, i) \in E$  and  $Q_i^{spill} = c_i^{spill} = 0$ . Hence,

$$c_{ij}^s = 0 \quad \forall j: (i, j) \in E$$

- Node  $i = s$ : In this case,  $c_{hi}^s = 0 \quad \forall h: (h, i) \in E$ . Hence,

$$c_{sj}^s = c_s^{spill} \cdot \frac{Q_s^{spill}}{Q_s^{spill} + \sum_{(s,j) \in E} Q_{sj}} \quad \forall j: (s, j) \in E$$

which means that the original contaminant concentration of the spill gets diluted by the total flow out of node  $s$

- Node  $i$  is downstream of node  $s$ : In this case,  $Q_i^{spill} = c_i^{spill} = 0$ . Hence,

$$c_{ij}^s = \frac{\sum_{(h,i) \in E} Q_{hi} c_{hi}^s}{\sum_{(i,j) \in E} Q_{ij}} \quad \forall j: (i, j) \in E$$

which means that the contaminant concentration downstream node  $i$  is a weighted average of the concentration along the incoming edges. In particular, if only one of the edges that enter node  $i$  is contaminated, then the flow of the other edges dilutes the contaminant concentration.

- Node  $i$  is downstream of node  $s$  and has only one incoming edge: In this case,  $Q_i^{spill} = c_i^{spill} = 0$  and

$$c_{ij}^s = c_{hi}^s \cdot \frac{Q_{hi}}{\sum_{(i,j) \in E} Q_{ij}} \quad \forall j: (i, j) \in E$$

Assuming that the flow leaving node  $i$  is equal to the flow entering the node, the above equation implies that  $c_{ij}^s = c_{hi}^s \quad \forall j: (i, j) \in E$ , i.e., there is no dilution in this case.

### 3. APPLICATION TO THE GRB

The Guadalquivir River, with a total length of 657 km, is the main river in Southern Spain. It supplies water to 4.3 million people and more than eight hundred thousand hectares of irrigated land. The GRB covers an area of 57,679 square km and has a total reservoir capacity of 8,155.4 hm<sup>3</sup> and an average flow rate (close to the RM) of 80.42 m<sup>3</sup>/s.

Figure 2 shows the different types of nodes that have been identified in the network together with their corresponding abbreviation. The figure also provides information on the number of nodes of each type. A total of 234 nodes have been considered. Note that the methodology is flexible and allows a coarser or a finer level of detail. This allows it to be adapted to the purpose of the study.

Figure 3 shows the actual GRB network, with its different types of nodes and with the width of the edges proportional to their total annual flow rate (i.e., the cumulative annual flow rate for year 2019). It can be seen how the width of the edges increases as we move downstream.

Figure 4 shows the average in-flows for each type of node. Note that although the RM node has a large in-flow value, it is not the largest. Actually, the largest in-flows correspond to RESV nodes, which are the main water supply source. The out-flows of those nodes (not shown in Figure 4) do not generally match these in-flows as water can be stored in them. Of the two main types of demand nodes, namely IA and DWTP + UWU, the former represents around double in terms of water consumption than the latter. However, while the latter demand is almost constant (i.e., fixed), the former can be regulated and eventually reduced in case of droughts. It can also be noticed that there is one IA that receives much more in-flow than the rest. Similarly, there is a large in-flow value for drinking water treatment plants and urban water usage (DWTP + UWU), which corresponds to Seville, the largest town along the Guadalquivir River (population of 690,000). Even more dispersion can be observed in the in-flows of reservoir (RESV) and streamflow gauging stations (SGS). In the latter, the smallest in-flows correspond to nodes upstream, while the larger in-flows correspond to downstream nodes, close to the RM. It is also

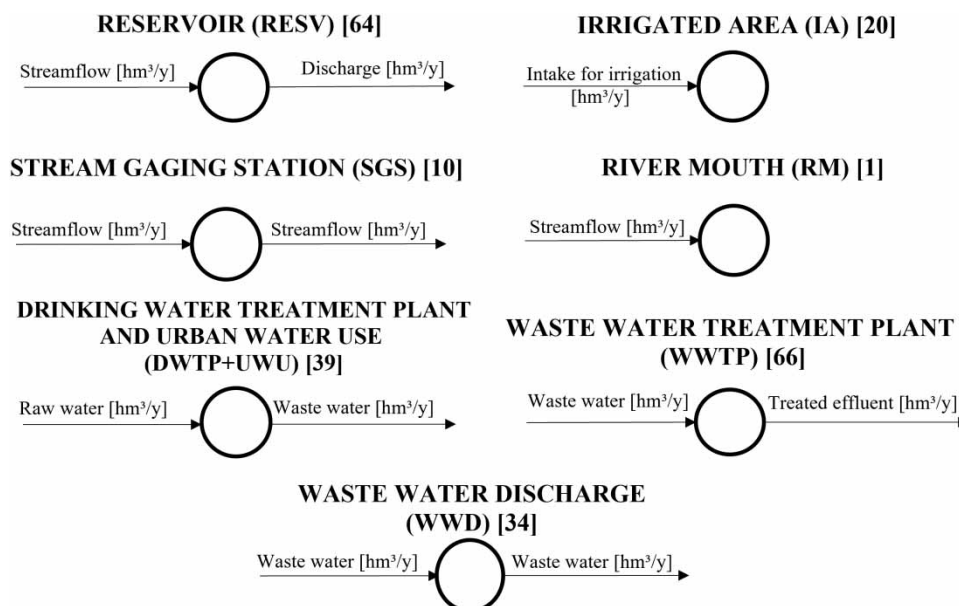


Figure 2 | Types of nodes in the GRB network.



