River networks in the framework of complex network theory

A. Sanchirico [1] and M. Fiorentino [2]

[1]*Dipartimento per l'Innovazione Umanistica, Scientifica e Sociale, University of Basilicata, Via Lanera 20, 75100 Matera, Italy*

[2]*Dipartimento di Ingegneria, University of Basilicata, Via dell'Ateneo Lucano 10, 85100 Potenza, Italy*

(Received 26 November 2024; revised 14 May 2025; accepted 5 June 2025; published 9 July 2025)

This study aims to conceptualize natural drainage networks as genuine complex networks. To overcome the limitation imposed by the regular connectivity structure of the traditional planted binary tree model—commonly employed to represent the planar configuration of fluvial systems—we introduce a line graph transformation. This mapping converts a planted binary tree into a plane tree, which can be regarded as a special class of loopless, directed graph characterized by inhomogeneous topology. From a geomorphological standpoint, our transformation naturally induces a hierarchical basin ordering that reflects the pioneering Gravelius stream order, offering insights into this scheme by projecting the hierarchy of tributary rivers onto a purely topological space. The application of complex network theory provides an explanatory perspective on river networks, yielding experimental findings that complement and extend established results in the geomorphological literature. In particular, our analysis of six catchments in southern Italy reveals previously unrecognized regularities, including a family of exact scaling laws and a surprisingly smooth width-type function that closely fits a Gamma distribution. These scaling laws emerge as power-law probability distributions for both the in-degree of nodes and the size of their in-components, which capture the scale invariance of tributary river lengths and drainage areas in tributary basins, respectively, thereby expanding the pantheon of known fluvial laws. Notably, these distributions represent sample distributions drawn from the complete populations of upstream lengths and drainage areas, with the sampling induced by our mapping through the selection of fluvial quantities corresponding to topologically well-defined elements of the plane tree. On the theoretical front, our approach is grounded in a model of directed, uncorrelated random graphs with arbitrary degree distributions. Among other things, by applying the generating function formalism to uncorrelated plane trees we derive an analytical expression for the distribution of drainage areas in tributary basins as a function of the distribution of tributary river lengths. By relying on average cluster properties, we predict the scaling exponent of the former distribution to be exactly 2 in the limit of large system sizes—a result that aligns well with both our empirical observations and previous findings in the complex network literature. Beyond the first-order approximation inherent in the model, our mapping also offers a coherent framework for interpreting the correlation structure consistently observed across all examined drainage networks.

DOI: 10.1103/kpbq-8wzb

I. INTRODUCTION

Networks with inhomogeneous topologies are ubiquitous in nature, representing a distinctive hallmark of complex systems and serving as the basic skeleton for their dynamic processes [1]. This prevalence largely arises from the exceptional flexibility of the mathematical abstraction inherent in modeling a system as a network, which is fundamentally represented by graph—a collection of vertices (or nodes) connected by edges (or links) [2]. This basic scheme transcends traditional disciplinary boundaries, enabling the representation of virtually any system of interacting elements while retaining only its essential traits. Accordingly, random graph models offer a unified framework for studying networks from a probabilistic perspective [3], independent of the physical systems they describe, facilitating the identification of common properties. Of particular relevance to this study are the scale-free degree distribution of nodes [4] and the correlations between the degrees of neighboring vertices [5–11], both of which are defining features that characterize a network as complex. The degree of a node is the number of edges incident to it, and the degree distribution represents the probability of a randomly chosen vertex having a specific number of edges.

This study pertains to directed graphs with complex topologies, where vertices are connected by directed edges [3,5,12,13]. Each vertex in a directed graph is characterized by both an in-degree, i, and an out-degree, u, representing the number of edges entering and leaving the vertex, respectively. The in-degree distribution, p_i, and the out-degree distribution, $\hat{p}_u$, describe the probability that a randomly chosen vertex has i incoming and u outgoing edges, respectively. Extensive analysis of numerous real-world network databases has revealed that these distributions typically exhibit a power-law tail, with

in-degree and out-degree exponents generally ranging from 2 to 3, although exceptions do exist [1].

While the node degree is an important local network statistic, it does not capture the global topological structure, which is more accurately described by connected components—clusters of vertices that are mutually reachable by at least one path along the edges. In directed graphs [3], this concept is extended into two categories: each vertex belongs to both an in-component and an out-component, which are the set of vertices from which a given node can be reached, and the set that is reachable from that node, respectively. The distribution, ψ_s, of in-component sizes represents the probability that a randomly chosen vertex belongs to an in-component of finite size s. Similarly, the probability distribution, $\hat{\psi}_t$, of out-component sizes gives the fraction of clusters, with t total vertices, which can be reached by a randomly chosen vertex. A power-law in-component size distribution, with scaling exponent approximately equal to 2, has been observed in measurements of the Internet topology [14]. The same inverse-square distribution has also been documented across various social [15,16] and ecological systems [17], where the original network is transformed into a hierarchical tree that reflects community structure. An asymptotic power-law tail of the form $\psi_s \sim s^{-2}$ has been derived using a model of growing random networks [18], and predicted through branching process theory [14,19] and tree models [20], suggesting the universality of this scaling behavior.

Another statistic that introduces a higher-order approximation with respect to the degree distribution is the joint degree-degree distribution, g_{hk}, which represents the probability that a randomly chosen edge leaves a vertex with in-degree k and points to a vertex with h incoming edges [21–23]. This distribution captures the network hierarchy, specifying how vertices from different degree classes pair up with each other. A network is said to be correlated when the joint distribution does not factorize into the product of its marginal distributions. An average correlation measure is given by the Pearson coefficient between the in-degrees of adjacent vertices, which has been found to be nonzero in many natural and artificial networks [5].

The degree distribution [3,24], degree-degree correlations [10,11,25–27], and edge direction [3,12,13], are key to the dynamics unfolding on top of networks, including critical phenomena [28] and spreading processes [29].

An important subclass of networks is represented by trees [30], which are connected graphs without cycles in which there exists exactly one path between any two nodes. Consequently, a tree has one more node than it has edges. A rooted tree is a tree with a designed node, known as the root, which imparts a natural orientation to the structure, with all edges considered as directed towards the root. This orientation reflects the asymmetric interactions among the system's elements, such as those arising from drainage directions in geomorphology. In combinatorics of trees, the in-degree of a node indicates the number of children it has, i.e., the number of nearest neighbors when edges are traversed in the backward direction. A node with an in-degree of 0 is termed a leaf; otherwise, it is referred to as an internal node. All but the root nodes have an out-degree of 1, while the root has zero out-going edges. The level of a node is the distance, measured by the number of edges along the path from the node to the root, with the root being the only node at level 0. A plane tree is a rooted tree in which the children of any node are ordered from left to right. Figure 1(a) illustrates a plane tree with five nodes. A special case is the binary tree, in which every internal node has exactly one left and one right child. A further variation is the planted binary tree—hereinafter referred to simply as a planted tree—a plane tree in which the root is the only node with a single child; in all other respects, the structure conforms entirely to that of a binary tree [31]. Figure 1(b) depicts a planted tree with five leaves.

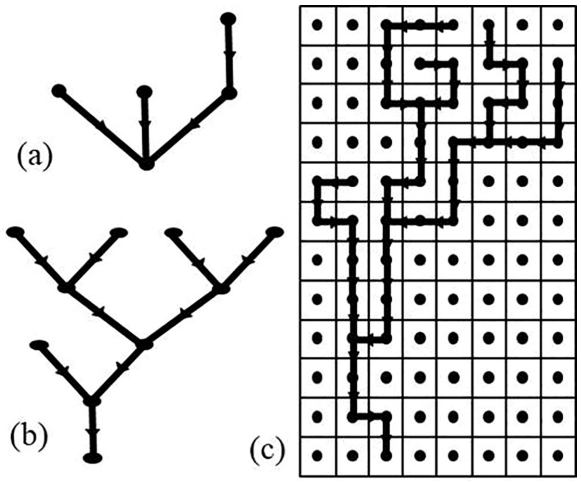

FIG. 1. (a) A plane tree with five nodes. (b) A planted tree with five leaves. (c) A spanning tree on a square lattice. Disregarding Euclidean edge lengths, the bifurcation structure of the spanning tree in panel (c) is mirrored by the planted tree in panel (b). The planted tree in (b) is mapped onto the plane tree in (a) via the line graph transformation formally defined in Sec. II A.

