A complex network analysis of Spanish river basins

R. Rodríguez-Alarcón
Confederación Hidrográfica del Guadalquivir, Avda. del Brillante, 57, 14012 Córdoba, Spain.
E-mail: rrodriguez@chguadalquivir.es
S. Lozano
Escuela Superior de Ingenieros, University of Seville, Isla de la Cartuja, 41092 Seville, Spain
E-mail: slozano@us.es (corresponding author)

ESTE TRABAJO SE PUBLICÓ EN LA REVISTA
JOURNAL OF HYDROLOGY (2019) vol. 578, 124065

https://doi.org/10.1016/j.jhydrol.2019.124065

ABSTRACT

This paper carries out a study of Spanish river basins for the period 2008-2014 using complex network analysis (CNA) tools. The purpose is to gain insight into the structure and characteristics of the national hydrological system with an emphasis on the interconnectivity between the different river basins and the extent to which the current IBT can mitigate the rainfall imbalances of the country, particularly in a scenario of climate change. Apart from the size of the corresponding catchment areas, data on water demand for irrigation, industrial and municipal water supply, historical catchment inflows, reservoir capacity and historical levels, interbasin transfer infrastructures and historical interbasin transfer (IBT) flows, and ocean discharges were collected. A weighted directed network is built with all this information and a number of CNA characterization measures. It has been found that the system has a two-tier structure with a few river basins (hubs) that supply IBT flows to a relatively large number of receiver river basins. Some of those receiver river basins have incoming links from more than one source river basin. This diversification of IBT sourcing is necessary since the availability of water for IBT from a single river basin is not guaranteed. The CNA results also indicate that the IBT infrastructure has been designed to supply water from the river basins with surplus reservoir capacity to river basins with water

deficits. The community structure of the system has also been determined with some groups of river basins forming separate, self-sufficient subsystems and other communities minimally connected by IBT links. It can be concluded that the topology and characteristics of the network are a consequence of the imbalances created by the varying climatic conditions of the river basins, the water storage capacity provided by the existing reservoir infrastructure, the geographical and orographic constraints of the country and the high cost of establishing links between neighbouring river basins.

Keywords: River basins; interbasin transfers; complex network analysis; community structure; assortativity; efficiency

1. Introduction

Water is a scarce, strategic natural resource and its effective management has become a necessity and a priority for governments in rich and poor countries alike. Climate change has added pressure and difficulty to this complex task. Water resource management requires balancing supply and demand. The dynamic and uncertain character of supply implies that the necessary reservoirs and, sometimes, interbasin transfer (IBT) infrastructure, needs to be developed and integrated within the existing river basins. This requires planning and has significant social, economic and environmental impacts. Emanuel et al. (2015) highlight the importance of the IBT of surface water resources globally and point to the dearth of information on IBT locations and characteristics, both at the regional and global scales. They also provide an insightful and critical assessment on the effectiveness of IBT to mitigate climate-driven water shortages.

In this paper, the Spanish river basin system is studied using the paradigm of complex networks analysis (CNA). CNA is a relatively young, multidisciplinary field whose object is the study of the structure and functions of networks. Although networks (and graphs) have been studied by mathematics and operations research for a long time, it has not been until the last part of the past century that researchers noted the presence and importance of a number of features and phenomena that occur in the real-world networks. New concepts such as small world, scale-free networks, clustering, assortativity, preferential attachment, cascading failures, etc. were introduced and proved very useful to characterize and understand better the growth and behaviour of real-world networks. Moreover, the immense power of networks to capture and model relationships between physical objects or between logical entities has led to an explosion in the number of applications of these techniques, including social networks, transportation networks, power grids, logistics networks, trade networks, tourism flows networks, citations networks, business collaboration networks, etc. Newman (2003) and da Fontoura Costa et al. (2011) provide useful introductory papers on the theory and applications, respectively, of these techniques.

CNA has also been applied to hydrology for different purposes. In fact, Sivakumar (2015) argues that the ability of the CNA paradigm to model any type of connection represents a generic theory for studying the connections and evolution of hydrologic systems. Sivakumar et al. (2018) present an overview of applications of CNA in hydrology. Thus, for example, CNA has been applied to Water Distribution Networks (WDN). Some of these studies consider undirected unweighted networks while others consider weighted directed networks. The latter requires knowing the network flows' directions and hence the

hydraulics of the system, for example, through EPANET (Soldi et al. 2015). Two main research topics studied in these CNA WDN papers are the structural properties and vulnerability (e.g. Yazdani and Jeffrey 2011, 2012, Yazdani et al. 2011, Shuang et al. 2014, 2015) and the segmentation of WDN for creating District Metered Areas (DMA), controlling leakage or for minimizing contamination risks (e.g. Scibetta et al. 2013, Diao et al. 2013, 2014, Giustolisi and Ridolfi 2014a, 2014b, Giustolisi et al. 2015, Ciaponi et al. 2016).

CNA has also been used to study streamflow and river runoff time series (e.g. Tang et al. 2010, Sivakumar and Woldemeskel 2014, Halverson and Fleming 2015, Braga et al. 2016, Serinaldi and Kilsby 2016, Fang et al. 2017, Han et al. 2018, Lange et al. 2018, Yasmin and Sivakumar 2018). These studies basically aim at streamflow modelling, classification of catchments and predictions in ungauged basins.

A third group of CNA applications to hydrology involves rainfall gauging stations (Scarsoglio et al. 2013, Jha et al. 2015, Sivakumar and Woldemeskel 2015, Jha and Sivakumar 2017, Agarwal et al. 2018, Naufan et al. 2018), satellite-based rainfall data (Malik et al. 2012, Boers et al. 2013, 2014, Marwan and Kurths 2015, Ozturk et al. 2018, Sun et al. 2018) or drought severity time series (Konapala and Mishra 2017). These applications study the climate dynamics and spatio-temporal rainfall modeling.

Finally, and this is the type of applications most closely related to the one presented in this paper, CNA has also been used to study water resource networks, involving both natural and built components. In this type of studies the nodes of the networks are the stream flows, the reservoirs and the water infrastructure elements, including aqueducts, canals and water supply pipelines (e.g. Porse and Lund 2015, 2016). The main purpose of these studies is to visualize and to study the resilience of the corresponding network.

This paper uses the CNA methodology to model a national river basin system and its associated IBT infrastructure. To the best of our knowledge a similar study has not been carried out in the literature. The purpose of this research is to gain insight into the structure and properties of the national hydrological system. In particular, we are interested in the interconnectivity between the different river basins and the extent to which the current IBT can mitigate the rainfall imbalances of the country, particularly in a scenario of climate change.

The structure of the paper is the following. Section 2 describes the Spanish river basin system. Section 3 describes the methods used in this research while Section 4 presents and discusses the results of the study. Finally, Section 5 summarizes and concludes.

2. Description of the Spanish river basin system

Spain has 15 river basins, some of which discharge into the Atlantic Ocean and others into the Mediterranean Sea. As shown in Figure 1, four of them, namely Miño-Sil, Duero, Tajo and Guadiana, extend to Portugal. Except for the existing engineered inter-basin transfer infrastructure, in principle, these river basins can be considered as autonomous systems, managed by the corresponding public body, called "Confederación Hidrográfica". The hydrological imbalance between different parts of the country, with some areas (mainly in the northern part of the country) showing excess rainfall and the rest (central, south and eastern parts) suffering rainfall deficits, chronic in some cases, has made it necessary to build an IBT infrastructure, managed by central government, which in case of need can decree the transfer of specific amounts of water from one river basin to another. The purpose of these water transfers, which are often met with social and political opposition from the source region, is to alleviate water shortages in the receiving basin.

========== Figure 1 ==========

Table 1 show some characteristics and time series (for the period 2008-2014) of the different river basins, such as the size of the corresponding catchment area, total water demand (including irrigation, industrial and municipal water supply), catchment inflows, reservoir capacity and levels, inter-basin transfer flows and ocean discharges. Note that there is a large disparity between the river basins in terms of their size as well as in the catchment and transfer inflows they receive.

========== Table 1 ==========

Table 2 shows the existing IBT infrastructures, indicating the connected drainage basins connected, the link length, the link capacity and the average annual flows during the period 2008-2014. Note that one of these pipelines is bidirectional (i.e. the water can flow from either of the two ends to the other) and another is disused. Also, in some cases, there is more than one pipeline between a given pair of river basins.

========== Table 2 ==========

Figure 2 shows the total demand faced by each river basin and its three components (irrigation, industrial use and municipal water supply) for the period 2008-2104. The needs of the river basins with largest demands are basically for irrigation. The river basins with smaller demands are basically for municipal water supply.