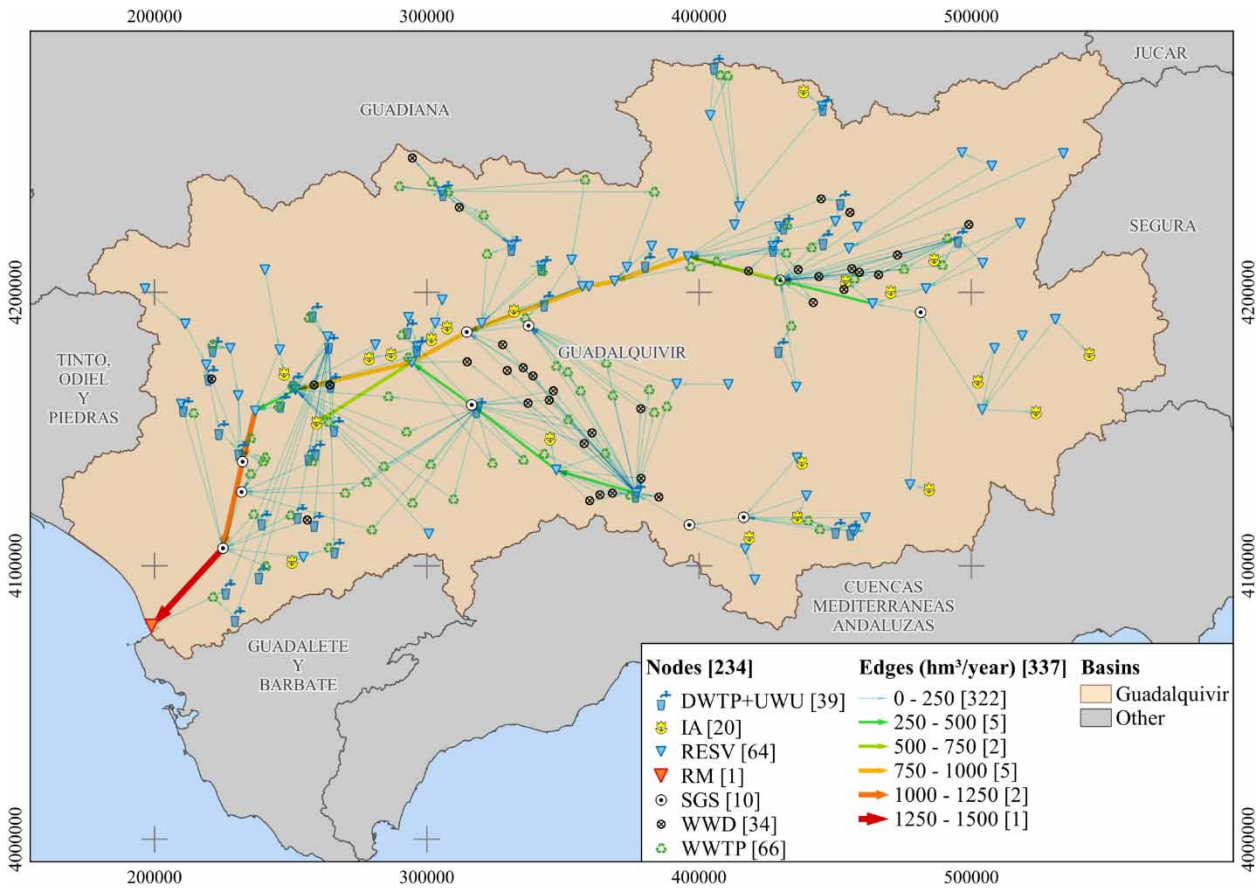


Figure 3 | Actual GRB network.

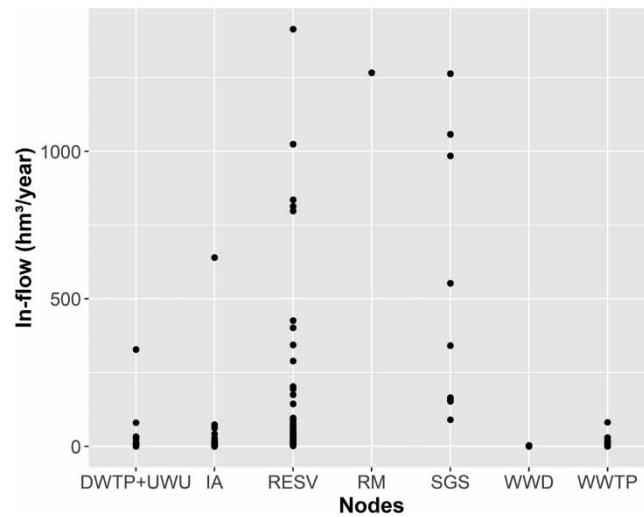
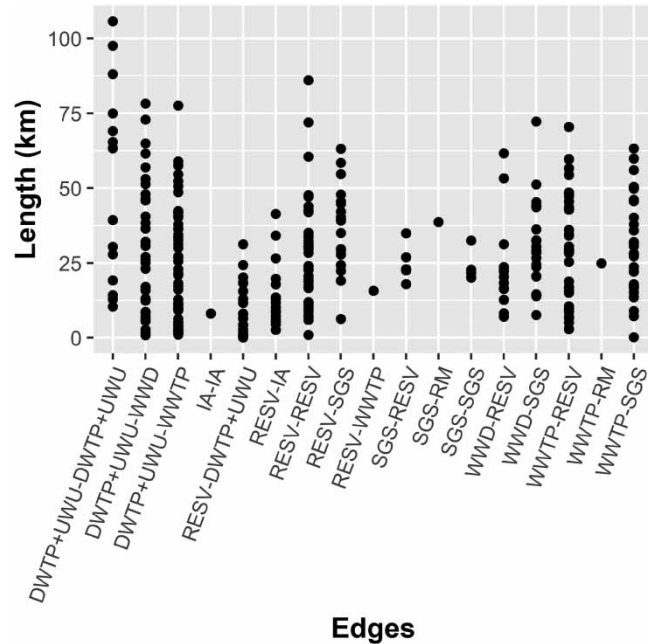


Figure 4 | Node in-flows for each type of node.

interesting to note that the in-flows of all the wastewater treatment plants (WWTP) are quite similar. Interestingly, they are also similar to those of the DWTP + UWU nodes.

Figure 5 shows the edge lengths of the different edge types. Note the relatively short links between RESV and DWTP + UWU nodes as well as between RESV and IA nodes. The other edge types have a larger dispersion of edge lengths. In general,



**Figure 5** | Edge lengths for different edge types.

there are two types of edges: those that correspond to natural streams (e.g., RESV-RESV, RESV-SGS, SGS-SGS and SGS-RM) and those that correspond to engineered water infrastructure (e.g., RESV-IA, RESV-DWTP + UWU and DWTP + UWU-WWD). The longest edges are of type DWTP + UWU-DWTP + UWU, which links and transports water between different cities and can be as long as 100 km. It is interesting to note also that the length of the RESV-IA edges is generally small and that is because precisely the construction of the RESV and the associated RESV-IA infrastructure were originally carried out as a single project (in other words, one of the main purposes of building the RESV was to provide water to those IAs and that influenced the selection of the RESV location not to be too far from the IA). It can be seen that, since each dot in [Figure 5](#) corresponds to a specific edge, a detailed analysis of each type of edge can be carried out.