Our analyses focus on river networks, which are commonly represented as spanning trees on a regular two-dimensional lattice [see Fig. 1(c)], with unit-area sites corresponding to pixels in the digital elevation map that discretizes the surface. Bonds between sites indicate drainage directions, determined by local elevation gradients [32].

Two fundamental geomorphological quantities can be associated with each site along the spanning tree: the drainage area and the upstream length. The drainage area a_x at a point x is defined as the total contributing area draining into x through the flow directions, expressed as the number of upstream sites connected to x [32–35]. The drainage area is a crucial hydrologic quantity, serving as a measure of the flow discharge at site x in the case of uniform rainfall injection over the basin [32,33,35]. For each site x, the upstream length l_x refers to the Euclidean length of the main stream, defined as the upstream drainage path from x to the basin boundary, following the sites with the largest drainage areas at each bifurcation [34,36]. Alternatively, one can consider the longest stream in the drainage region, measured from point x to the divide [32,35,37].

By providing the most immediate view of the planar configuration, the spanning-tree representation of drainage networks quickly became dominant in fluvial analysis, leading to the identification of several scaling laws that govern the hierarchical structure of river systems [38]. Within the scope of

this study, the most relevant are the power-law probability distributions of upstream lengths and drainage areas, formulated as $\Pi_l \sim l^{-\gamma}$ and $\Phi_a \sim a^{-\tau}$, respectively, along with Hack's law [37], expressed as $l \sim a^{\eta}$, where length l and area a are measured at any channeled site within the basin [39]. Extensive comparative analyses of real catchments [38] have shown that the respective scaling exponents typically fall within the ranges $1.7 \leqslant \gamma \leqslant 1.9$ [36], $1.41 \leqslant \tau \leqslant 1.45$ [33–36], and $0.57 \leqslant \eta \leqslant 0.6$ [34,36,40,41], with Hack's exponent reflecting the tendency of natural watersheds to elongate as they grow larger [36]. Furthermore, through standard transformations of random variables, Hack's exponent can be derived as $\eta = (\tau-1)/(\gamma-1)$, thereby indicating a coherent scaling behavior [34,36].

Since the early days of geomorphology, these scaling laws have spurred extensive scientific activity aimed at unraveling the fundamental physical mechanisms underlying scale invariance in river networks. This line of inquiry has drawn upon a broad spectrum of approaches, including dynamical models of landscape evolution [32,35,42,43] and optimal channel networks [32,35,43–45].

By focusing exclusively on bifurcation patterns, the spanning tree is reduced to a planted tree, in which the root node serves as the river's outlet, the leaves correspond to sources, and the internal nodes correspond to fluvial confluences; the links between node pairs ultimately represent the individual stream channels. This discrete structure enables the classification of river networks by assigning a stream order to channel segments throughout the basin [46–48] and lends itself naturally to the statistical analysis of fluvial quantities [40,46,48–52].

However, while the spanning-tree model and its associated planted tree effectively capture the scale invariance observed in river networks, they suffer from a crucial limitation in revealing their scale-free degree distribution, due to the inherent absence of a heterogeneous connectivity structure. Planted trees, in fact, qualify as regular graphs, where the in-degree of all nodes—except for the root—is restricted to either 2 or 0, a constraint that also inhibits the emergence of nontrivial correlations between the in-degrees of adjacent vertices. Consequently, the planted tree fails to represent natural drainage networks as irregular graphs, characterized by heavy-tailed degree distributions and degree–degree correlations—two features widely recognized as unmistakable statistical signatures of complex networks.

As a result, despite the identification of a wide array of scaling laws in fluvial systems nearly half a century before the concept of complex networks gained prominence—and in spite of significant progress in unifying these laws into a coherent framework [40]—a formal treatment of river networks as genuine complex networks is still lacking. It is paradoxical that, more than two decades after the inception of complex network science, its relevance to fluvial geomorphology remains largely unexplored—especially considering that river networks have long been recognized as prototypical examples of loopless tree graphs.

To bridge this gap, we introduce a line graph transformation [53] that maps a planted tree with n leaves onto a plane tree with n vertices, whose internal nodes can inherently accommodate an arbitrary number of incoming edges. In this transformation, tributary rivers are mapped to vertices, while the internal nodes along these rivers are converted into directed edges in the line graph. Since tributaries in the planted tree may host a variable number of confluences, the corresponding vertices in the line graph exhibit a heterogeneous in-degree distribution, thereby enabling the application of complex network theory to the analysis of fluvial systems.

Within this framework, our approach relies on a model of random graphs with arbitrary degree distributions [3], known as the configuration model [54], which extends the classical Erdős-Rényi model [55] to networks with heterogeneous topology. Except for the constraint imposed by the prescribed degree distribution, the model assumes the graph is entirely random, with edges between vertices assigned independently of their degrees. Despite being a first-order approximation, this null model has been shown to describe many interesting properties of real-world complex networks [1,3].

We present our main results while outlining the structure of the paper as follows. In Sec. II A we describe our mapping and illustrate how it inherently captures the hierarchical organization of tributary rivers through topological distances in the plane tree. Notably, this mapping leads to the pioneering Gravelius stream order [56]—originally devised for cartographic applications—which here emerges naturally from a graph-theoretical framework. Section II B presents an experimental analysis of six catchments located in southern Italy. Our findings provide compelling evidence that the line graphs of river networks exhibit power-law distributions in both the in-degree of nodes and the size of in-components, corresponding respectively to the probability distributions of tributary river lengths and the drainage areas of tributary basins in the underlying planted tree description. These, in turn, represent sample distributions drawn from the full populations of upstream lengths and drainage areas, with data points restricted to the planted subtrees corresponding to the plane tree vertices under our mapping. Specifically, the sample comprises the n planted subtrees drained by their respective tributary rivers, selected from a total of $2n-1$ planted subtrees identifiable within a given catchment. This sampling results in scaling exponents that deviate from the canonical values commonly reported in geomorphological literature. Remarkably, the sample distribution of drainage areas displays no observable cutoff. These results reveal a family of scaling laws emerging from the underlying complex network topology, wherein fluvial random variables are intrinsically associated with well-defined topological elements in the line graph. Importantly, the scaling exponent for the in-component size distribution remains consistently close to 2 across all the catchments examined, aligning closely with findings from previous studies [14–20]. We then turn our attention to the distribution of levels in the line graph—namely, the out-component size distribution—which closely fits a Gamma distribution across all analyzed basins. This uncovers a previously unreported regularity in river network topology, which is also reflected in the observed power-law decay of the branching ratio, describing the growth rate of vertices along the plane tree levels. On the theoretical side, in Sec. III we introduce a mathematical formulation of river networks based on a model of directed, uncorrelated random graphs with arbitrary degree distributions [3]. This model

leads to a closed-form expression that relates the distribution of drainage areas in tributary basins to the generating function of the tributary river length distribution, thereby laying the groundwork for a combinatorial interpretation of Hack's law. Furthermore, by exploiting the average properties of clusters, this expression predicts a scaling exponent for the drainage area of tributary basins exactly equal to 2, in precise agreement with our experimental observations (Sec. III D). Moving beyond the first-order approximation, Section IV presents empirical evidence that natural drainage networks exhibit positive correlations between the in-degrees of adjacent vertices, with Pearson coefficients consistent with values reported in the complex network literature. In Sec. V we revisit our approach in light of branching process theory [57], which offers an alternative interpretation for some of our results. Finally, Sec. VI is devoted to our conclusions.

II. COMPLEX RIVER NETWORKS

The regular topological structure of planted trees has, until now, kept fluvial geomorphology outside the domain of complex network science, in contrast to its widespread application in other fields centered around networked systems. To characterize natural drainage networks as true complex networks with scale-free topologies, we introduce a line graph transformation that replaces the planted tree with a plane tree, where internal nodes may have an arbitrary number of children. This mapping allows for the application of complex network formalism to the empirical analysis of fluvial systems, offering an explanatory perspective that uncovers previously unrecognized topological patterns in river networks.

A. The river network as a plane tree

A planted tree with n leaves contains exactly $2n-1$ links—a relationship known in geomorphology as Melton's law [47], where n indicates the number of sources in the river network, also referred to as the magnitude of its terminal link. Each internal node v, except for the root of the entire planted tree, acts as the common root of the left and right planted subtrees of v, which include its left and right children, respectively. In total, $2n-1$ planted subtrees can be extracted from the entire tree, each having as its own terminal link a unique link from the planted tree.