========== Figure 2 ==========

3. Methods

In this section, the CNA metrics and algorithms used in this study are introduced and explained. The metrics are of two types: global (i.e. for the whole network) and local (i.e. node-level). The basic global network measures are its order (i.e. the number of nodes, n), its size (i.e. its number of arcs, m) and its density (i.e. the fraction of possible arcs that actually exist, ρ) For a weighted directed network like the one considered in this paper, its density is given by $\rho = \frac{m}{n \cdot (n-1)}$. Other global measures of interest are:

- Reciprocity (`igraph_reciprocity` function of igraph package): This is the probability that the opposite counterpart of an existing directed edge also exists, i.e. $R = \frac{1}{m}\sum_{i,j} A_{ij} \cdot A_{ji} = \frac{1}{m} Tr\left(A^2\right)$, where $A = \left(A_{ij}\right)$ is the binary adjacency matrix whose elements indicate whether each arc (i, j) exists or not. As its name indicates, this measure reflects the degree of bidirectionality of the connections. In an undirected network all the connections are bidirectional and the reciprocity is one. If none of the connections is bidirectional then the reciprocity is zero.
- Transitivity (`igraph_transitivity_undirected` function of igraph package): Similar to the clustering coefficient, the transitivity measures the probability that two neighbors of a vertex are connected but it is computed at the global level as the ratio of the triangles and the connected triples in the graph. This measure considers the network as undirected. The corresponding formula is

$$T = \frac{3 \cdot \sum_{h>j>i} A_{ij} A_{ih} A_{jh}}{\sum_{h>j>i} \left(A_{ij} A_{ih} + A_{ji} A_{jh} + A_{hi} A_{hj}\right)} = \frac{\frac{1}{2} Trace\left(A^3\right)}{\sum_{j>i} \left(A^2\right)_{ij}}$$

- Centralization index (`igraph_centralization_degree` function of igraph package): This is a measure of the extent to which the network topology resembles that of a star. The star is used as reference because its hub-spoke topology in which the central node is connected with all other nodes and these are only connected with the central node. In a star, the central node has $n-1$ connections and the $n-1$ nodes have just 1 connection. For an undirected network the centralization index is measured as $CI = \frac{\sum_i \left(k^{max} - k_i \right)}{(n-1)(n-2)}$, where k_i refers to the degree of each node i (i.e. the number of connections it has) and $k^{max} = \max_i k_i$ is the largest degree in the network. For a directed network, the centralization index can be computed using the in-degree or the out-degree of the nodes, $k_i^{in} = \sum_j A_{ji}$ and $k_i^{out} = \sum_j A_{ij}$, respectively.
- Average path length (`igraph_average_path_length` function of igraph package): This is the average geodesic distance between any pair of nodes. The geodesic distance between two nodes (`igraph_shortest_paths` function of igraph package) d_{ij} is the length of the shortest directed path from node i to node j. Therefore, the average path length is $\bar{d} = \frac{1}{n \cdot (n-1)} \cdot \sum_i \sum_{j \neq i} d_{ij}$. Note that if two nodes are not connected then, by convention, their geodesic distance is infinite. In that case, the average path length can be computed taking into account only between the pair of nodes that are effectively connected. Note also that in the case of a weighted network the geodesic distance generally uses the lengths of the arcs given by the arc weights W_{ij}. The lower the average path length, the easiest and the least costly is going (or moving goods or transferring information) between any pair of nodes in the network.
- Diameter (`igraph_diameter` and `igraph_diameter_dijkstra` functions of igraph package): This is the largest geodesic distance between connected nodes $D = \max_{\{(i,j):d_{ij}<\infty\}} d_{ij}$. It corresponds to the worst case, i.e. geodesic distance between the nodes that are farthest apart.
- Network efficiency: This is an alternative to the average path length that does not have any problem when some pairs of nodes are not connected (and hence their geodesic distance is infinite). Instead of using the arithmetic average, the network efficiency is computed as the harmonic mean of the

geodesic distances, i.e. $\eta = \frac{1}{n \cdot (n-1)} \cdot \sum_i \sum_{j \neq i} d_{ij}^{-1}$. The higher the geodesic distances between the nodes, the lower the efficiency of the movements (or the flow of goods or of the communications) over the network.

- Assortativity coefficient (`igraph_assortativity` function of igraph package): The assortativity of the network with respect to a certain node attribute is the Pearson correlation coefficient between the value of that attribute for the two nodes linked by each arc. Thus, if each node i has an attribute x_i then

$$r_x = \frac{\frac{1}{m} \cdot \sum_{\{(i,j):A_{ij}=1\}} x_i x_j - \left(\frac{1}{m} \cdot \sum_{\{(i,j):A_{ij}=1\}} x_i \right) \left(\frac{1}{m} \cdot \sum_{\{(i,j):A_{ij}=1\}} x_j \right)}{\sqrt{\frac{1}{m} \cdot \sum_{\{(i,j):A_{ij}=1\}} x_i^2 - \left(\frac{1}{m} \cdot \sum_{\{(i,j):A_{ij}=1\}} x_i \right)^2} \cdot \sqrt{\frac{1}{m} \cdot \sum_{\{(i,j):A_{ij}=1\}} x_j^2 - \left(\frac{1}{m} \cdot \sum_{\{(i,j):A_{ij}=1\}} x_j \right)^2}}$$

  This assortativity coefficient measures whether nodes with a large value of the attribute are connected with nodes that also has a large value of that attribute and nodes with a low value are connected with nodes with low value. If that is the case, the assortativity coefficient (i.e. the correlation) is positive and there is assortativity. Conversely, if nodes with a large value of the attribute are connected with nodes with a low value of the attribute then the correlation is negative and there is disassortativity.

- Degree-degree correlation (`igraph_avg_nearest_neighbor_degree` and `igraph_assortativity_degree` functions of igraph package): When the attribute considered in the assortativity coefficient above is the node degree then we can measure the degree-degree correlation between the nodes. When the network is directed then the attribute considered in each end of the links can be the in- or the out-degree of the corresponding node. The usual way of measuring this type of directed degree-degree correlations is through the distribution of the corresponding average nearest neighbor degree. Thus, for example, if at the origin node of a link we consider its out-degree and at the destination node of the link we consider its in-degree, we would be testing whether nodes with large out-degree are connected with nodes with large in-degree (assortativity) or with nodes with low in-degree (disassortativity). To that end, for each node i we

compute the average out-degree of the nodes that are connected to i, i.e. $k_{nn,i}^{out,in} = \frac{1}{k_i^{in}} \sum_j A_{ji} k_j^{out}$ and we average these values for all the nodes with the same value of the in-degree, i.e. $k_{nn}^{out,in}(k^{in}) = \frac{1}{\left|\left\{i : k_i^{in} = k^{in}\right\}\right|} \cdot \sum_{\left\{i:k_i^{in}=k^{in}\right\}} k_{nn,i}^{out,in}$. Plotting $k_{nn}^{out,in}(k^{in})$ as a function of k^{in} allows detecting the existence of assortativity or disassortativity between those two degree variables depending on whether the function is increasing or decreasing. If the degree variables whose correlation we want to compute are the out-degree of the origin node and the out-degree of the destination node then the procedure is similar and the corresponding mathematical formulae would be $k_{nn,i}^{out,out} = \frac{1}{k_i^{out}} \sum_j A_{ij} k_j^{out}$ and $k_{nn}^{out,out}(k^{out}) = \frac{1}{\left|\left\{i : k_i^{out} = k^{out}\right\}\right|} \cdot \sum_{\left\{i:k_i^{out}=k^{out}\right\}} k_{nn,i}^{out,out}$.

As regards the node-level CNA measures, apart from the in- and out-degree already mentioned, the following indexes can be computed:

- Strength (`igraph_strength` function of igraph package): This is the sum of the weights of the arcs that leave a node, out-strength, $s_i^{out} = \sum_j W_{ij}$ or of the arcs that enter the node, in-strength $s_i^{in} = \sum_j W_{ji}$
- Node out-efficiency: This is the harmonic mean of the geodesic distance from a given node to every other node in the network, i.e. $\eta_i^{out} = \frac{1}{n-1} \cdot \sum_{j \neq i} d_{ij}^{-1}$. It measures how easy it is to reach any node from that node. The average of the node out-efficiencies is equal to the network efficiency, i.e. $\eta = \frac{1}{n} \cdot \sum_i \eta_i^{out}$.
- Node in-efficiency: This is the harmonic mean of the geodesic distance to a given node from every other node in the network, i.e. $\eta_i^{in} = \frac{1}{n-1} \cdot \sum_{j \neq i} d_{ji}^{-1}$. It measures how easy it is to reach that node from

any node. The average of the node in-efficiencies is also equal to the network efficiency, i.e. $\eta = \frac{1}{n} \cdot \sum_i \eta_i^{in}$.

- Betweenness centrality (`igraph_betweenness` function of igraph package): The normalized betweenness centrality of a node i corresponds to the fraction of shortest paths between each pair of nodes of the network that pass through node i averaged for all such pairs of nodes (that are distinct from i). The unnormalized measures is the sum of the number of shortest path between any pair of nodes that pass through node i, i.e. $\beta_i^{unnormalized} = \sum_{\substack{j \neq p \\ j \neq i \neq p}} \sigma(j,p|i)$ where $\sigma(j,p|i)$ is the number of shortest paths from j to p that pass through i.

Although, as indicated above, the in- and out-degree and the in- and out-strength are local measures, their distribution is a global feature. Thus, for example, some complex networks have a scale free topology with a power law degree distribution in which most nodes have a very small degree and a few nodes (called hubs) has a very large degree. In other cases, like in random networks, the degree distribution is more homogeneous, with most node having a degree around the average degree.

When not all the nodes in a network are connected them the network can be partitioned into separate components (`igraph_clusters` function of igraph package). A component is a connected subgraph (i.e. a subset of nodes so that from each of them one can go to any other) of maximal cardinality (i.e. if a node that belongs to the component has a link with some other node, that other node also belongs to the component). In directed networks one can distinguish between strongly connected components and weakly connected components. The former take into account the direction of the arcs while the latter correspond to the associated undirected network formed ignoring the direction of the arcs.