[Table 1](#) shows some global metrics of the GRB network. The network density, transitivity and centralization indexes are all low. This indicates that the network is sparse, does not have a central point (like a star has) and closures of transitive triangles are infrequent. The sparsity of the network can be explained, on one hand, by the tree-like topology of river networks and, on the other, by the cost of building man-made water infrastructure edges. This leads to river basins not requiring a large number

**Table 1** | Some global metrics of the GRB network

Nodes	234
Edges	337
Density	0.0062
Transitivity	0.0155
Centralization index (in-degree)	0.0754
Centralization index (out-degree)	0.1139
Average path length (unweighted)	4.52
Diameter (unweighted)	14
Average path length (km)	105.09
Diameter (km)	432.23
Network efficiency (unweighted)	0.0022

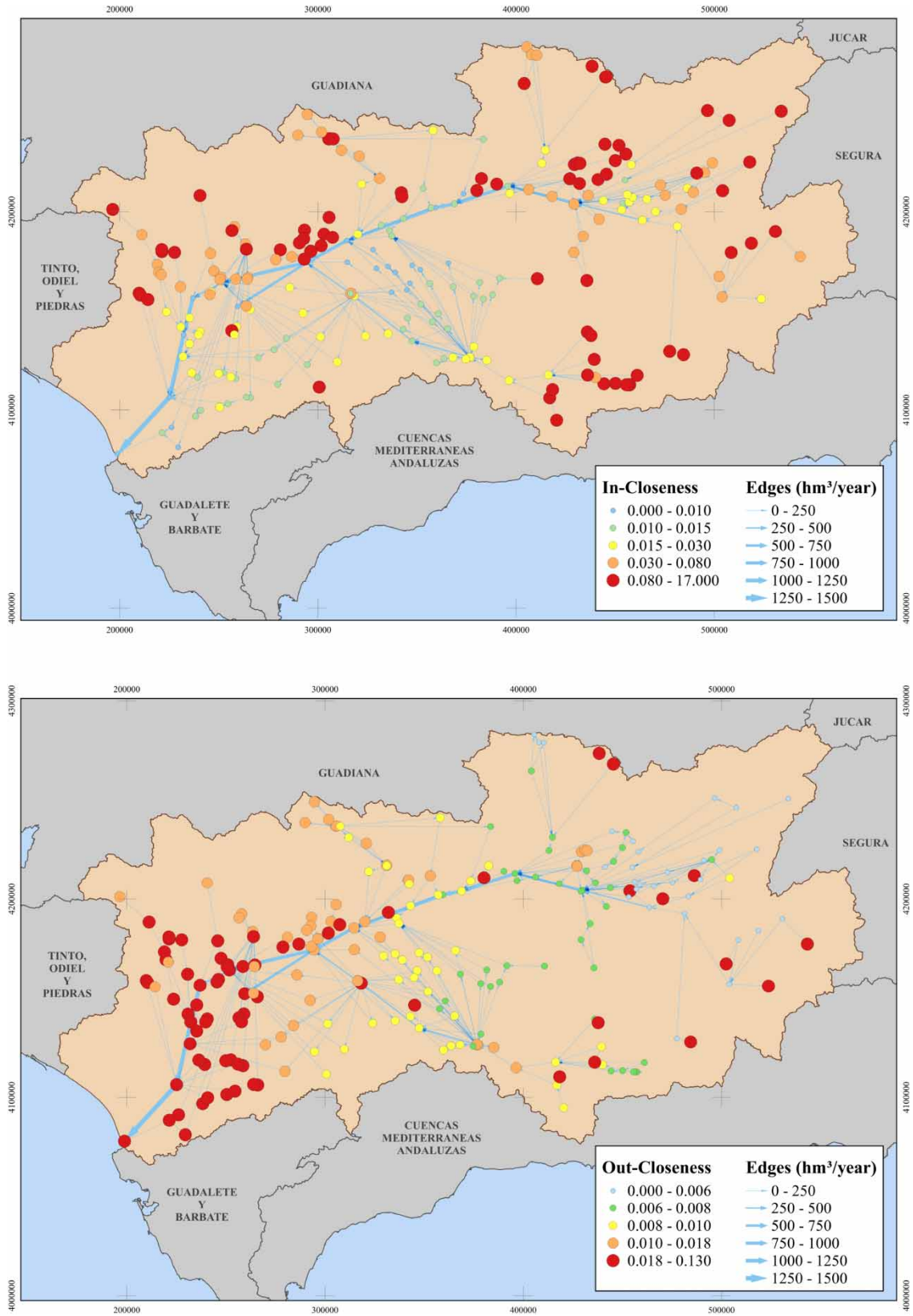
of edges to perform their function. We can refer to this concept as network effectiveness, to distinguish it from the more conventional, geodesic distance-based network efficiency concept commented below. The average path length between connected nodes is small (4.52), i.e., there are between four and five edges on average between any two connected nodes. The diameter is 14, which corresponds to a path involving up to 6% of the total number of nodes. Hence, it can be said that the GRB network is a small world and that it has an elongated topology. The average path length and diameter using actual distances are 105 and 432 km, respectively. The fact that there are only directed paths from each node to only a subset of nodes (i.e., only to those that are downstream) is reflected in the low network efficiency (0.0022 for the unweighted network). All these global characterization measures are of a descriptive nature and may be important when comparing different river basins.

As shown in the Supplementary material, the in- and out-degrees of the nodes follow a power-law distribution (with exponents of 2.74 and 2.72, respectively). The power law is a non-homogeneous, fat-tailed distribution, i.e., most nodes have a very low degree but a few nodes (labeled as hubs) have a very large degree. The power-law distribution is characteristic of scale-free networks and indicates that the average path length will not grow much with the network size, i.e., a river basin whose network had many more nodes should not have a much larger average path length. This is a prediction that, of course, needs to be verified by analyzing the networks associated with other river basins. It is remarkable also that a scale-free network appears in this river basin context formed by a majority of edges that correspond to natural streams but complemented with a number of man-made edges. The most common mechanism able to generate scale-free networks, i.e., preferential attachment (Barabási & Albert 1999), does not seem to be applicable in this context, so some alternative mechanism (to be investigated further) must be at play here. In any case, the scale-free nature of a river basin network also means that it is both 'robust yet fragile', i.e., robust to random failures but vulnerable to targeted attacks (Albert *et al.* 2000). In this context, random failures and targeted attacks should be interpreted as referring to both the man-made edges and to the natural streams, although the latter cannot be so easily 'deleted' from the network as the former. A clear example of this vulnerability is the breakdown or attack on a DWTP + UWU node, which would significantly affect the water supply of a whole city. For example, the DWTP + UWU node labeled Iznájar supplies drinking water to several nearby towns with a total population of 254,700.