The drainage area and upstream length at any given point along the spanning tree can be reformulated in terms of nodes and links in the corresponding planted tree. Let z be a link from node y to node x in a planted tree. The drainage area a at link z is defined as the total contributing area that drains through z, and it is assumed to be equivalent to the total number of links in the planted subtree rooted at x that contains y. Using Melton's law, the area a is given by $a = 2s-1$, where s represents the magnitude of link z, i.e., the number of leaves ultimately tributary to it. The main stream length in a planted subtree rooted at x is defined as the number of links in the backward path from x to a leaf, selecting the upstream link with the largest total contributing area at each internal node. In this study, however, we refer to the length l of the longest stream, defined as the number of links in the backward path from x to the leftmost leaf at the maximum level. Finally, the probability distribution of topological distances—known as the width function—represents the normalized number of links leaving all nodes at a given distance d from the root [45,49,51,58]. This corresponds to the level distribution in the planted tree, where, in geomorphological contexts, levels are conventionally assigned to the links emanating from the nodes, rather than to the nodes themselves. Acting as a proxy for the local contributing areas at a distance from the basin's outlet, the width function is a key hydrological statistic that reflects the geomorphological response of a catchment under simplified assumptions [59].

Compared to the spanning-tree model, the planted tree scheme introduces no significant physical differences, apart from the exclusion of Euclidean lengths. Specifically, it provides an alternative computational method for estimating the probability distributions of drainage areas and upstream lengths, which are now defined for any planted subtree within the channel network. As we will explain shortly, the key innovation of our approach lies in considering a strict subset of planted subtrees. This subset is shown to be naturally induced by our mapping, reflecting the control exerted by the inherent complexity of river networks on the statistics of the random variables that characterize a fluvial system.

On the other hand, as discussed in the Introduction, the main factor that excludes current descriptions of river networks as relevant models for complex network science is the inherently regular topological structure of the planted tree. This structure results in a bimodal in-degree distribution of nodes, which, in the large n limit, can be expressed as $p_i = [\delta(i) + \delta(i-2)]/2$, where $\delta(\cdot)$ denotes the Kronecker delta function. This characteristic sets planted trees apart from the broader class of complex networks typically found in nature, which exhibit a much richer and more diverse topology of connections.

To overcome this limitation, drawing on line graph transformations [53], we introduce a nonbijective map from planted trees with n leaves to plane trees with n vertices, with no restrictions on the in-degree of nodes. The plane tree resulting from this mapping will also be referred to as the line graph of the planted tree. Our map is defined recursively as follows: we begin by considering the path from the leftmost leaf at the maximum level to the root of a planted tree, which represents the longest river in the entire tree. Each internal node x, excluding the root, along this path is treated as the outlet of a tributary river, which is identified by the longest stream within the associated planted subtree, i.e., the planted subtree rooted at x that contains the tributary river. We then proceed recursively, considering progressively smaller tributary rivers feeding into the internal nodes (excluding the outlet) along the traced longest rivers, continuing until the tributary rivers are reduced to a single link in the final step. The entire planted tree is then transformed into a plane tree by mapping tributary rivers onto vertices, with a directed edge from vertex w to vertex v in the line graph if and only if the outlet of the tributary river corresponding to w coincides with an internal node of the river corresponding to v. Each tributary river originates from a distinct leaf of the planted tree, so the plane tree has the same total number of vertices as there are leaves in the planted tree. Furthermore, the leaves of the plane tree form a subset of the planted tree's leaves,

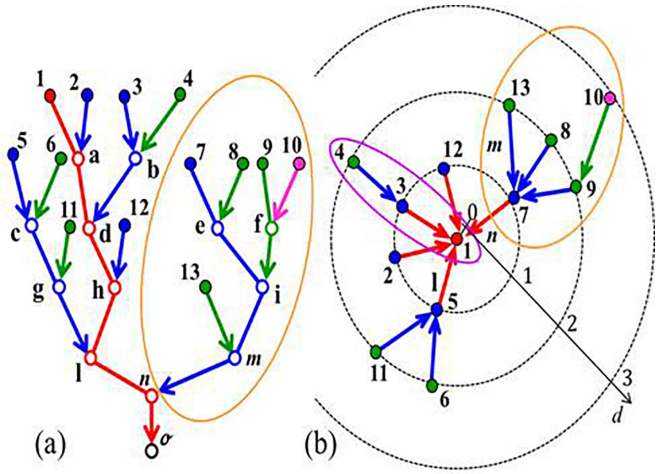

FIG. 2. (a) A planted tree with 13 leaves and the root at node o. Leaves and internal nodes are labeled with Arabic numerals and lowercase letters, respectively. The longest river in the entire basin, originating at leaf 1 and terminating at the outlet o, is marked in red and assigned hierarchical order 0. All internal nodes along this drainage path—excluding the root o—serve as the outlets for tributary rivers that flow directly into the main river; these tributaries are assigned hierarchical order 1 and are depicted in blue. Tributary rivers of orders 2 and 3 are colored green and magenta, respectively. (b) The line graph of the planted tree shown in panel (a). In this transformation, tributary rivers are mapped to vertices, while internal nodes are represented as directed edges. For instance, the root of the plane tree, vertex 1, corresponds to the main river in the planted tree. Similarly, the tributary river originating at leaf 5 and terminating at node l corresponds to vertex 5 in the line graph, while node l maps to the directed edge from 5 to 1. By comparing both panels, it is evident that all tributary rivers of hierarchical order d correspond to vertices in the plane tree located at a distance d from the root node 1. The orange oval in panel (b) encloses the five vertices comprising the in-component of vertex 7, i.e., the plane subtree rooted at vertex 7. This subtree corresponds to the right planted subtree rooted at node n in panel (a), also enclosed in orange, which represents the tributary tree drained by the river originating at leaf 7 and ending at node n. Its leaves define the magnitude of the terminal link from node m to node n. Note that the left planted subtree of node n, which is drained by the fluvial segment from leaf 1 to node n—thus overlapping with the main river in the catchment—does not correspond to any plane subtree rooted at a vertex in the line graph. The indigo oval in panel (b) encompasses the three vertices forming the out-component of node 4. This vertex lies at a level one less than the size of the cluster to which it belongs (node 4 belongs to an out-component of size $t = 3$ and lies at level $d = 2$). The level of node 4 also determines the hierarchical order of the corresponding tributary river that originates at leaf 4 in the planted tree shown in panel (a).

specifically those that give rise to tributary rivers consisting of a single link. Figure 2 illustrates a planted tree with thirteen leaves alongside the corresponding plane tree with thirteen vertices, where tributary rivers identified at the same recursion step are colored identically. In this construction, the tributary rivers within the planted subtrees they drain are assumed to be the dominant streams in the channel network, reflecting the strength of the connections in the plane tree. The in-degree of a vertex represents the length of the corresponding tributary river: the longer a tributary river, the more numerous the fluvial confluences it encounters, and consequently, the higher the in-degree of the vertex in the line graph. The longest river of the entire planted tree is then mapped onto the root of the plane tree, making it the node with the highest number of incoming edges.

It is important to emphasize that our mapping selects exactly one of the two planted subtrees rooted at any given node x within the entire tree. Specifically, it chooses the subtree whose longest stream corresponds to the tributary river flowing into x, since only tributary rivers are mapped to vertices in the line graph—the discrete structure central to this study. We refer to the planted subtree identified in this way as a *tributary tree*, and to the subbasin it drains as a *tributary basin*. For example, in Fig. 2(a), only the right planted subtree rooted at node n is considered, as it is drained by the tributary river originating at leaf 7 and terminating at n, whereas the left subtree—whose longest stream overlaps with the river identified in the previous recursion step—is excluded. As a result, our map selects n tributary trees from the total of $2n-1$ planted subtrees.

This is an important point to grasp, as we are interested in the distributions of the in-degree of vertices and the size of their in-components within the line graph of a river network with n sources. In the next section, these distributions will be interpreted as the probability distributions of tributary river lengths and drainage areas of tributary basins, respectively. As a result, while the random variables in the line graph pertain to the n vertices of the plane tree, their fluvial counterparts correspond to n-element samples drawn from the entire populations of upstream lengths and drainage areas across the basin. Consequently, our sample distributions are expected to differ from the full probability distributions of upstream lengths and drainage areas measured in traditional geomorphological analyses [38], which typically reference all $2n-1$ planted subtrees in the discretized channel network.