Finally, most real word networks have a community structure and that is also a global feature of the network. A community, also labeled a module, corresponds to a cluster of nodes that have more connections among themselves what with nodes outside the community. There are many methods for community detection in complex networks (see Fortunato 2010 for a survey). Some of those methods, like Clauset et al. (2004), `igraph_community_fastgreedy` function of igraph package, or Blondel et al. (2008), `igraph_community_multilevel` function of igraph package, are based on the optimization of modularity (Newman 2006). The modularity function, `igraph_modularity`

function of igraph package, is a measure that indicates to what extent a given partition of the nodes represents the actual community structure of the network. Normally, to identify the community structure an undirected network is considered. Let a_c the number of edges within each community c and b_c the sum of the degree of all the nodes in community c. Then, the modularity is computed as $Q = \sum_c \left[\frac{a_c}{m} - \left(\frac{b_c}{2m} \right)^2 \right]$. The first term within the brackets is the fraction of edges that actually fall within each community while the second represents the expected number that should fall in that community according to the corresponding null model (i.e. a network with the same number of nodes and edges and the same degree distribution). The modularity of the partition that has all nodes as a single community is zero while the modularity of a partition in which each node constitutes a community is negative. The larger the value of the modularity of a partition, the better that partition matches the actual community structure of the network. Since the modularity cannot exceed one, a partition with a modularity value above 0.3 is generally considered a good fit.

4. Results

The Spanish river basin network has 15 nodes and 14 arcs. It is a weighted directed network where the weights considered depend on the purpose of the analysis. Thus, it can be either the physical length, the capacity of the transfer link or the transfer flow in a specific year or over the whole period. Table 3 shows some global network characterization measures. Note that there is no transitivity (i.e. there are no triangles in the network), which is reasonable as this would involve expensive and unnecessary connections between already connected nodes. There is a small reciprocity, due to the existence of some bidirectional links. Bidirectional links allow either of the two river basins connected to transfer water to the other. Almost all links are, however, unidirectional and hence the source and destination of the transfers are fixed. The network is sparse, i.e. the number of links is relatively small. Recall that each link involves a significant investment and hence their number is limited. The in-degree centralization index is very low but the out-degree centralization index is not so. Recall that the centralization index measures the extent to which the network resembles a star with a central, well-connected node, surrounded by many nodes that are not directly connected among themselves. Therefore, the Spanish river basin network shows some similarity with an out star.

========== Table 3 ==========

The diameter of the network is rather large (498.1 km). This is the largest geodesic distance between any pair of connected nodes. The diameter corresponds to the length of the longest transfer that may be carried out in the network. The average path length, i.e. the average of the geodesic distances (between connected nodes) is 167.3 km. Measured in terms of number of links, i.e. considering the unweighted network, the diameter is 4 and the average path length is 1.67. This means that the longest possible transfer would involve four consecutive IBT links. Table 4 shows the geodesic distances between each pair of nodes. The entries correspond to the unconnected pairs of nodes are left blank. The bold entry is the network diameter, which corresponds to the length of the shortest path between the TAJ and GDQ nodes. That shortest path is shown in red in Figure 3.

========== Table 4 ==========

========== Figure 3 ==========

It can be seen both in Figure 3 and in Table 4 that most pairs of nodes are unconnected. Actually, the network has four weakly connected components, one of them (namely DUE) is a singleton. The other three components are {TAJ, GDN, TOP, GDQ, CMA, GBB, JUC, SEG}, i.e. the river basins in the Centre and South of Spain, {EBR, CIC, COC, COR}, i.e. the river basins in the North and Northeast of Spain, and {GLC, MÑS}, in the Northwest of the country.

Figure 4 shows the community structure identified using the fast greedy hierarchical agglomerative algorithm proposed in Clauset et al. (2004). The right panel of the figure shows the dendrogram that summarizes the agglomerative process in which the nodes are progressively merged into larger clusters by greedily maximizing the modularity improvement in each step. The level of the dendrogram that gives the maximum value of modularity ($Q = 0.53$) corresponds to the seven clusters shown on the left panel. Note that two of the river basins, namely DUE and CIC belong to independent, single-node clusters. DUE has no connections with any other river basin. CIC, which lies on the Northeast of Spain, receives transfers just from EBR, which is grouped in a different cluster, together with COR and COC, both on the North of Spain. There is another three-node cluster that groups three river basins to the South, namely GDQ, CMA and GBB. The other river basin on the Southwest of pain, TOP, is clustered with the river basin that feeds it, GDN. The two Eastern reiver basins, JUC and SEG, which are the ones with largest rainfall deficits in the country, are grouped with the river basin that supplies to them, TAJ. It is interesting to note that there is a single arc connecting each of these three communities,

indicating that the IBT infrastructure aims at maintaining connectivity at minimum cost. Finally, the two Northwest river basins, GLC and MÑS, form a separate, self-sufficient subsystem. Recall that the Northwest is the Spanish region with the highest precipitation level.

========== Figure 4 ==========

Figure 5 shows the distribution of the in- and out-degree. It can be seen that, due to the high cost of the links, most nodes have a low in-degree. The out-degree is, however, relatively high for some donor river basins (namely TAJ and EBR), part of whose hydrological surplus can be transferred to other regions. Therefore, the in-degree distribution shows that the river basins receive transfer from one or at most two other basins while the out-degree distribution shows that some river basins with surplus water can transfer it to four other river basins (the out-degree of EBR is five but some of its outgoing links go to the same river basins, CIC and COR).

========== Figure 5 ==========

Very informative also is the strength distribution. Thus, Figure 6 shows, for the period 2008-2104, the in- and out-strength of the different river basins, in decreasing order. This allows the easy identification of those river basins that receive the largest amounts of water transfers and those that are the largest donors. Note that although EBR has a larger out-degree, it is TAJ that has the largest out-strength. These two river basins, and to a lesser extent GDN and CMA, constitute the main IBT sources. As regards in-strength, the main receivers of IBT flows are SEG, in Eastern part of Spain, followed by COR (in the North) and TOP (in the Southwest). These correspond with the river basins with the largest deficit between catchment inflows and local demand and that is why they require the largest IBT inflows. Note also that some river basins with non-zero in-degree (namely EBR, MÑS, COC and GDQ) have zero in-strength. This means that that the corresponding IBT infrastructure was not used at all during the period 2008-2014.

========== Figure 6 ==========

As regards degree-degree correlations, Figure 7 shows the average out-degree of linking neighbours versus the in-degree of a node and the average out-degree of linked neighbours versus the out-degree of

a node. The positive slope in the first case (left panel) is indicative of assortativity and means that nodes with larger in-degrees receive connections from nodes with higher out-degrees, i.e. the river basins with more incoming IBT links feed from the river basins with more outgoing IBT links. In the second case (right panel), the slope is negative and this disassortativity means that the river basins with more outgoing IBT links feed river basins with low outgoing IBT links. All this suggests that the system has a two-tier structure, with some river basins functioning as IBT sources (that supply multiple basins) and other river basins functioning as IBT receivers (that get IBT flows from more than one source). In this regard, note that although having an in-degree larger than one adds flexibility to the IBT system, it is costly and hence it is justified only when the availability of water for IBT from a single river basin is not guaranteed and hence it is necessary to diversify the IBT sources.

========== Figure 7 ==========

We can also look at the correlation between linked nodes in terms of some relevant node attributes such as, for example, catchment inflows, IBT inflows, reservoir capacity, volume stored in reservoir and total water demand. In complex network analysis this is called assortative mixing. As shown in Table 1, the reservoir capacity has not changed during the period under study and that the annual demand for each river basin is known on average. Therefore, for these two node attributes a single correlation coefficient for the whole period is computed. For the other river basin attributes the correlation can be computed also for each year. As shown in Table 5, most of the assortativity coefficients are negative, indicating a disassortative mixing. This means that nodes with large demand are often connected with nodes of low demand, nodes of large reservoir capacity are often connected with nodes of low reservoir capacity, nodes with large volume of water stored in reservoirs are connected with nodes with low volume of water stored, and nodes with large IBT inflows are connected with nodes with low IBT inflows. All this indicates that the IBT infrastructure has been designed to supply water from the river basins with surplus reservoir capacity to river basins with water deficits. In the case of catchment inflows there are both negative and positive assortativity coefficients, depending on the year. Most years (all but 2009 and 2010) there was a positive correlation between the catchment inflows of the connected river basins (both facing rainfall shortfall or surplus) while in 2009 and 2010 the correlation between the catchment inflows of each pair of connected river basins was negative (with one of the ends of the connection experiencing more rainfall than the other). In this regard, note that, because of the short-term character of the public perception of the water scarcity problem, the social and political

resistance to IBT normally arises when the donor river basin experiences low rainfall and occurs even if the receiving river basin experiences an even more acute water shortage.

========== Table 5 ==========

Another interesting CNA characterization measure is the network efficiency. In the case of the river basin network under study and using the actual lengths (in km) of the different links, it results a rather low network efficiency (0.0026). This is not surprising given to the high cost of building this type of links, which precludes the existence of many connections and shortcuts between the nodes. Figure 8 shows the in- and out-efficiency of the different nodes, in decreasing order, and their average (i.e. the network efficiency). The river basins with the highest in- and out-efficiency coincide and their out-efficiency is slightly larger than their in-efficiency, i.e. their outgoing links are generally shorter than their incoming links. There are some exceptions to this rule, however. Thus, for GDN, CIC, MÑS and DUE their out-efficiency is slightly larger than their in-efficiency.

========== Figure 8 ==========

Figure 9 shows the unnormalized betweenness centrality of each node versus its in- and out-degrees (left and right panels, respectively). The nodes with the highest betweenness centrality (namely GDQ, TOP and CMA) are located in the Southern part of the country and have an in-degree equal to one or two and an out-degree equal to one. These river basins often lie on the shortest path between two other river basins. This means that the water transferred between those two river basins should go through the river basin with the high betweenness centrality. This implies that the corresponding IBT infrastructure is doubly utilized: by the IBT flows whose destination is the river basin with high betweenness centrality and by the IBT flows that pass through that river basin with a different destination.