As regards node centrality measures, Figures 6 and 7 show the in- and out-closeness centrality and the betweenness centrality and PageRank indexes, respectively. In both figures, the color of the nodes indicates their corresponding centrality, while the edge widths are proportional to their flow rates. Note that the in-closeness centrality decreases as we move downstream and the opposite happens with out-closeness centrality. Since the in-closeness is the inverse of the average geodesic distance from the nodes upstream of a given node, the largest values correspond to nodes that are close to the headwaters, often connected to RESV. In this regard, it has been mentioned before the economic advantages of having water infrastructures (such as IA and DWTP + UWU) as close as possible to RESV. In the case of the out-closeness centrality, which is the inverse of the geodesic distance of a node to those that are downstream, the largest values correspond to the nodes in the lower course of the river, closer to the RM. In this section of the river, the population density (and hence related water infrastructure) is higher and the nodes are relatively close to each other.

The betweenness centrality, as shown in Figure 7, is the largest for the nodes in the middle section of the river mainstream. That is because a large number of geodesic paths between nodes have to go through those nodes. These nodes generally correspond to SGS nodes located in the mainstream and those measure flow rates after the confluence from certain tributaries. These nodes can be important for monitoring and early warning of flood events. Because they generally have important and stable flow rates, they are also good potential locations for establishing run-of-river dams. As regards the PageRank index, it is computed following the direction of the edges and it measures the likelihood/ease with which a certain node can be reached following the directed paths of the network. Hence, it tends to increase as we move downstream. It is, however, affected by the pattern of connections and, particularly, by how the out-flow of a node is divided among the different edges that leave the node. This is because the calculation of the PageRank index assumes that each of the arcs that leave a node is selected with a probability proportional to its relative flow rate. When a terminal node with no out-going edges (e.g., the RM or IA nodes) is reached, then a new random node is selected to start a new directed path downstream. This directed 'random walk' process (random navigation is more appropriate in this case) is simulated for a long period and the fraction of times (i.e., the relative frequency) that each node is visited represents its PageRank index. Thus, as in the case of betweenness centrality, nodes with high PageRank tend to be also in the mainstream, but not in the middle section but in the lower course. These nodes are more