Note that the map we introduce results in a hierarchy of tributary rivers that essentially mirrors the Gravelius stream order [56], reflecting the traditional naming conventions of rivers. In this scheme, the longest stream in the planted tree is assigned order 0, while a tributary river receives order d if it flows into another of order $d-1$. Our approach offers additional insights into this ordering, revealing that the hierarchical relationships among tributary rivers are inherently captured through topological distances in the line graph. Specifically, counting the number of vertices at each level in the plane tree leads to a width-type function, whereby vertices are naturally ordered according to their topological distance from the root. The topological ordering induced by the mapping therefore corresponds to the hierarchical stream order of tributary rivers in the planted tree. As shown in Fig. 2, the number of vertices at level d in the plane tree matches the number of tributary rivers of order d in the original planted tree. Moreover, due to the properties of line graph transformations [53]—which convert incidence relationships into adjacency relationships—the topological distance from a vertex v to the root of the plane tree reflects the edge-to-edge distance in the planted tree. This distance is measured by the number of tributary rivers—plus one—that separate the river corresponding to v from the longest stream in the entire basin.

The edge-to-edge distance, in turn, reflects the nested structure of tributary basins: a tributary river flows into another if and only if its tributary basin is contained within the other.

This nested architecture can be formally expressed as an inclusion relation as follows: let the set of tributary basins be endowed with the partial order defined by the inclusion relation between pairs of subsets. It is well known that any partially ordered set can be represented as a directed acyclic graph, where the vertices correspond to the elements of the collection, and a directed edge from vertex w to vertex v exists if and only if the elements are ordered such that $w \preccurlyeq v$ (i.e., w is included in v). As a result, the line graph of the river network can be interpreted as the Hasse diagram of the partial order, which represents the inclusion relations between tributary basins through the reachability relation between the corresponding vertices.

These considerations highlight a significant distinction between the Gravelius stream order and the Horton-Strahler ordering scheme [46], the latter providing a measure of complexity in the bifurcation patterns of natural drainage networks and also being used to describe the hierarchical organization of binary trees as a reflection of the community structure in underlying complex networks [15,60]. Despite its widespread adoption, however, the Horton-Strahler taxonomy fails to establish a consistent hierarchy among subbasins in terms of topological distances. For example, in this scheme, any link originating from a source is classified as a first-order segment, irrespective of the order of the fluvial segment it merges with. As a result, two first-order links are incorrectly assigned the same edge-to-edge distance from the highest-order stream segment that ultimately collects both. In contrast, the Gravelius stream order can be explicitly conceived as an inclusion relation between subsets, inducing a topological distance between vertices in the line graph while capturing the nested architecture of the corresponding tributary rivers. Accordingly, our mapping allows the Gravelius stream order to be regarded as a topological ordering in the sense proposed by Shreve [48], while also offering a generalization of the hierarchical view of Horton-Strahler [46].

Before concluding this section, it is worth noting that a similar approach has been employed in the analyses of urban street networks, where axial lines representing streets are mapped as nodes, and intersections between pair of roads are treated as edges in the dual graph representation [61].

B. Fluvial distributions and their complex counterparts in network theory

The map we introduce enables the examination of three fundamental fluvial quantities through the framework of complex network theory: the probability distributions of tributary river lengths and drainage areas of tributary basins, as well as the distribution of the number of tributary rivers at a given edge-to-edge distance from the longest river in the basin—i.e., the number of tributary rivers of the same Gravelius stream order. In accordance with the constraints imposed by our mapping, these discrete distributions are defined solely at the n points corresponding to the vertices of the plane tree. Consequently, while analogous to those typically considered in traditional geomorphological studies, our sample distributions exhibit key differences. As a result, the distributions and scaling properties we derive are expected to deviate from those known so far for river networks [38].

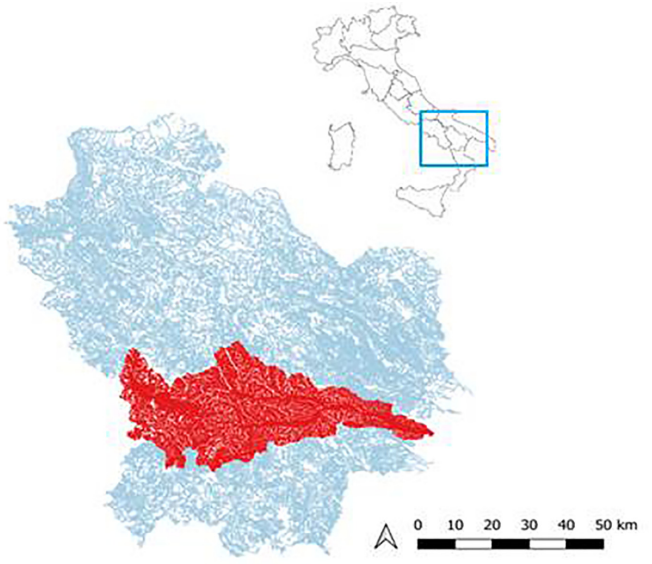

FIG. 3. Geographical location of the study area in southern Italy, with the Agri River in the Basilicata region marked in red.

In this section, while discussing theoretical aspects, we also present results from an experimental analysis conducted on six catchments in southern Italy, located approximately at the 41st parallel (see Fig. 3). The six basins, which are adjacent to each other, differ for climate and range in size from around 380 to 3000 square kilometers.

Relevant characteristics of these catchments are provided in Table I. The raw data consist of a scalar elevation field on a regular square grid with a cell size of 5×5 m^2 [62]. Drainage directions were extracted using the steepest descent method with GRASS QGIS software, assuming a flow accumulation threshold of 1000 cells to determine channel initiation [63].

As explained in the previous section, our mapping transforms tributary rivers into vertices, while the internal nodes (excluding the outlet) along the rivers are converted into directed edges leading to the vertices. Consequently, the length l of a tributary river is equal to the in-degree i of the corresponding vertex in the plane tree, plus one, with the additional factor of 1 accounting for the terminal link entering the river's outlet. Accordingly, the probability distribution Π_l^* of the lengths of tributary rivers is then mapped onto the in-degree distribution p_i of nodes in the line graph, such that $p_i = \Pi_l^*$, with $l = i + 1$ for any in-degree $i \geqslant 0$. As a result, the in-degree of a vertex can take any non-negative integer value, depending on the length of the corresponding tributary river, effectively circumventing the constraint imposed by the regular in-degree distribution of nodes in planted trees.

In Fig. 4, a log-log plot illustrates the exceedance probability distribution of tributary river lengths for the Agri River, representing the probability that a given length exceeds a certain value, with length measured by the number of links. This analysis reveals that the line graph of the Agri River is a complex network with scale-free in-degree distribution of nodes, characterized by a scaling exponent derived from the

TABLE I. Summary of the general characteristics of the six river networks analyzed. For each line graph of the river network, the following attributes are reported: catchment area (in square kilometers); total number of vertices, n; in-degree of the root node, i_r; total number of leaf nodes, n_0; height of the tree, Δ; observed scaling exponent α of the power-law in-degree distribution of nodes; observed scaling exponent β of the power-law distribution of in-component sizes; average topological level, E[d]; variance of the topological distances across the plane tree, var[d]; Hack's exponent η, estimated from the log-log plot of tributary river lengths versus drainage areas in tributary basins; Hack's exponent η, estimated via the relation $\eta = (\beta-1)/(\alpha-1)$; scale parameter γ of the Gamma distribution fitted to the empirical topological distance distribution using the method of moments; scale parameter γ estimated using the log-log plot as described in Eq. (45) (Sec. III C); scaling exponent τ of the power-law decay of the branching ratio, estimated by data regression [Eq. (46), Sec. III C]; scaling exponent β of the in-component size distribution, computed via Eq. (55) (Sec. III D); Pearson correlation coefficient R for the in-degrees of adjacent vertices (Sec. IV A); and the estimated average branching ratio, $\langle B_d \rangle$ (Sec. V).