========== Figure 9 ==========

Finally, Figure 10 shows a visualization of the network where the nodes have been placed in the centroid of each river basin. The size of each node is proportional to the average catchment inflows (during the period under study 2008-2014). Thus, larger nodes correspond to river basins with large

catchment inflows. The node colour shows the average percentage of utilization of the corresponding reservoir capacity. As regards the arcs, their size is proportional to the average flow transferred while their colour shows the percentage of utilization of the corresponding IBT infrastructure. It can be seen that the largest IBT flows correspond to the TAJ-SEG, EBR-COR, EBR-CIC, GDN-TOP and CMA-GBB links. Of these, the EBR-COR and GDN-TOP are those that make a higher utilization of their IBT capacity. The two river basins which supply the largest IBT flows are TAJ and EBR, with the latter having a higher reservoir capacity utilization. Actually, the fact that TAJ supplies large amounts of IBT flows to many neighbouring river basins in spite of its relatively low level of reservoir utilization generates concerns in the population of that region. It is remarkable that DUE, which has the largest catchments inflows of all Spanish river basins, is an isolated node, unconnected with the rest of the system. This is due to the orography of this mountainous region, which makes it technically and economically unfeasible too build any IBT infrastructure. Note also that the North of Spain has more catchments inflows than the South and the East of the country, which are the driest regions. Most of these river basins have low levels of reservoir capacity utilization. It can also be seen in Figure 10 why, as indicated in Figure 9, TOP, GDQ and CMA have high betweenness centrality, followed by GDN and EBR. The two-tier structure of the system, with EBR and TAJ each acting as centre of an out star with peripheral nodes that in some cases have more than one incoming link, is also perceptible. The different connected components are clearly visible and also, with the help of Figure 4, the different communities that form the two largest components.

========== Figure 10 ==========

5. Conclusions

In this paper, CNA tools and concepts have been used to study and visualize the structure and interrelationships of the Spanish river basins and the existing IBT infrastructure. This is a novel approach that provides a holistic perspective of the system and at the same time allows assessing the role and centrality of the different nodes of the network. Thus, a number of network characterization measures can be computed, such as the density, reciprocity, centralization index, average path length and diameter, efficiency, etc. The picture these measures draw is of a sparse network, with low reciprocity, zero transitivity, four components (one of them formed by just a single unconnected node), relatively large average path length and diameter, and low efficiency. Most of these features can be explained by the high cost of establishing links between the nodes. The out-degree distribution is

broader (maximum out-degree=5) than the in-degree distribution (maximum in-degree=2), i.e. there are a few river basins (namely, EBR and TAJ) that each supply IBT flows to a relatively large number of receiver river basins while those receiver river basins have incoming links from at most two of these source river basins. The CNA results also indicate that the IBT infrastructure has been designed to supply water from the river basins with surplus reservoir capacity to river basins with water deficits and that the system has a two-tier structure, with some river basins functioning as IBT sources (that supply multiple basins) and other river basins functioning as IBT receivers (that get IBT flows from more than one source). This diversification of IBT sourcing implies that the availability of water for IBT from a single river basin is not guaranteed. A neat community structure involving seven small clusters (of two or three nodes generally) has also been identified. It is interesting to note that in most cases there is a single arc connecting these communities, indicating that the IBT infrastructure aims at maintaining connectivity at minimum cost. The in- and out-strength distributions clearly identify the main source and recipient river basins of the IBT flows. They also show that some river basins with incoming IBT links have not made any use of them during the period under study. The in- and out-efficiency of each node and its contribution to the overall network efficiency have also been determined and the nodes with high betweenness centrality have been identified.

In summary, this study has carried out a thorough and insightful analysis of the Spanish river basin system and their IBT interconnections. It can be concluded that the topology and characteristics of the network are a consequence of the imbalances created by the varying climatic conditions of the river basins, the water storage capacity provided by the existing reservoir infrastructure, the geographical and orographic constraints of the country and the high cost of establishing links between neighbouring river basins.

As a continuation of this research, a more fine-grained study considering the different sub-drainage basins that form each drainage basin is planned. Also, a similar study on a larger scale (e.g. the European river basins) would be desirable. In this regard, it is anticipated that the largest difficulty that would be faced is that of data collection.

Acknowledgements

This research was carried out with the financial support of the Spanish Ministry of Science and the European Regional Development Fund (ERDF), grant DPI2017-85343-P. The authors would like to thank the associate editor and two anonymous reviewers for their constructive and helpful comments and suggestions.

References

Agarwal, A., Marwan, N., Maheswaran, R., Merza, B. and Kurths, J., "Quantifying the roles of single stations within homogeneous regions using complex network analysis", *Journal of Hydrology*, 563 (2018) 802-810

Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E., "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and Experiment*, 10 (2008) P10008

Boers, N., Bookhagen, B., Marwan, N., Kurths, J. and Marengo, J., "Complex networks identify spatial patterns of extreme rainfall events of the South American Monsoon System", *Geophysical Research Letters*, 40 (2013) 4386-4392

Boers, N., Rheinwalt, A., Bookhagen, B., Barbosa, H.M.J., Marwan, N., Marengo, J. and Kurths, J., "The South American rainfall dipole: A complex network analysis of extreme events", *Geophysical Research Letters*, 41 (2014) 7397-7405

Braga, A.C., Alves, L.G.A., Costa, L.S., Ribeiro, A.A., de Jesus, M.M.A., Tateishi, A.A. and Ribeiro, H.V., "Characterization of river flow fluctuations via horizontal visibility graphs", *Physica A: Statistical Mechanics and its Applications*, 444 (2016) 1003-1011

Ciaponi, C., Murari E. and Todeschini, S., "Modularity-Based Procedure for Partitioning Water Distribution Systems into Independent Districts", *Water Resources Management*, 30 (2016) 2021-2036

Clauset, A., Newman, M.E.J. and Moore, C., "Finding community structure in very large networks", *Physical Review E*, 70 (2004) 066111

da Fontoura Costa, L., Oliveira Jr, O.N., Travieso, G., Rodrigues, F.A., Ribeiro Villas Boas, P., Antiqueira, L., Palhares Viana, M. and Correa Rocha, L.E., "Analyzing and modeling real-world phenomena with complex networks: a survey of applications", *Advances in Physics*, 60, (2011) 329-412

Diao, K., Zhou, Y. and Rauch, W., "Automated Creation of District Metered Area Boundaries in Water Distribution Systems", *Journal of Water Resources Planning and Management*, 139, 2 (2013) 184-190

Diao, K., Farmani, R., Fu, G., Astaraie-Imani, M., Ward S. and Butler, D., "Clustering analysis of water distribution systems: identifying critical components and community impacts", *Water Science & Technology*, 70, 11 (2014) 1764-1773

Emanuel, R.E., Buckley, J.J., Caldwell, P.V., McNulty, S.G. and Sun, G., "Influence of basin characteristics on the effectiveness and downstream reach of interbasin water transfers: displacing a problem", *Environmental Research Letters*, 10 (2015) 124005

Fang, K., Sivakumar, B. and Woldemeskel, F.M., "Complex networks, community structure, and catchment classification in a large-scale river basin", *Journal of Hydrology*, 545 (2017) 478-493

Fortunato, S, "Community detection in graphs", *Physics Reports*, 486 (2010) 75-174

Giustolisi, O. and Ridolfi, L., "New Modularity-Based Approach to Segmentation of Water Distribution Networks", *Journal of Hydraulic Engineering*, 140, 10 (2014a) 04014049

Giustolisi, O. and Ridolfi, L., "A novel infrastructure modularity index for the segmentation of water distribution networks", *Water Resources Research*, 50 (2014b) 7648-7661

Giustolisi, O., Ridolfi, L. and Berardi, L., "General metrics for segmenting infrastructure networks", *Journal of Hydroinformatics*, 17, 4 (2015) 505-517

Halverson, M.J. and Fleming, S.W., "Complex network theory, streamflow, and hydrometric monitoring system design", *Hydrology and Earth System Sciences*, 19 (2015) 3301-3318

Han, X., Sivakumar, B., Woldmeskel, F.M. and Guerra de Aguilar, M., "Temporal dynamics of streamflow: application of complex networks", *Geoscience Letters*, 5:10 (2018)

Jha, S.K., Zhao, H., Woldemeskel, F.M. and Sivakumar, B., "Network theory and spatial rainfall connections: an interpretation", *Journal of Hydrology*, 527 (2015) 13-19

Jha, S.K. and Sivakumar, B., "Complex networks for rainfall modeling: spatial connections, temporal scale, and network size", *Journal of Hydrology*, 554 (2017) 482-489

Konapala, G. and Mishra, A., "Review of complex networks application in hydroclimatic extremes with an implementation to characterize spatio-temporal drought propagation in continental USA", *Journal of Hydrology*, 555 (2017) 600-620

Lange, H., Sippel, S. and Rosso, O.A., "Nonlinear dynamics of river runoff elucidated by horizontal visibility graphs", *Chaos*, 28 (2018) 075520

Malik, N., Bookhagen, B., Marwan, N. and Kurths, J., "Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks", *Climate Dynamics*, 39 (2012) 971-987

Marwan, N. and Kurths, J., "Complex network based techniques to identify extreme events and (sudden) transitions in spatio-temporal systems", *Chaos*, 25 (2015) 097609