**Figure 6** | In-closeness centrality (top panel) and out-closeness centrality (bottom panel).

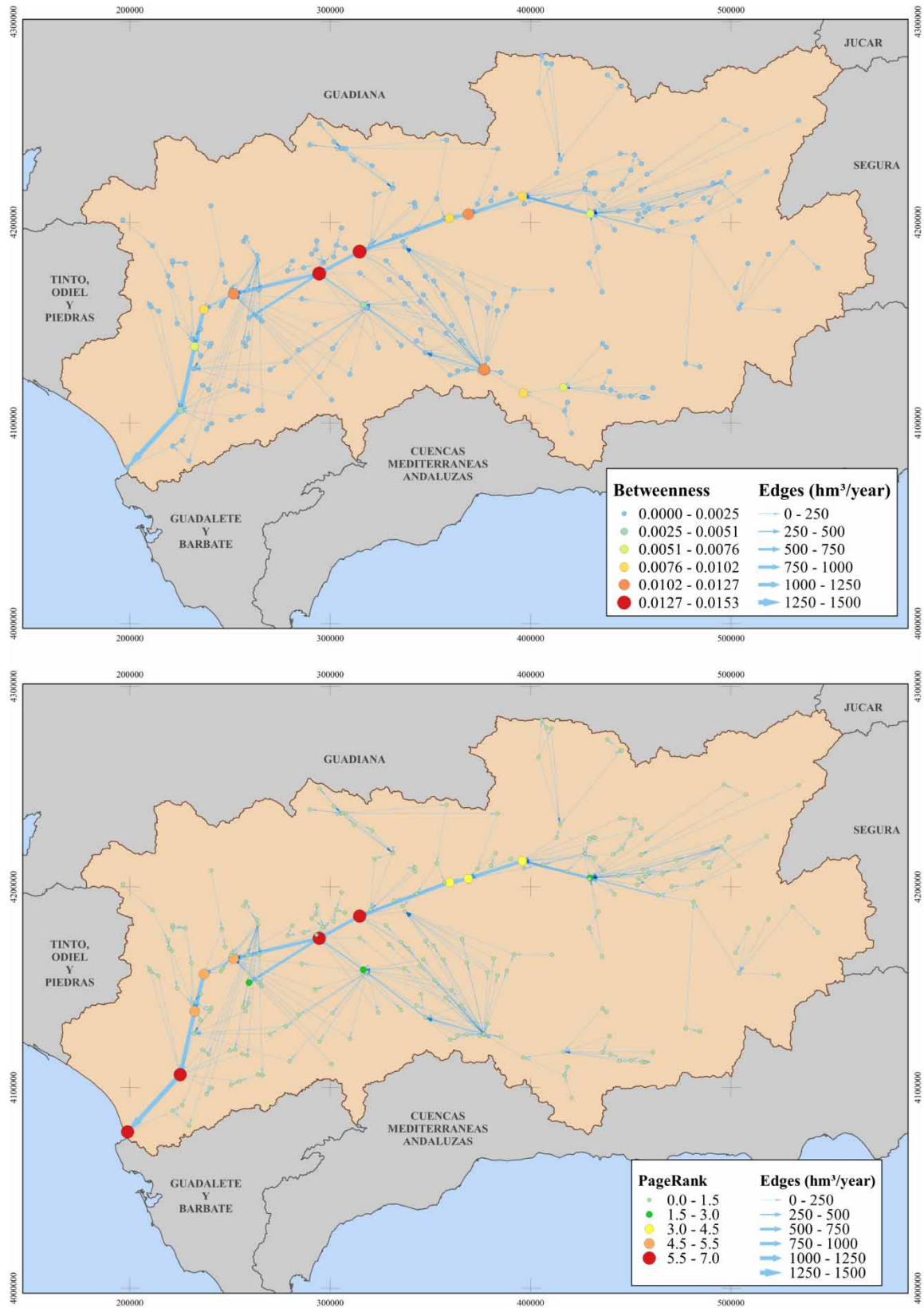


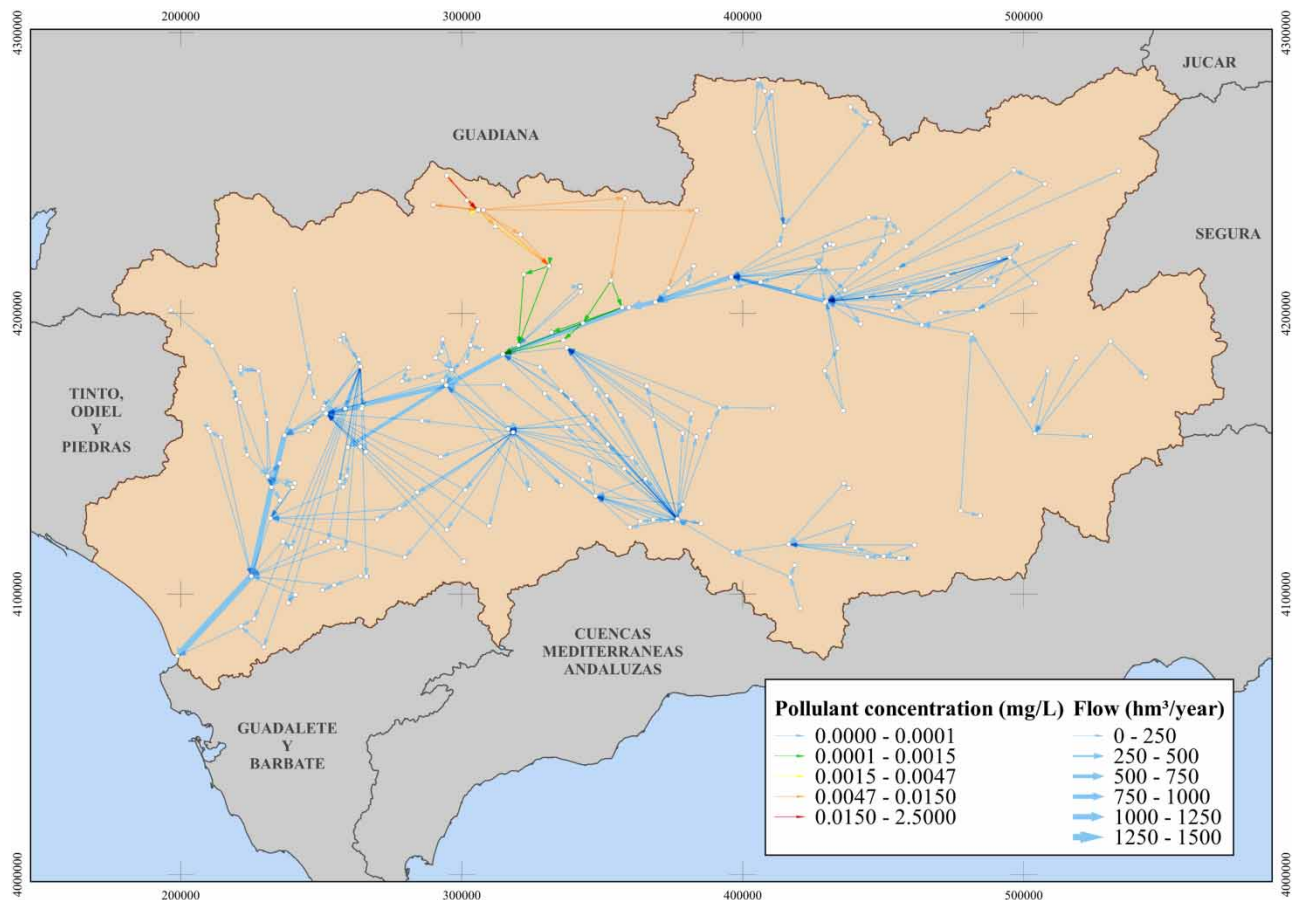
Figure 7 | Betweenness centrality (top panel) and PageRank index (bottom panel).

likely to be affected by floods as the network topology and the natural flow along the network tend to convey water toward them in a larger proportion.

As regards edge centrality measures, a main one is the edge flow rate, which has been shown in many previous figures and which increases as we move downstream. We can also consider the edge betweenness centrality. As shown in the Supplementary material, the edges with the highest betweenness lay in the middle section of the river with betweenness decreasing as we move upstream or downstream. Actually, the edges with high betweenness are located after the confluences of major tributaries and, as with the nodes with high betweenness, can play an important role in flood monitoring.