River name	Bradano	Agri	Basento	Sinni	Cavone	Noce
Area [km^2]	3000	1723	1535	1360	684	380
n	31 010	21 788	18 980	16 523	7277	3705
i_r	898	821	976	727	538	355
n_0	24 228	17 142	14 771	12 850	5704	2841
Δ	9	8	9	8	6	7
α	2.50	2.56	2.54	2.52	2.46	2.50
β	1.99	1.99	2.01	1.99	1.98	1.97
E[d]	3.30	3.24	3.06	3.16	2.76	2.77
var[d]	1.19	1.30	1.23	1.10	0.92	1.07
η	0.69	0.68	0.68	0.68	0.69	0.67
$(\beta-1)/(\alpha-1)$	0.67	0.63	0.66	0,65	0.67	0.64
γ	9.11	8.06	7.58	9.04	8.25	7.20
γ [Eq. (45)]	8.67	8.52	7.53	8,32	8.39	8.02
τ [Eq. (46)]	3.97	4.19	3.74	4.14	4.72	4.21
β [Eq. (55)]	2.02	2.03	2.04	2.03	2.05	2.07
R	0.205	0.236	0.233	0.211	0.230	0.244
$\langle B_d \rangle$	100.95	103.79	109.4	91.89	90.74	51.48

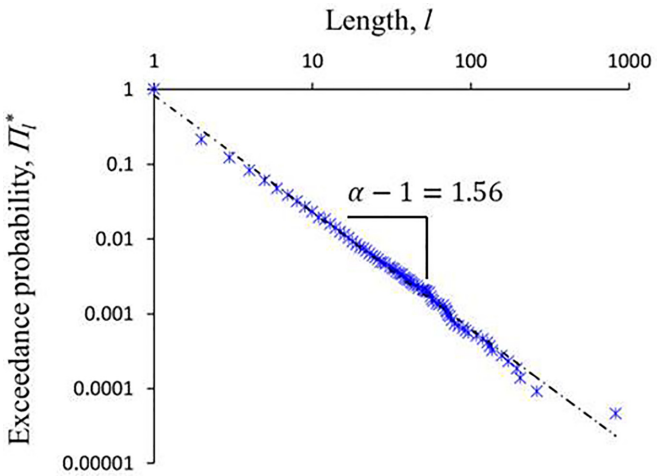

FIG. 4. Log-log plot of the complementary cumulative distribution of tributary river lengths for the Agri River, corresponding to the in-degree distribution of the nodes in its line graph representation.

power-law cumulative distribution of tributary river lengths. As shown in Table I, the scaling exponent values for the probability distribution of the six basins analyzed range from 2.46 to 2.56, which is in good agreement with the commonly observed interval (2 to 3) of power-law exponents in real-world networks [1]. This finding enables river networks to be treated as directed graphs with heterogeneous topologies, analogous to other complex networks studied across various fields, whose far-reaching physical implications were discussed in the Introduction.

On the other hand, as expected, the observed scaling exponents for the distribution of tributary river lengths deviate from those characterizing the distribution of all upstream lengths, which typically hover around 1.8 [36]. To provide a more comprehensive discussion of this difference, we will revisit it after examining the scaling exponents associated with the probability distribution of drainage areas in tributary basins, where analogous behavior is observed. At this stage, we simply note that our analysis is restricted to the n tributary rivers corresponding to the vertices of the plane tree—a necessary condition for interpreting fluvial lengths as the in-degrees of nodes. All other drainage paths identifiable within the river network do not correspond to any elements of the plane tree and, therefore, lose significance from a graph-theoretical perspective. In contrast, traditional geomorphological analyses typically consider all $2n-1$ longest streams, each associated with a different planted subtree of the entire planted tree representing the drainage network. As a result, our sampling leads to a faster decay in the exceedance probability distribution of tributary river lengths compared to that of all upstream lengths.

Another important fluvial quantity that can be reframed within the context of complex network theory is the distribution of drainage areas in tributary basins. To this end, we begin by revisiting the notion of an in-component in a tree with respect to a given node. As outlined in the Introduction, the in-component of a vertex v in a directed graph is defined as the set of all vertices from which v can be reached by following a path of directed edges [3]. When the graph is reduced to a tree, the in-component of v corresponds precisely to the subtree rooted at v. Accordingly, the distribution of the total number s of vertices from which a given vertex can be reached also stands for the distribution of subtree sizes—that is, the probability of encountering a subtree containing s vertices.

We now let x be the outlet of a tributary river, let $\mathcal{B}_x$ denote the tributary tree rooted at node x and drained by the tributary river, and consider the vertex v in the plane tree, which corresponds to the tributary river in question. Since the map we introduce converts incidence relationships into adjacency relationships, it transforms $\mathcal{B}_x$ into the plane subtree rooted at v, denoted by $\mathcal{P}_v$. The total number s of vertices in $\mathcal{P}_v$—which represents the size of the in-component of root v—hence corresponds to the number of leaves in $\mathcal{B}_x$. According to Melton's law [47], the number of links in $\mathcal{B}_x$—that is, the total contributing area at its terminal link—is given by $a = 2s-1$. Therefore, our mapping translates the drainage area of the tributary basin drained by $\mathcal{B}_x$ into the size of the in-component of vertex v. Consequently, the probability distribution ψ_s of in-component sizes in a plane tree corresponds to the probability distribution Φ_a^* of drainage areas in tributary basins,

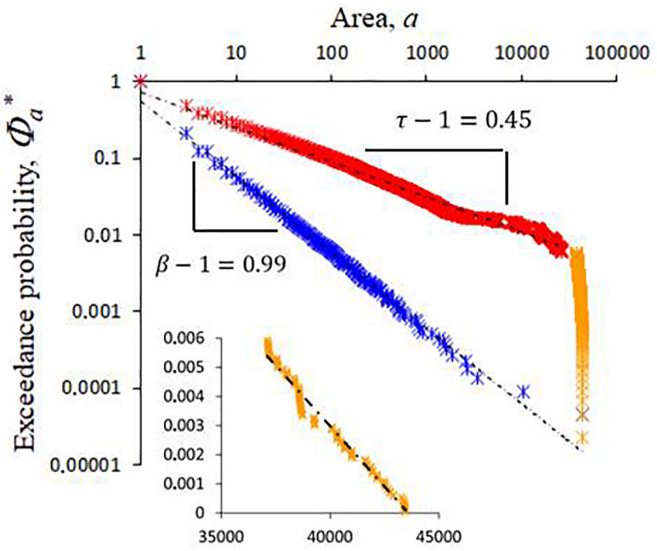

FIG. 5. Log-log plot of the cumulative probability distribution of drainage areas in the tributary basins of the Agri River (blue straight line), reflecting the distribution of in-component sizes in the corresponding line graph. The red-orange curve shows the exceedance probability distribution of all drainage areas, accounting for every planted subtree within the Agri River. An abrupt cutoff is observed at the transition from 26 649 to 37 987 links, highlighted by the change in color from red to orange. The area jump occurs at the confluence where the main river of the Agri basin meets its largest first-order tributary, as indicated in Fig. 6. Inset: Linear-scale view showing the linear decay in the exceedance distribution of all drainage areas within the cutoff region.

such that $\psi_s = \Phi_a^*$, with $a = 2s-1$ for any in-component size $s \geqslant 1$. In turn, the distribution Φ_a^* also represents the probability of encountering a tributary tree with a total of $a+1$ nodes in the planted tree, just as the distribution ψ_s stands for the probability of finding a plane subtree with s vertices.

The blue curve in Fig. 5 represents the exceedance probability distribution of drainage areas for tributary basins in the Agri River, plotted on a log-log scale. Here the total contributing area is treated as a discrete random variable, measured by the number of links in the n tributary trees that drain the respective subbasins. According to our mapping, the observed power-law regime in the cumulative distribution of drainage areas in tributary rivers implies a scale-free distribution for the in-component sizes of vertices in the plane tree. This analysis was similarly applied to all the real basins considered, and, as shown in Table I, the measured scaling exponent values for the probability distribution consistently approach 2. As will be discussed in Sec. III D, this result will also find important theoretical support. As expected, the cumulative distribution of the in-component sizes falls off more slowly than the exceedance probability of node in-degrees, due to the lower bound imposed by the vertex degree on the size of the in-component to which the vertex belongs [64].

Similar to the scaling of tributary river lengths, the observed value of 2 for the scaling exponent of drainage areas in tributary basins is consistent with results from the broader graph literature [14–20]. This exponent clearly differs from those reported in previous geomorphological studies [33–36], where values typically range around 1.43, as it refers specifically to tributary trees, rather than to all possible planted subtrees, which are commonly considered in those contexts.