Naufan, I., Sivakumar, B., Woldemeskel, F.M., Raghavan, S.V., Vu, M.T. and Liong, S.Y., "Spatial connections in regional climate model rainfall outputs at different temporal scales: Application of network theory", *Journal of Hydrology*, 556 (2018) 1232-1243

Newman, M.E.J., "The Structure and Function of Complex Networks", *SIAM Review*, 45, 2 (2003) 167-256

Newman, M.E.J., "Modularity and community structure in networks", *Proceedings of the National Academy of Sciences*, 103 (2006) 8577-8582

505 Ozturk, U., Marwan, N., Korup, O., Saito, H., Agarwal, A., Grossman, M.J., Zaiki, M. and Kurths, J.,
506 “Complex networks for tracking extreme rainfall during typhoons”, *Chaos*, 28 (2018) 075301
507 Porse, E. and Lund, J., “Network structure, complexity, and adaptation in water resource systems”,
508 *Civil Engineering and Environmental Systems*, 32, 1-2 (2015) 143-156
509 Porse, E. and Lund, J., “Network Analysis and Visualizations of Water Resources Infrastructure in
510 California: Linking Connectivity and Resilience”, *Journal of Water Resources Planning and*
511 *Management*, 142, 1 (2016) 04015041
512 Scarsoglio, S., Laio, F. and Ridolfi, L., “Climate Dynamics: A Network-Based Approach for the
513 Analysis of Global Precipitation”, *PLoS ONE*, 8, 8 (2013) e71129
514 Scibetta, M., Boano, F., Revelli, R. and Ridolfi, L., "Community detection as a tool for complex pipe
515 network clustering", *EPL*, 103 (2013) 48001
516 Serinaldi, F. and Kilsby, C.G., “Irreversibility and complex network behavior of stream flow
517 fluctuations”, *Physica A: Statistical Mechanics and its Applications*, 450 (2016) 585-600
518 Shuang, Q., Zhang, M. and Yuan, Y., “Node vulnerability of water distribution networks under
519 cascading failures”, *Reliability Engineering and System Safety*, 124 (2014) 132-141
520 Shuang, Q., Yuan, Y., Zhang, M. and Liu, Y., “A Cascade-Based Emergency Model for Water
521 Distribution Network”, *Mathematical Problems in Engineering*, 2015 (2015) 827816
522 Sivakumar, B., “Networks: a generic theory for hydrology?”, *Stochastic Environmental Research and*
523 *Risk Assessment*, 29, 3 (2015) 761-771
524 Sivakumar, B. and Woldemeskel, F.M., “Complex networks for streamflow dynamics”, *Hydrology and*
525 *Earth System Sciences*, 18, 11 (2014) 4565-4578
526 Sivakumar, B. and Woldemeskel, F.M., “A network-based analysis of spatial rainfall connections”,
527 *Environmental Modelling & Software*, 69 (2015) 55-62
528 Sivakumar, B., Puente, C.E. and Maskey, M.L., “Complex Networks and Hydrologic Applications”, in:
529 Tsonis, A.A. (Ed.), *Advances in Nonlinear Geosciences*, Springer, Cham (2018) 565-586
530 Soldi, D., Candelieri, A. and Archetti, F., “Resilience and vulnerability in urban water distribution
531 networks through network theory and hydraulic simulation”, *Procedia Engineering*, 119 (2015)
532 1259-1268
533 Sun, A.Y., Xia, Y., Caldwell, T.G. and Hao, Z., “Patterns of precipitation and soil moisture extremes in
534 Texas, US: A complex network analysis”, *Advances in Water Resources*, 112 (2018) 203-213
535 Tang, Q., Liu, J. and Liu, H., “Comparison of different daily streamflow series in US and China, under
536 a viewpoint of complex networks”, *Modern Physics Letters B*, 24, 14 (2010) 1541–1547

Yasmin, N. and Sivakumar, B., “Temporal streamflow analysis: Coupling nonlinear dynamics with complex networks”, *Journal of Hydrology*, 564 (2018) 59-67

Yazdani, A. and Jeffrey, P., “Complex network analysis of water distribution systems”, *Chaos*, 21, 1 (2011) 016111

Yazdani, A., Otto, R.A. and Jeffrey, P., “Resilience enhancing expansion strategies for water distribution systems: A network theory approach”, *Environmental Modelling & Software*, 26 (2011) 1574-1582

Yazdani, A. and Jeffrey, P., “Water distribution system vulnerability analysis using weighted and directed network models”, *Water Resources Research*, 48 (2012) W06517

List of figures and table captions

572

573

574

575 Figure 1. Spanish river basins

576

577

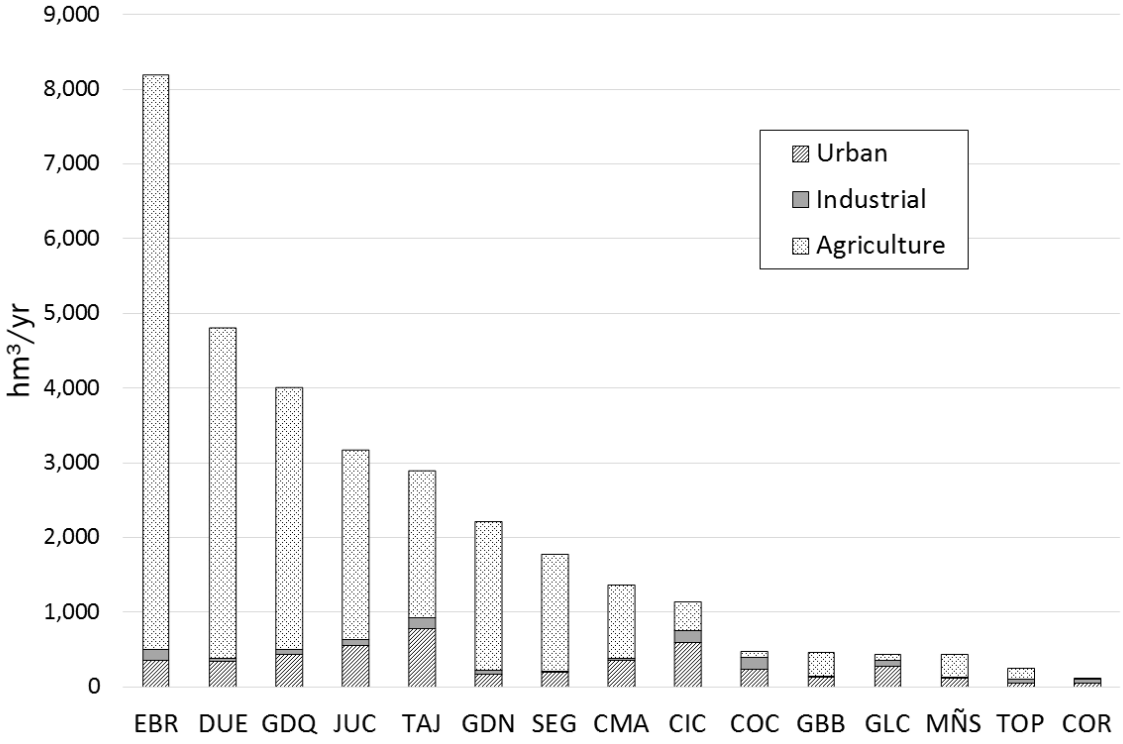


Figure 2. Volume and composition of river basin demands (2008-2014)

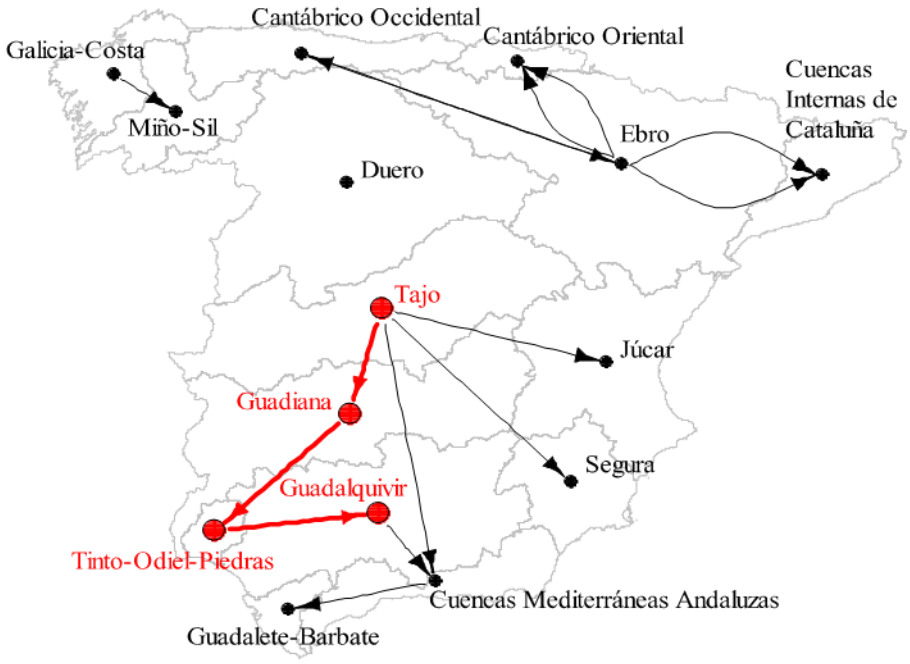


Figure 3. Shortest path between TAJ and GDQ nodes (its length corresponds to the network diameter)