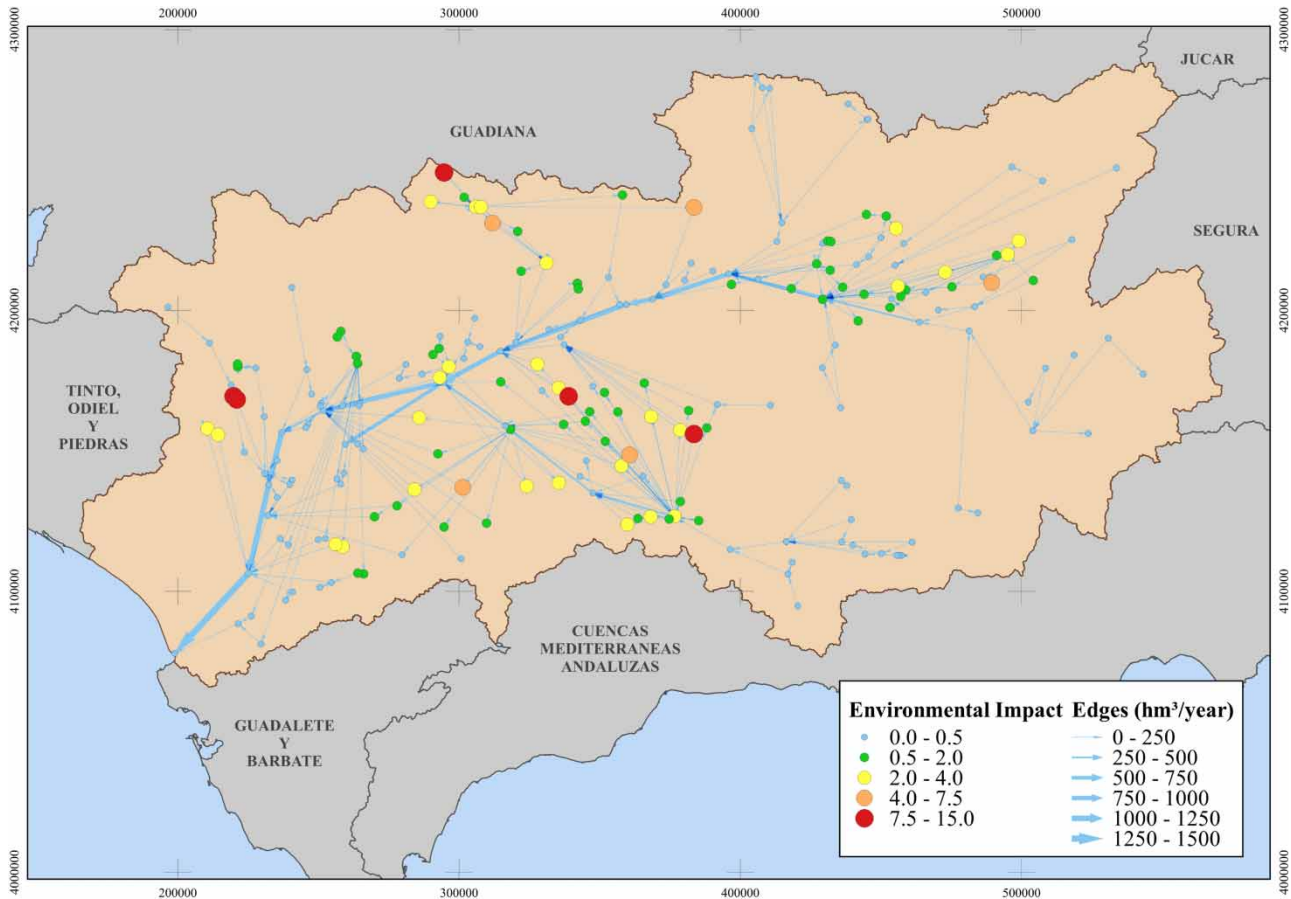
Finally, let us see an example of how the new node centrality index calculates the environmental impact that a contaminant spill originated in a given node would have on the whole river (measured as a weighted average of the concentration of the contaminant along each edge downstream of the spill source). Figure 8 shows an example of the simulation of the pollution propagation from a certain source node (namely, La Granjuela, Córdoba). The values of the spill flow and pollutant concentration at the source node are  $Q_s^{spill} = 30 \text{ L/s}$  and  $c_s^{spill} = 0.1 \text{ mg/L}$ , respectively. The edge width is proportional to its corresponding flow rate. The edge color depends on the computed pollutant concentration so that red and orange mean very high to high concentration, yellow and green mean intermediate to low concentration and blue means zero concentration. It can be seen that the pollutant concentration is high in the edges immediately downstream of the source node and that it gets quickly diluted as soon as the polluted stream reaches the main river stream.

Once the contamination propagation from a given source is simulated and the pollutant concentration along the river has been computed, the environmental impact index of that source node is determined (see Section 2). We have normalized the environmental impact centrality of the nodes so that their sum is equal to the number of nodes. In this way, a value greater than 1 means a significant environmental impact centrality, while a value lower than 1 indicates a below-average



**Figure 8** | Contamination propagation from a given source node. Please refer to the online version of this paper to see this figure in color: <http://dx.doi.org/10.2166/hydro.2022.148>.



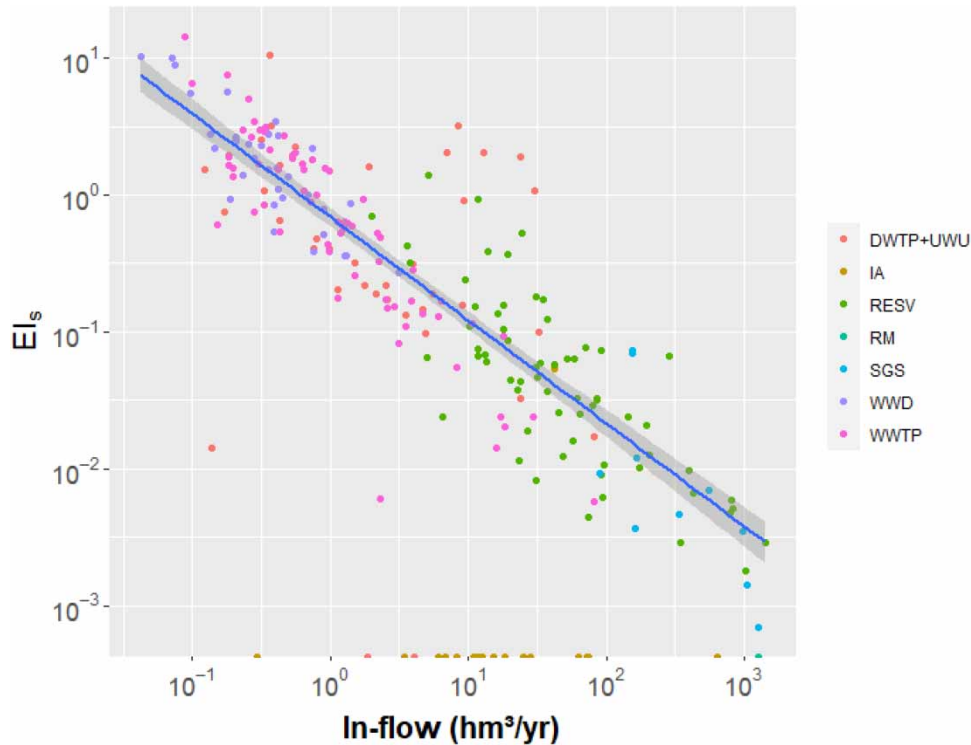


**Figure 9** | Environmental impact centrality.

environmental impact centrality. **Figure 9** shows the computed environmental impact centrality for all the nodes of the network. As before, the color of the nodes indicates their corresponding centrality, while the edge widths are proportional to their flow rates. Since a dilution of the contaminant concentration occurs whenever a contaminated stream merges with an uncontaminated one, the contaminant concentration decreases as we move downstream. The overall impact of the contaminant spill depends on the node on which it originates and is influenced in a complex way by the network topology and the associated flow pattern.