For the sake of clarity, we benchmark our findings against established approaches, with specific reference to river network scheme proposed in this study. Our primary objective is to describe the topological structure of the line graph of a river network with n sources, that is, a plane tree of size n. Consequently, when analyzing the probability distributions of the in-degree of nodes and the size of in-components, the relevant random variables are defined at each vertex in the plane tree. It follows that, when tracing back to the underlying planted tree, the corresponding fluvial quantities are associated with a strict subset of n tributary trees, selected from the overall set of $2n-1$ planted subtrees. Only the tributary trees correspond to the n vertices in the line graph through their respective tributary rivers. Therefore, the probability distributions of tributary river lengths and drainage areas of tributary basins are derived from n-element samples drawn from the full populations of upstream lengths and drainage areas, which are typically evaluated at each planted subtree within the basin in traditional geomorphological analyses [38]. Note that our n-subsets of random variables are not random samples, but rather a rational selection induced by the mapping, which determines how subtrees are identified and how many are included in the analysis, based on the vertex-tributary river correspondence. Planted subtrees that are not tributary trees are associated with fluvial paths and drainage regions in the river network, which have no topological meaning in the line graph. They do not have a correspondence in the plane tree and are therefore excluded from our calculations. For comparison, Table II presents all the random variables of interest for the synthetic planted tree depicted in Fig. 2(a).

On the other hand, when estimating the full distributions of upstream lengths and drainage areas, we immediately recover the typical scaling exponent values reported in literature [38]. For instance, the red-orange curve in Fig. 5 shows the exceedance probability distribution of drainage areas for the Agri River, plotted on a log-log scale and based on all $2n-1$ planted subtrees in the river network. The empirical data clearly follow a power-law distribution, with the typical abrupt cutoff at larger scale [33–36]. In this case, the scaling exponent, estimated via the least-squares method over the linear regime, is approximately $\tau - 1 = 0.45$, in good agreement with established geomorphological results [33–36].

Figure 5 highlights an additional significant finding from our analysis: the absence of a cutoff in the cumulative distribution of drainage areas in tributary basins, in contrast to the exceedance distribution of all drainage areas, which exhibits a sharp cutoff at the maximum system size—a phenomenon commonly attributed to the finite-scale effect [34,35]. Our analysis reveals that the differences between the two distributions arise from the distinct collections of subbasins. Including all $2n-1$ planted subtrees introduces a redundancy effect, particularly pronounced near the catchment's closure, where the drainage network typically narrows into a main channel that collects small tributary branches. In this region, the total contributing area increases by nearly a constant amount at each fluvial junction [65], leading to a uniform discrete density function for the drainage areas of all subbasins.

TABLE II. The first column lists the outlets of all subbasins identifiable within the entire catchment: each internal node, except for the root ϕ of the full planted tree, serves as the common root of its left and right planted subtrees. The left and right children of these roots are listed in the second column. The third and fourth columns report, respectively, the longest stream lengths and the total contributing areas for each planted subtree. The fifth column indicates the leaves from which tributary rivers originate; their lengths and the drainage areas of their respective tributary basins are again reported in the third and fourth columns. Vertices in the plane tree—corresponding to tributary rivers—are labeled according to their associated leaves and therefore reappear in the fifth column. The sixth and seventh columns indicate, respectively, the in-degree and the in-component size of each vertex to which a tributary river is mapped. River lengths and drainage areas in tributary basins can be recovered using the relationships $l = i + 1$ and $a = 2s-1$, as described in the text. Not all planted subtrees listed in the first column identify a tributary tree—that is, a vertex of the plane tree—in the fifth column.

ID outlet	ID child	l	a	ID leaf	i	s
a sx	1	1	1	–	–	–
a dx	2	1	1	2	0	1
b sx	3	1	1	–	–	–
b dx	4	1	1	4	0	1
c sx	5	1	1	–	–	–
c dx	6	1	1	6	0	1
e sx	7	1	1	–	–	–
e dx	8	1	1	8	0	1
f sx	9	1	1	–	–	–
f dx	10	1	1	10	0	1
g dx	11	1	1	11	0	1
h dx	12	1	1	12	0	1
m sx	13	1	1	13	0	1
d sx	a	2	3	–	–	–
d dx	b	2	3	3	1	2
g sx	c	2	3	–	–	–
h sx	d	3	7	–	–	–
i sx	e	2	3	–	–	–
i dx	f	2	3	9	1	2
l sx	g	3	5	5	2	3
l dx	h	4	9	–	–	–
m dx	i	3	7	–	–	–
n sx	l	5	15	–	–	–
n dx	m	4	9	7	3	5
ϕ	n	6	25	1	5	13

This results in the corresponding linear exceedance distribution collapsing into an abrupt cutoff on a bilogarithmic scale. Supporting this observation, the inset in Fig. 5 illustrates the linear decay of the cumulative distribution of all drainage areas within the cutoff region for the Agri River, represented on a linear scale. Notably, the cutoff begins when the total contributing area sharply increases from 26 649 to 37 987 links, as indicated by the shift from the red to the orange segments of the curve in Fig. 5. This transition occurs at the confluence where the longest river of the entire basin meets its largest first-order tributary. Beyond this point, the Agri River resembles a backbone, with small tributary branches feeding into the main stream (see Fig. 6).

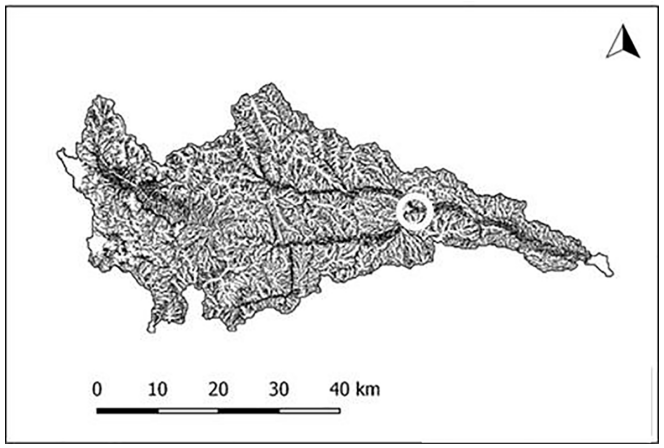

FIG. 6. The Agri River catchment: the white circle marks the confluence where the main river meets its longest tributary.

While the abrupt cutoff appears to be a numerical by-product of the excessive number of areal elements near their maximum size, our mapping avoids the redundancy of subbasins that affects previous geomorphological analyses by selecting only the planted subtrees corresponding to tributary trees. As a result, no cutoff is observed when analyzing the line graph of a river network. The plane tree scheme seems to capture the river network topology in its most essential form, leading to a smoother power-law distribution for drainage areas when mapped onto the sizes of the in-components of vertices in the line graph. All other drainage areas across the river network, which do not correspond to plane subtrees, seem to introduce a disturbance effect manifesting as a cutoff. This suggests that a line graph representation of river networks may offer a natural framework for investigating them.

Let us now turn to the analysis of out-components in directed graphs [3], and let t denote the size of the set of vertices reachable from a given vertex v. When the graph is reduced to a tree, where every vertex except the root has exactly one out-going edge, the total number t of vertices reachable from v corresponds to the level $d = t-1$ of that vertex [see Fig. 2(b)]. Therefore, up to a unit shift, the probability distribution $\hat{\psi}_t$ of the out-component sizes, as defined in the Introduction, can be interpreted as the probability of finding a vertex at distance d from the root. That is, $\hat{\psi}_t = \Theta_d$, for any out-component size $t = d + 1 \geqslant 1$, where Θ_d denotes the level distribution.

The distribution Θ_d of vertices that are d steps away from the root of a plane tree corresponds, in turn, to the distribution of d-order tributary rivers in the underlying planted tree. This distribution can then be interpreted as a form of width function, where the mass elements at each level of the plane tree represent tributary rivers of given order and varying lengths throughout the catchment. In this scheme, the edge-to-edge distance reflects the number of tributary rivers that the flow originating from a headwater must traverse to reach the longest river in the basin, which acts as the terminal outlet. Figure 7 shows the plot of Θ_d as a function of d for the Agri River. The data exhibit good agreement with a Gamma distribution, whose parameters were estimated using the method of moments. Notably, the distribution displays a peak around its mean. Similar patterns are observed in the other analyzed

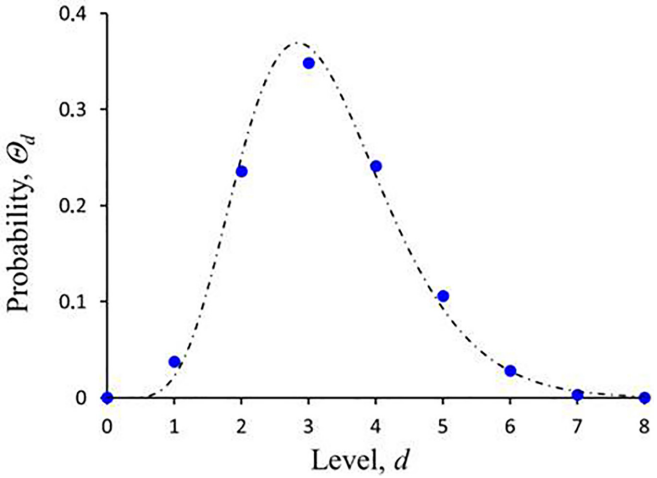

FIG. 7. Empirical distribution of topological distances in the plane tree, or, equivalently, the distribution of tributary rivers sharing the same Gravelius stream order in the original planted tree. The blue points represent the observed number of vertices at distance d from the root node in the line graph of the Agri River, while the dashed curve shows the Gamma fit to the data.