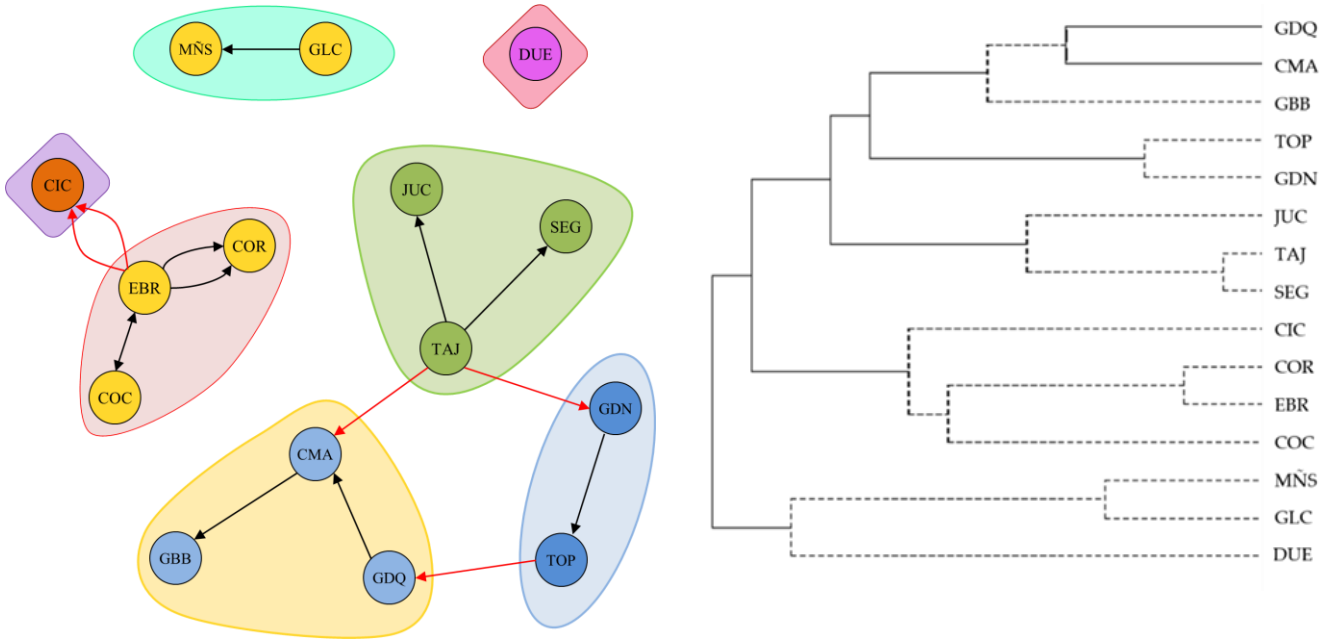
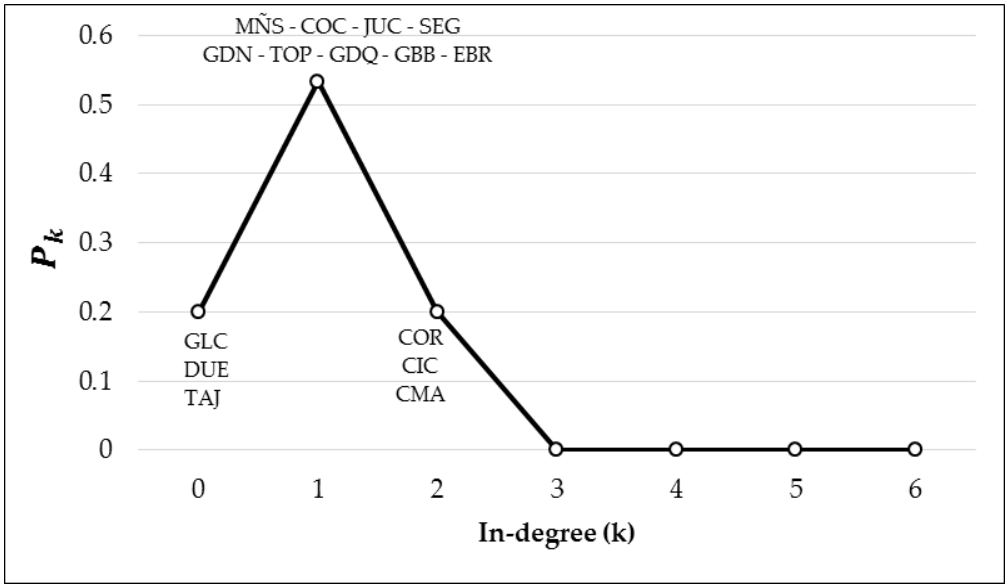


Figure 4. Community structure (fast greedy algorithm)



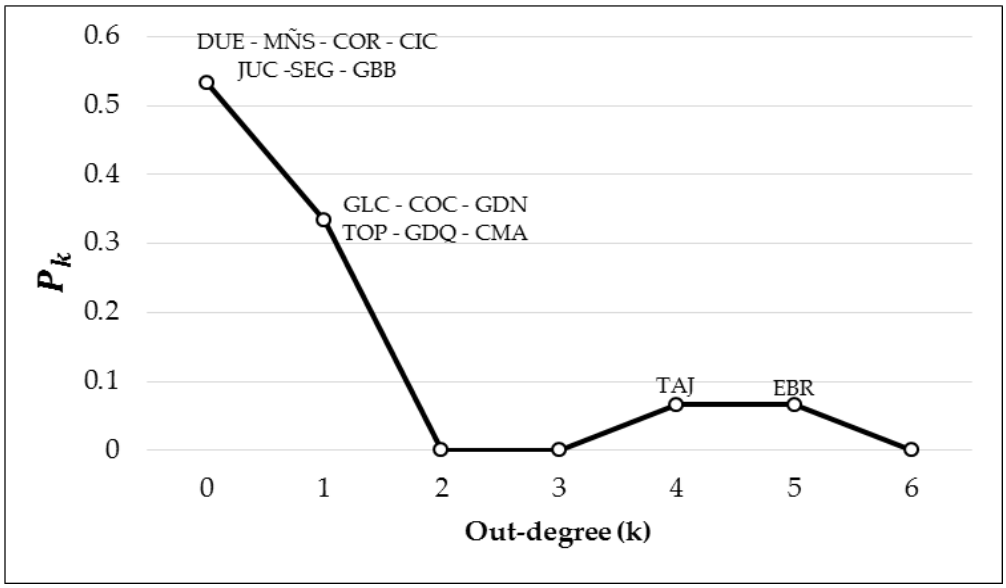
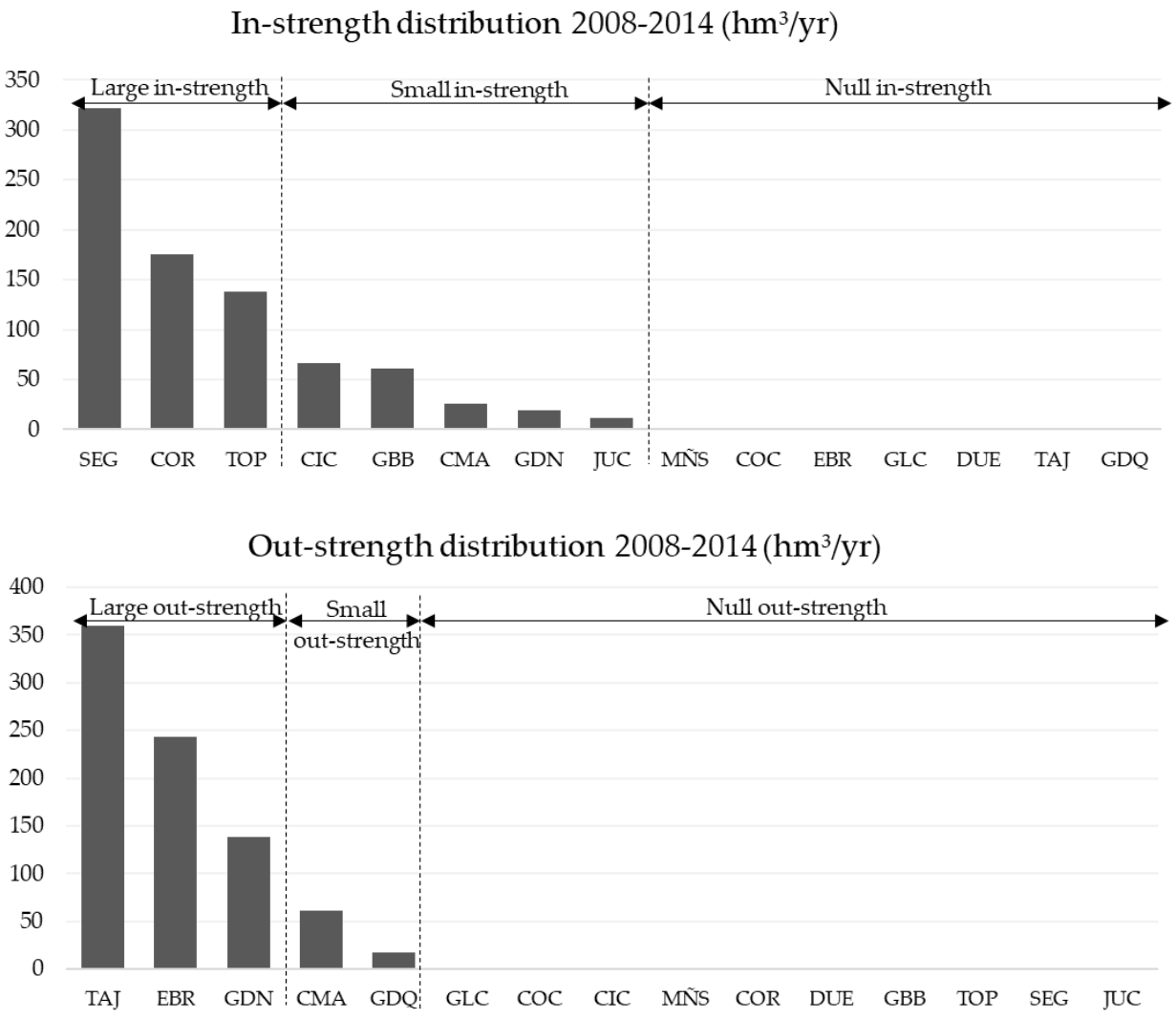