**Figure 10** shows the environmental impact centrality index  $EI_s$  versus the in-flow of each node, distinguishing by type of node. It can be seen that the nodes with the largest environmental impact are of types WWTP, WWD and DWTP + UWU. In the Supplementary material, the 20 nodes with the largest values of the environmental impact centrality are listed. Note that since the normalization used for this index is such that the average is 1, values much larger than 1 imply a significant environmental impact centrality. On the other hand, although for different reasons, nodes of types IA and SGS have low environmental impact centralities. In the case of IA, it is because these nodes are sinks and hence do not propagate the spill downstream. In the case of SGS, the low environmental impact is due to the fact that these nodes are always located in the mainstream and have significant in- and out-flow, which means that the contamination can get diluted relatively quickly and therefore does not propagate downstream very far. Hence, the environmental impact centrality index decreases with the in-strength of the nodes. Interestingly, this decrease seems to follow a power law function.

Finally, as shown in the Supplementary material, the proposed environmental impact centrality is not correlated with the conventional centrality indexes (namely, in- and out-closeness, betweenness centrality and PageRank). This indicates that this way of measuring centrality is different from the existing centrality indexes or, in other words, that the conventional centrality indexes are not suitable (because they have not been designed for that) for measuring the environmental impact due to a spill originated in a certain node.



**Figure 10** | Environmental impact centrality index for each type of node.

#### 4. CONCLUSIONS

In this paper, a CNA of a river basin is proposed. The first step involves identifying the different types of nodes in the river basin and their connections and collecting the data on the corresponding runoffs and streamflow data. It also involves deciding the level of detail that is appropriate, which depends on the purpose of the study. Once the network is built, its analysis using CNA tools is facilitated by the existence of a number of software packages like Pajek, UCINET, Gephi or igraph. Global characterization measures (like density, transitivity, centralization, average path length and network efficiency) can be easily calculated and provide useful information. In the application to the GRB, the network is sparsely connected, has small transitivity and centralization, small average path length, relatively large diameter (relative to the total number of nodes) and a low network efficiency. It is a conjecture whether these characteristics will be shared by other river basin networks.

The distribution of the edge lengths and flow rates in and out of the nodes provides additional information on the structure and topology of the network. Distinguishing by node and edge types is particularly interesting. Another important information that CNA provides is the distribution of the in- and out-degree and in- and out-strength. In the case of the GRB, the network has a power-law degree distribution, with most nodes having small degrees but a few nodes having very large degrees. This indicates a scale-free character and this is important because scale-free networks have many features that are of interest in a river basin context. Thus, for example, scale-free networks are small worlds (i.e., the average geodesic distances grow much slower than network size), are robust to random node failures and are vulnerable to targeted attacks. It is a conjecture whether other river basin networks also share this scale-free character.

Finally, both node and edge centrality indexes can be used to identify those that are more important for specific purposes (e.g., flood monitoring, flood prevention and run-of-river dam location). Thus, in the case of edges, both their flow rate and their betweenness centrality are of interest and can be easily calculated. Similarly, in the case of nodes, closeness and betweenness centralities can be calculated and used to rank the nodes and assess their relative position within the river basin network. Finally, a new node centrality index has been developed based on the environmental impact that a contaminant spill would have on the whole network. The larger the contaminant concentration on the different edges of the network as a result of the pollution propagation, the larger the environmental impact centrality of the source node of the spill. It has been observed that the different centrality measures mentioned above are not correlated. This is because they provide



information that is complementary and that together allows assessing the role and significance of the different elements (nodes and edges) of the network. In fact, we believe that these rich topological results can only be obtained using CNA.

The proposed methodology can, in principle, be applied to any river basin, provided the corresponding data are available. In this regard, there is flexibility in the level of detail that should be included in the analysis. The methodology harnesses the proven power of CNA tools and techniques, taking advantage of the fact that networks are the natural paradigm to study streams and river basins. The information and insights provided by this type of analysis represent a fresh perspective and an open venue for further research. Thus, apart from applying the methodology to other river basins, we think that developing additional, specific centrality indexes like the environmental impact centrality proposed in this paper is an interesting research topic. Also, vulnerability assessment and community detection can be carried out on this type of river basin network. Another promising topic is extending the static character of this study to multiperiod data. Finally, another issue, which we have also explored but which is definitely more complex, is that of modeling underground water streams and reservoirs as a complex network, on a different layer but nevertheless connected to the aboveground water streams and reservoirs modeled in this paper. This means building a multilayer network, which allows studying a whole new set of system features (see, e.g., Boccaletti *et al.* 2014; Kivelä *et al.* 2014). The problem with this approach is that underground water resources are not yet sufficiently known and inventoried and data are rather incomplete and unreliable but we hope this may change in the future.

## ACKNOWLEDGEMENTS

This research was carried out with the financial support of the Spanish Ministry of Science, Innovation and Universities, grant no. PGC2018-095786-B-I00. The funding source had no involvement in the writing of the report nor in the decision to submit the article for publication. The authors are also grateful to the three anonymous reviewers for their constructive remarks and suggestions, which have helped to improve the paper.

## DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

## REFERENCES

- Agarwal, A., Marwan, N., Maheswaran, R., Merza, B. & Kurths, J. 2018 [Quantifying the roles of single stations within homogeneous regions using complex network analysis](#). *Journal of Hydrology* **563**, 802–810.
- Albert, R., Jeong, H. & Barabási, A. L. 2000 [Error and attack tolerance of complex networks](#). *Nature* **406**, 378–382.
- Barabási, A. L. & Albert, R. 1999 [Emergence of scaling in random networks](#). *Science* **286**, 509–511.
- Boccaletti, S., Bianconi, G., Criado, R., del Genio, C. I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z. & Zanin, M. 2014 [The structure and dynamics of multilayer networks](#). *Physics Reports* **544**, 1–122.
- Braga, A. C., Alves, L. G. A., Costa, L. S., Ribeiro, A. A., de Jesus, M. M. A., Tateishi, A. A. & Ribeiro, H. V. 2016 [Characterization of river flow fluctuations via horizontal visibility graphs](#). *Physica A: Statistical Mechanics and Its Applications* **444**, 1003–1011.
- Büttner, K., Krieter, J., Traulsen, A. & Traulsen, I. 2013 [Static network analysis of a pork supply chain in Northern Germany – characterisation of the potential spread of infectious diseases via animal movements](#). *Preventive Veterinary Medicine* **110**, 418–428.
- Campbell, E., Izquierdo, J., Montalvo, I., Ilaya-Ayza, A., Pérez-García, R. & Tavera, M. 2016 [A flexible methodology to sectorize water supply networks based on social network theory concepts and multi-objective optimization](#). *Journal of Hydroinformatics* **18** (1), 62–76.
- Diao, K., Farmani, R., Fu, G., Astaraie-Imani, M., Ward, S. & Butler, D. 2014 [Clustering analysis of water distribution systems: identifying critical components and community impacts](#). *Water Science & Technology* **70** (11), 1764–1773.
- Fang, K., Sivakumar, B. & Woldemeskel, F. M. 2017 [Complex networks, community structure, and catchment classification in a large-scale river basin](#). *Journal of Hydrology* **545**, 478–493.
- Giustolisi, O. & Ridolfi, L. 2014a [New modularity-based approach to segmentation of water distribution networks](#). *Journal of Hydraulic Engineering* **140** (10), 04014049.
- Giustolisi, O. & Ridolfi, L. 2014b [A novel infrastructure modularity index for the segmentation of water distribution networks](#). *Water Resources Research* **50**, 7648–7661.
- Giustolisi, O., Ridolfi, L. & Berardi, L. 2015 [General metrics for segmenting infrastructure networks](#). *Journal of Hydroinformatics* **17** (4), 505–517.
- Han, X., Sivakumar, B., Woldemeskel, F. M. & Guerra de Aguilar, M. 2018 [Temporal dynamics of streamflow: application of complex networks](#). *Geoscience Letters* **5**, 10.
- Jha, S. K. & Sivakumar, B. 2017 [Complex networks for rainfall modeling: spatial connections, temporal scale, and network size](#). *Journal of Hydrology* **554**, 482–489.

- Jha, S. K., Zhao, H., Woldemeskel, F. M. & Sivakumar, B. 2015 Network theory and spatial rainfall connections: an interpretation. *Journal of Hydrology* **527**, 13–19.
- Joo, H., Kim, H. S., Kim, S. & Sivakumar, B. 2021 Complex networks and integrated centrality measure to assess the importance of streamflow stations in a river basin. *Journal of Hydrology* **598**, 126280.
- Kim, K., Joo, H., Han, D., Kim, S., Lee, T. & Kim, H. S. 2019 On complex network construction of rain gauge stations considering nonlinearity of observed daily rainfall data. *Water* **11**, 1578.
- Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y. & Porter, M. A. 2014 Multilayer networks. *Journal of Complex Networks* **2**, 203–271.
- Lange, H., Sippel, S. & Rosso, O. A. 2018 Nonlinear dynamics of river runoff elucidated by horizontal visibility graphs. *Chaos* **28**, 075520.
- Marwan, N. & Kurths, J. 2015 Complex network based techniques to identify extreme events and (sudden) transitions in spatio-temporal systems. *Chaos* **25**, 097609.
- Naufan, I., Sivakumar, B., Woldemeskel, F. M., Raghavan, S. V., Vu, M. T. & Liang, S. Y. 2018 Spatial connections in regional climate model rainfall outputs at different temporal scales: application of network theory. *Journal of Hydrology* **556**, 1232–1243.
- Payen, A., Tabourier, L. & Latapy, M. 2019 Spreading dynamics in a cattle trade network: size, speed, typical profile and consequences on epidemic control strategies. *PLoS ONE* **14** (6), e0217972.
- Porse, E. & Lund, J. 2015 Network structure, complexity, and adaptation in water resource systems. *Civil Engineering and Environmental Systems* **32** (1–2), 143–156.
- Porse, E. & Lund, J. 2016 Network analysis and visualizations of water resources infrastructure in California: linking connectivity and resilience. *Journal of Water Resources Planning and Management* **142** (1), 04015041.
- Rodríguez-Alarcón, R. & Lozano, S. 2019 A complex network analysis of Spanish river basins. *Journal of Hydrology* **578**, 124065.
- Serinaldi, F. & Kilsby, C. G. 2016 Irreversibility and complex network behavior of stream flow fluctuations. *Physica A: Statistical Mechanics and Its Applications* **450**, 585–600.
- Shuang, Q., Yuan, Y., Zhang, M. & Liu, Y. 2015 A cascade-based emergency model for water distribution network. *Mathematical Problems in Engineering* **2015**, 827816.
- Simone, A., Ciliberti, F. G., Laucelli, D. B., Berardi, L. & Giustolisi, O. 2020 Edge betweenness for water distribution networks domain analysis. *Journal of Hydroinformatics* **22** (1), 121–131.
- Sivakumar, B. 2015 Networks: a generic theory for hydrology? *Stochastic Environmental Research and Risk Assessment* **29** (3), 761–771.
- Sivakumar, B. & Woldemeskel, F. M. 2014 Complex networks for streamflow dynamics. *Hydrology and Earth System Sciences* **18** (11), 4565–4578.
- Sivakumar, B. & Woldemeskel, F. M. 2015 A network-based analysis of spatial rainfall connections. *Environmental Modelling & Software* **69**, 55–62.
- Sivakumar, B., Puente, C. E. & Maskey, M. L. 2018 Complex networks and hydrologic applications. In: *Advances in Nonlinear Geosciences* (Tsonis, A. A., ed.). Springer, Cham, pp. 565–586.
- Tang, Q., Liu, J. & Liu, H. 2010 Comparison of different daily streamflow series in US and China, under a viewpoint of complex networks. *Modern Physics Letters B* **24** (14), 1541–1547.
- Tumiran, S. A. & Sivakumar, B. 2021 Community structure concept for catchment classification: a modularity density-based edge betweenness (MDEB) method. *Ecological Indicators* **124**, 107346.
- Yasmin, N. & Sivakumar, B. 2018 Temporal streamflow analysis: coupling nonlinear dynamics with complex networks. *Journal of Hydrology* **564**, 59–67.
- Yazdani, A., Otto, R. A. & Jeffrey, P. 2011 Resilience enhancing expansion strategies for water distribution systems: a network theory approach. *Environmental Modelling & Software* **26**, 1574–1582.
- Zuluaga, V., García, J. M., Mesa, M. J. & Acosta-Minoli, C. A. 2020 Use of network theory to model water quality parameters in a hydrological network. *Journal of Physics: Conference Series* **1448**, 012006.

First received 8 November 2021; accepted in revised form 26 February 2022. Available online 10 March 2022