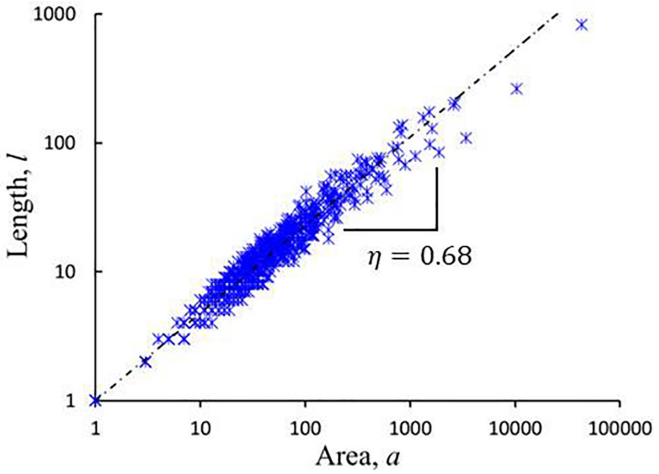

FIG. 8. Hack's law for the set of tributary basins. Hack's exponent η is estimated by fitting tributary river lengths versus drainage areas of the tributary basins in a log-log plot.

basins, with their respective means and variances summarized in Table I. Interestingly, a comparable statistical regime has been identified in the distribution of distances between pairs of nodes in real-world Internet maps [9]. While the hydrological implications of this probability distribution lie beyond the scope of this work, it is worth noting that the Gamma distribution represents the well-known instantaneous unit hydrograph introduced by Nash [66].

The distributions of in-component and out-component sizes are related when averaged over all vertices in the plane tree. To illustrate this, let n indicate the total number of vertices in a plane tree. Let s_j denote the size of the in-component of a given node, say, v_j, and let Λ_d be the set of all vertices that are d steps away from the root. Looking at Fig. 2(b), it is straightforward to recognize that

$$\sum_{j=1}^{n} s_j = \sum_{d\geqslant 0}\sum_{v_j\in\Lambda_d} s_j = \sum_{d\geqslant 0}\sum_{L\geqslant d} N_L = \sum_{d\geqslant 0}(d+1)N_d, \quad (1)$$

where the last equality comes from rearranging terms in the summation, and N_d indicates the number of vertices at the $d$ht level. On the other hand, the l.h.s. of Eq. (1) can be recast by grouping vertices into classes of the same in-component size, so that the identity chain (1) reads as

$$\sum_{s\geqslant 1} s\mathcal{N}_s = \sum_{d\geqslant 0}(d+1)N_d, \quad (2)$$

where $\mathcal{N}_s$ denotes the number of in-components of size s, i.e., the number of plane subtrees with s vertices. Hence, dividing both sides by n, one obtains

$$\langle s\rangle_{\psi} = \langle d\rangle_{\Theta} + 1 = \langle t\rangle_{\hat{\psi}}. \quad (3)$$

Finally, by using the relation $\psi_s = \Phi^*_a$, where $a = 2s-1$, the first equality in Eq. (3) can be rewritten as

$$\langle d\rangle_{\Theta} = \tfrac{1}{2}(\langle a\rangle_{\Phi^*} - 1). \quad (4)$$

This provides a useful relation for estimating the mean level in a plane tree, given that the average value of the drainage area of tributary basins is known. By the same argument, an identity similar to Eq. (3) has also been derived in Ref. [49] for the special case of a planted tree. As we will see in Sec. III D, Eq. (3) directly follows from the generating function formalism applied to plane trees.

Finally, we tested the consistency of our approach with respect to Hack's law. As shown in Fig. 8, this relationship holds for our collection of tributary basins, thereby relating the in-degree of a vertex in the plane tree to the size of its in-component.

Hack's exponent for the Agri River attests around 0.68, and similar exponent values are observed for all the basins examined (see Table I). On the other hand, denoting α and β as the scaling exponents for tributary river lengths and drainage areas of tributary basins, respectively, the relation $\eta = (\beta-1)/(\alpha-1)$, derived in [34,36], yields Hack's exponent values (η), which are in excellent agreement with those estimated through data regression, reflecting a unified scaling behavior in the line graph of river networks.

III. RANDOM PLANE TREES WITH ARBITRARY DEGREE DISTRIBUTIONS

In this section, we adopt the approach introduced in Ref. [3], employing a model of directed, uncorrelated random graphs with arbitrary in-degree distributions to characterize both local and global statistical properties of the line graph of river networks. Our analysis reveals that the model effectively captures key behaviors of fluvial systems in the limit of infinite network size, while, in some cases, theoretical predictions do not align with empirical evidence—possibly due to degree-degree correlations, which we experimentally observe in natural drainage networks but which are not accounted for by the first-order random graph model. Section III D is the key part, where we derive a closed-form expression that relates the scaling exponent of drainage areas in tributary basins to that of tributary river lengths by applying the generating function

formalism to the in-component size problem in uncorrelated plane trees. This expression yields values that closely match the experimental findings discussed in Sec. II C. To address the limitations of the null model in the out-component analysis, we propose in Sec. IV a theoretical framework based on a tree-level dependent correlation measure.

A. Statistical distributions for a plane tree

As discussed in the previous section, the orientation of the vertices towards the root makes a plane tree a specific type of directed network. Its topological structure can be described by the joint degree distribution, denoted as f_{ui}, which represents the probability that a randomly selected node in the tree has out- and in-degree equal to u, and i, respectively [3,12,13]. Let n denote the total number of vertices in the plane tree. Referring to the scheme in Fig. 2(b), it follows by construction that $f_{0i} = \delta(i - i_r)/n$ and $f_{1i} = p_i[1-\delta(i - i_r)]$, where $\delta(\cdot)$ indicates the Kronecker delta function, i_r represents the number of incoming edges of the root (i.e., the maximum in-degree of the tree), and p_i is the in-degree distribution of nodes, which can take any discrete probability function. The joint degree distribution thus satisfies the following equation:

$$f_{ui} = \frac{1}{n}\delta(u)\delta(i - i_r) + p_i\delta(u - 1)[1 - \delta(i - i_r)], \quad (5)$$

with the normalization constraint given by $\sum_{ui} f_{ui} = 1$. By using the condition $p_{i_r} = 1/n$, the marginal distributions of in- and out-degrees are derived by summing the joint probability over its indices. Specifically, $p_i = \sum_u f_{ui}$ and

$$\hat{p}_u = \sum_i f_{ui} = \frac{1}{n}\delta(u) + z\delta(u - 1), \quad (6)$$

where $m = n-1$ is the total number of edges and $z = m/n$ is the mean degree of the tree, defined as the average number of both incoming and out-going edges. This is consistent with the condition that every edge in a directed graph must start and end at a vertex [3]. Note that in the limit of large system size, $z = 1$. Equation (6) provides the probability that a randomly chosen node in the plane tree has u out-going edges. All other distributions of interest can be immediately derived from Eq. (5). In particular, for the number of out-going edges, one obtains

$$\hat{r}_u = \sum_i \frac{uf_{ui}}{z} = \frac{u\hat{p}_u}{z} = \delta(u - 1) \quad (7)$$

and

$$\hat{q}_u = \sum_i \frac{if_{ui}}{z} = \frac{i_r}{m}\delta(u) + \frac{m - i_r}{m}\delta(u - 1), \quad (8)$$

which give the probabilities that a randomly chosen edge comes from and leads to a node with out-degree u, respectively. For the number of incoming edges, Eq. (5) gives

$$r_i = \sum_u \frac{if_{ui}}{z} = \frac{ip_i}{z} \quad (9)$$

and

$$q_i = \sum_u \frac{uf_{ui}}{z} = \frac{p_i}{z}[1 - \delta(i - i_r)] = \frac{p_i}{z}, \quad (10)$$