Figure 5. In- and out-degree distribution

599

600



601 Figure 6. In- and out-strength distribution (average 2008-2014)

602

603

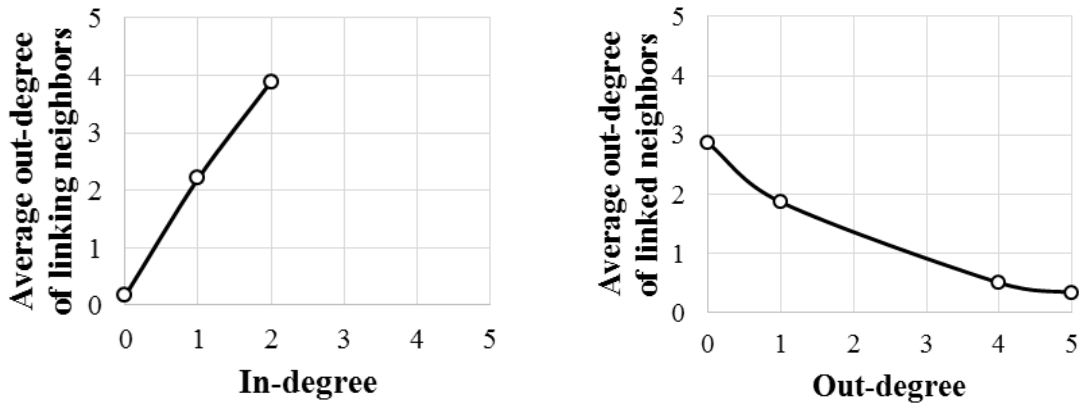
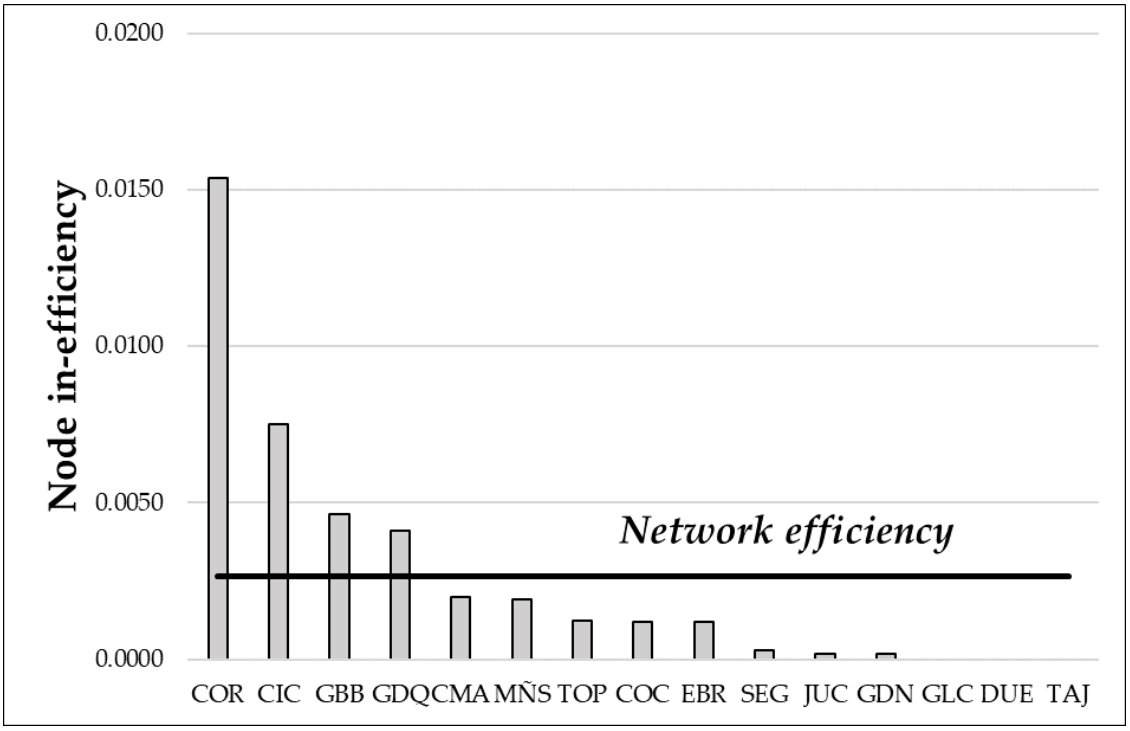


Figure 7. Degree-degree correlations



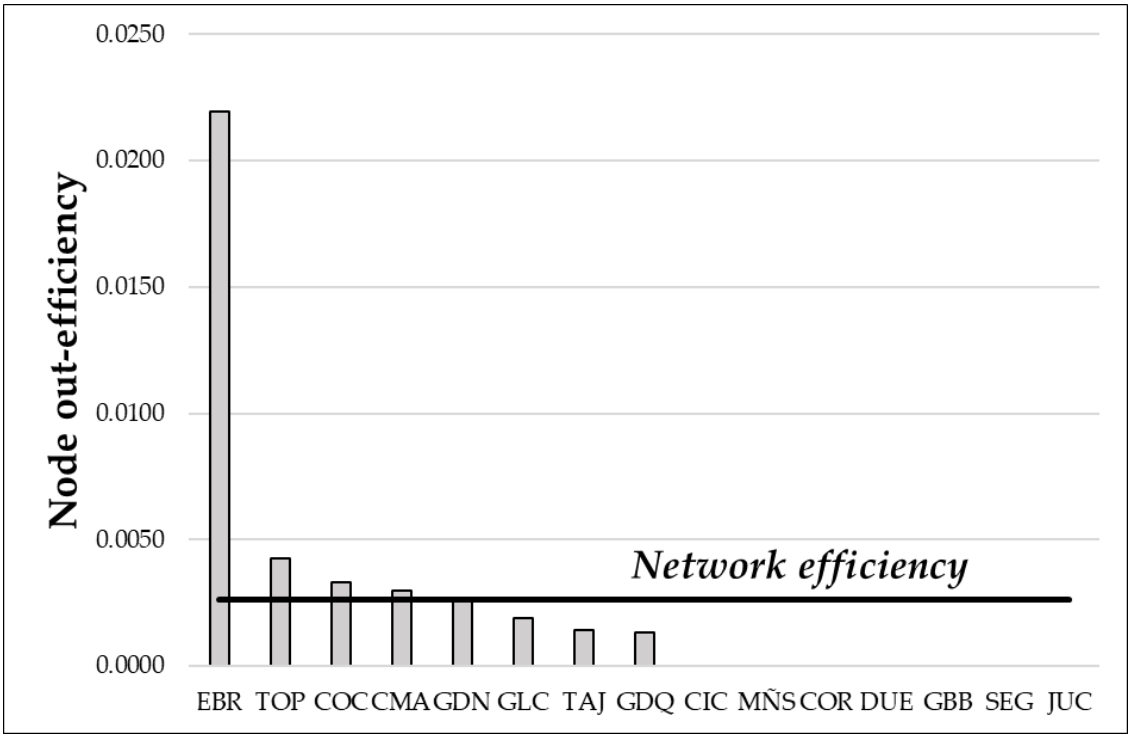


Figure 8. Network and node in- and out-efficiency

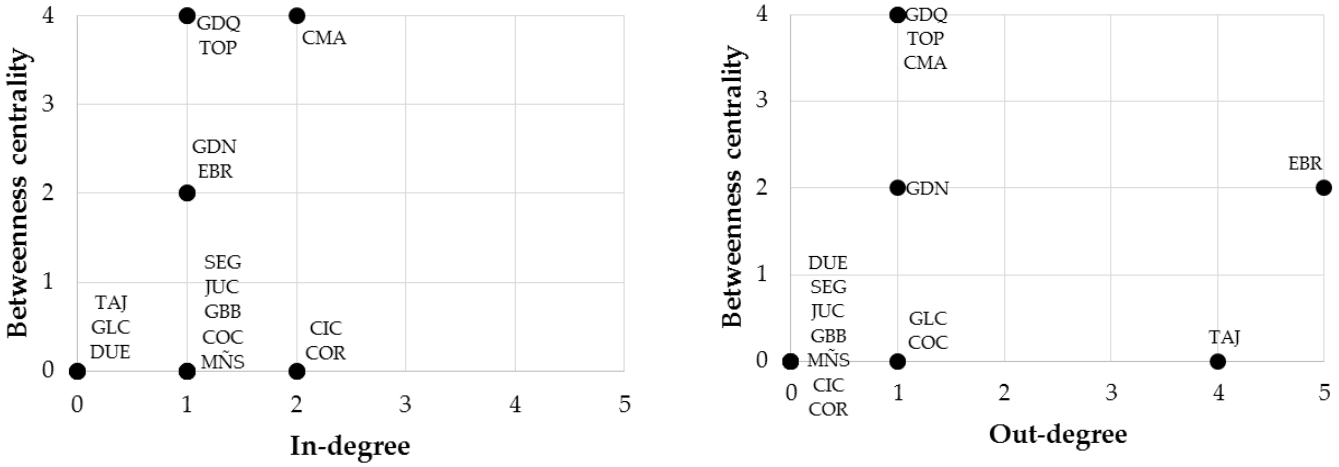


Figure 9. Betweenness centrality versus in- and out-degree

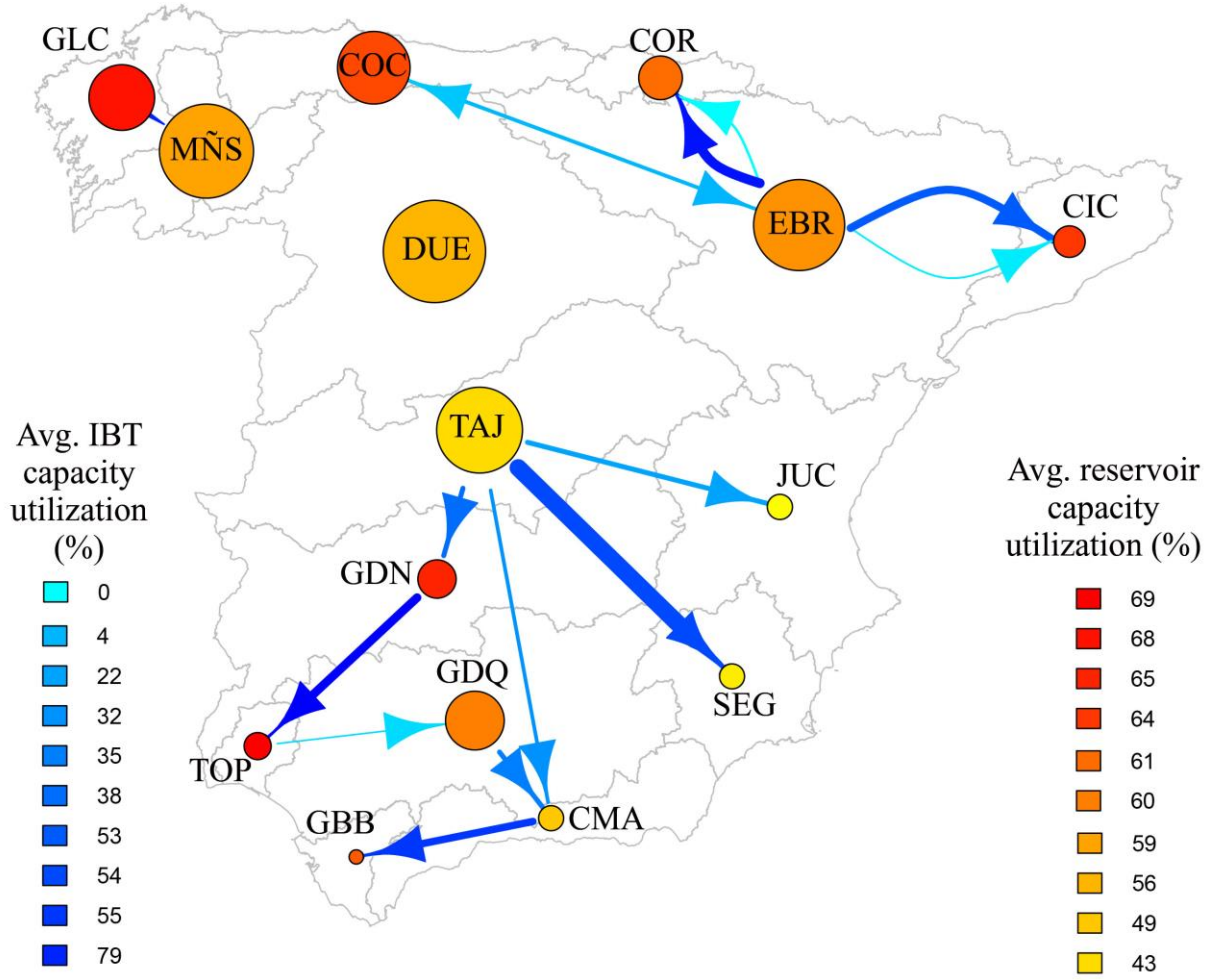


Figure 10. Visualization of the network. Node size proportional to average catchment inflows. Node colour shows average reservoir capacity utilization. Arc width proportional to average IBT flow. Arc colour shows IBT capacity utilization.

630 Table 1. Basic data of Spanish river basins

					2008-2014				
#	River basin	Label	Catchment area (km^2)	Reservoir capacity (hm^3)	Average annual catchment inflows (hm^3)	Average annual transfer inflows (hm^3)	Average annual discharges to ocean (hm^3)	Average annual demand	Average volume stored in reservoirs (hm^3)
1	Cuencas Internas de Cataluña	CIC	16,495.43	740.00	1,162.75	66.54	898.66	1,046.40	514.86
2	Miño.Sil	MÑS	17,603.37	3,030.00	10,733.26	1.58	12,721.47	436.02	1,695.29
3	Galicia.Costa	GLC	13,079.92	684.00	6,112.31	0.00	5,221.04	434.92	388.71
4	Cantábrico Oriental	COR	5,783.90	79.00	2,314.36	175.56	1,118.01	273.03	50.86
5	Cantábrico Occidental	COC	17,417.20	554.00	6,393.38	0.95	8,084.39	769.00	325.86
6	Duero	DUE	78,889.78	7,507.00	12,700.55	0.00	8,680.15	4,800.00	4,194.00
7	Tajo	TAJ	55,780.76	11,012.00	8,940.42	0.00	6,085.06	2,800.02	5,361.71
8	Guadiana	GDN	55,512.10	9,266.00	2,514.89	18.75	815.54	2,217.13	5,914.57
9	Guadalquivir	GDQ	57,187.61	8,101.00	4,184.71	0.00	1,988.30	4,007.73	5,242.71
10	Cuencas Mediterráneas Andaluzas	CMA	17,941.97	1,177.00	1,063.29	17.28	1,085.54	1,392.60	682.29
11	Guadalete.Barbate	GBB	5,956.26	1,651.00	333.11	500.87	NA	438.25	1,139.00
12	Tinto.Odiel.Piedras	TOP	4,762.34	229.00	861.47	138.53	1,000.00	264.68	165.86
13	Segura	SEG	19,024.80	1,141.00	728.67	321.54	7.96	1,722.50	568.57
14	Júcar	JUC	42,729.38	3,330.00	743.90	11.16	453.88	3,240.81	1,358.00
15	Ebro	EBR	85,562.79	7,511.00	10,060.26	1.67	11,063.40	8,385.99	4,038.00

631

632

Table 2. Data on existing inter-basin transfer infrastructure

Name	Source basin	Receiving basin	Length (km)	Maximum flow rate (l/s)	Transfer capacity (hm^3/y)	Average annual transfer volume (hm^3)	Remarks
Ebro – Besaya	EBR	COC	60.00	4,000	27	0.95	Bidirectional
Cerneja – Ordunte	EBR	COR	5.00	160	13	0.00	-
Zadorra – Arratia	EBR	COR	45.39	9,000	190	175.56	-
Ebro–Tarragona	EBR	CIC	97.94	NA	126	66.54	-
Ciurana – Ruidecañas	EBR	CIC	11.00	NA	7	0.00	Disused
Tajo – Segura	TAJ	SEG	248.72	NA	600	321.54	-
Tajo – Segura – Júcar	TAJ	JUC	381,28	NA	NA	11.16	-
Tajo – Segura – Guadiana	TAJ	GDN	408.63	NA	50	18.75	-
Tajo – Segura – CMA	TAJ	CMA	297.58	NA	27	8.67	-
Guadiana – TOP	GDN	TOP	67.00	NA	75	138.53	-
Eiras – Porriño	GLC	MÑS	37.38	NA	NA	1.58	-
TOP – Guadalquivir	TOP	GDQ	22.48	NA	5	0.00	-
Guadiaro – Majaceite	CMA	GBB	24.01	30,000	110	60.76	-
Negratín – Almanzora	GDQ	CMA	96.00	NA	50	17.28	-

Table 3. Some network characterization measures

Nodes	15
Arcs	14
Transitivity	0.00
Reciprocity	0.13
Density	0.067
Centralization index (in-degree)	0.071
Centralization index (out-degree)	0.286
Average path length (unweighted)	1.67
Diameter (unweighted)	4
Average path length (km)	167.33
Diameter (km)	498.11

Table 4. Geodesic distances (km) between nodes (network diameter shown in bold)

To / From	CIC	MÑS	GLC	COR	COC	DUE	TAJ	GDN	GDQ	CMA	GBB	TOP	SEG	JUC	EBR
CIC															
MÑS															
GLC		37.38													
COR															
COC	71.00			65.00											60.00
DUE															
TAJ								408.63	**498.11**	297.58	321.59	475.63	248.72	381.28	
GDN									89.48	185.48	209.49	67.00			
GDQ										96.00	120.01				
CMA											24.01				
GBB															
TOP									22.48	118.48	142.49				
SEG															
JUC															
EBR	11.00			5.00	60.00										

Table 5. Assortative coefficient for some relevant node attributes

Period	Catchment inflows	IBT inflows	Volume stored in reservoirs	Reservoir capacity	Total demand
2008	0.185	-0.136	-0.586	-	-
2009	-0.396	-0.179	-0.352	-	-
2010	-0.428	-0.209	-0.209	-	-
2011	0.419	-0.201	-0.245	-	-
2012	0.072	0.006	-0.471	-	-
2013	0.123	-0.038	-0.443	-	-
2014	0.118	-0.040	-0.502	-	-
2008-2014	0.013	-0.114	-0.401	-0.34	-0.457