which stand for the probabilities that an edge randomly extracted arrives at and runs out of a node with in-degree i, respectively. The last equality in (10) holds for any $i \neq i_r$, and it indicates that, up to the mean degree z, the probability that a randomly chosen edge comes from a node with in-degree i is the same as the probability of finding a vertex with i in-coming edges. Adopting the terminology from Ref. [5], in a directed graph, every edge has two types of ending vertices: a source node, from which the edge emanates, and a target node, to which the edge leads. Thus, the probabilities in Eqs. (7) and (8) represent the out-degree distributions of the source and target nodes of a randomly chosen edge, respectively. Similarly, Eqs. (9) and (10) give the in-degree distributions of the target and source nodes of an edge selected at random. Using Eq. (10), the mean in-degree of the source nodes of a randomly chosen edge can be expressed as

$$\langle i\rangle_q = \sum_i iq_i = \frac{m - i_r}{m} = \sum_u u\hat{q}_u = \langle u\rangle_{\hat{q}}, \quad (11)$$

which represents the ratio of the total number of edges to the total number vertices effectively contributing to the mean. This interpretation arises because the probability that a randomly chosen edge comes from a node with in-degree i_r is zero. As a result, all edges entering the root must be excluded when averaging with respect to the probability q_i, and the root node itself must also be omitted. The second last equality in Eq. (11) indicates that the same expression applies when determining the mean out-degree of the target nodes of a randomly chosen edge, according to Eq. (8). This outcome is a general requirement for oriented graphs in absence of correlations between the degrees of adjacent vertices. In this case, $z\langle u\rangle_{\hat{q}}$ and $z\langle i\rangle_q$ represent the average number of the second neighbors reachable from a randomly chosen node and from which a random node can be reached, respectively. Both quantities are equal to the average value of the product of in-degree i and out-degree u at a given vertex in the graph [3], which is defined as $\langle ui\rangle_f = \sum_{ui} uif_{ui}$. Finally, regarding the correlations between in- and out-degrees of a given node, comparing Eqs. (5) and (6) reveals that the joint degree distribution does not factorize into product of separate degree distributions. In fact, the covariance function between in- and out-degrees is given by

$$\mathrm{cov}_{ui} = \langle ui\rangle_f - \langle u\rangle_{\hat{p}}\langle i\rangle_p = \frac{m}{n}\left[\frac{m - i_r}{m} - \frac{m}{n}\right] < 0. \quad (12)$$

Negative correlations are a straightforward consequence of our mapping, which generates a plane tree where the maximum number of incoming edges occurs at the root node. However, as n increases, the covariance approaches zero at a rate of $-i_r/m$, indicating that the plane tree becomes uncorrelated. Alternatively, in the limit of large system size, the joint degree distribution simplifies to $f_{ui} = p_i\delta(u-1) = p_i\hat{p}_u$, for any $i \neq i_r$.

Now, let us focus on the probability distribution of the number of incoming edges to a randomly chosen vertex. As demonstrated in Sec. II C through empirical analysis, the in-degree distribution of nodes in line graph of river networks follows a power-law, with a nonzero probability of finding a vertex with zero incoming edges. This probability, which

accounts for the number n_0 of leaves in the plane tree, must be specified separately. Therefore, the expression for the power-law degree distribution is modified as follows:

$$p_i = p_0\delta(i) + (1-p_0)\frac{i^{-\alpha}}{\zeta(\alpha)}[1-\delta(i)], \tag{13}$$

where $p_0 = n_0/n$, and $\zeta(\alpha)$ is the Riemann Zeta function. According to Eq. (13), the mean degree of the plane tree is given by

$$z = (1-p_0)\frac{\zeta(\alpha-1)}{\zeta(\alpha)}. \tag{14}$$

Moreover, by substituting Eq. (13) into Eqs. (9) and (10), and using Eq. (14), one obtains

$$r_i = \frac{i^{1-\alpha}}{\xi(\alpha-1)}[1-\delta(i)] \tag{15}$$

and

$$q_i = \frac{p_0}{z}\delta(i)[1-\delta(i-i_r)] + \frac{i^{-\alpha}}{\zeta(\alpha-1)}[1-\delta(i)][1-\delta(i-i_r)]. \tag{16}$$

Equation (15) reflects the fact that there are no edges leading to the leaves of the tree. In all other cases, the in-degree distribution of the target nodes for a randomly chosen edge follows a power law, as expected. Moreover, by definition of the Kronecker delta, Eq. (16) is consistent with the fact that there are no edges coming from the root. Thus, except for the root node and up to the mean degree of the network, the in-degree distribution of the source nodes of a randomly selected edge is the same as the in-degree distribution of nodes, as stated in Eq. (10).

B. The generating function formalism for plane trees

Once all the distributions of interest are at our disposal, we can calculate their probability generating functions as follows. Let us denote by

$$\omega(x,y) = \sum_{ui} f_{ui}x^u y^i \tag{17}$$

the generating function for the joint degree distribution of nodes, for u and i ranging from 0 to infinity. Following closely the arguments in Ref. [3], the generating functions for the probability distributions $\hat{p}_u$, $\hat{r}_u$, and $\hat{q}_u$ are given by

$$\omega(x,1) = \sum_{u=0}^{\infty}\hat{p}_u x^u = \hat{F}_0(x), \tag{18}$$

$$\frac{x}{z}\frac{\partial\omega(x,y)}{\partial x}\bigg|_{y=1} = \sum_{u=0}^{\infty}\hat{r}_u x^u = \hat{F}_1(x) = \frac{x\hat{F}_0'(x)}{z}, \tag{19}$$

and

$$\frac{1}{z}\frac{\partial\omega(x,y)}{\partial y}\bigg|_{y=1} = \sum_{u=0}^{\infty}\hat{q}_u x^u = \hat{R}_1(x). \tag{20}$$

Similarly, the generating functions for the probability distributions p_i, q_i, and r_i are

$$\omega(1,y) = \sum_{i=0}^{\infty}p_i y^i = F_0(y), \tag{21}$$

$$\frac{y}{z}\frac{\partial\omega(x,y)}{\partial y}\bigg|_{x=1} = \sum_{i=0}^{\infty}r_i y^i = F_1(y) = \frac{yF_0'(y)}{z}, \tag{22}$$

and

$$\frac{1}{z}\frac{\partial\omega(x,y)}{\partial x}\bigg|_{x=1} = \sum_{i=0}^{\infty}q_i y^i = R_1(y). \tag{23}$$

In our case, using Eq. (5), Eq. (17) simplifies to

$$\omega(x,y) = xF_0(y) + \frac{y^{i_r}}{n}(1-x), \tag{24}$$

and consequently, the equations from (18) to (20) specialize as follows:

$$\hat{F}_0(x) = \frac{1}{n} + \frac{m}{n}x, \tag{25}$$

$$\hat{F}_1(x) = x, \tag{26}$$

and

$$\hat{R}_1(x) = \frac{i_r}{m} + \frac{m-i_r}{m}x, \tag{27}$$

respectively, where we have used the well-known conditions $F_0(1) = 1$ and $F_0'(1) = z$.

On the other hand, substituting Eq. (13) into Eq. (21) immediately gives the generating function for the in-degree distribution of nodes:

$$F_0(y) = p_0 + (1-p_0)\frac{\mathrm{Li}_\alpha(y)}{\zeta(\alpha)} = p_0 + z\frac{\mathrm{Li}_\alpha(y)}{\zeta(\alpha-1)}, \tag{28}$$

where the last equality comes from Eq. (14), and $\mathrm{Li}_p(y)$ is the pth polylogarithm of y. Then, by taking the first derivative of Eq. (28) with respect to y, the last equality in Eq. (22) gives

$$F_1(y) = \frac{\mathrm{Li}_{\alpha-1}(y)}{\zeta(\alpha-1)}, \tag{29}$$

where we have used the relation $y\partial\mathrm{Li}_\alpha(y)/\partial y = \mathrm{Li}_{\alpha-1}(y)$. Finally, by taking the first derivative of Eq. (24) with respect to x, evaluated at $x = 1$, and substituting into Eq. (23), we obtain

$$R_1(y) = \frac{F_0(y)}{z} - \frac{y^{i_r}}{m} = q_0 + \frac{\mathrm{Li}_\alpha(y)}{\zeta(\alpha-1)} - \frac{y^{i_r}}{m}, \tag{30}$$

where the last equality follows from Eq. (28) and from Eq. (10) evaluated at $i = 0$. Note that, up to the mean degree z, Eqs. (28) and (30) differ only by the term $-y^{i_r}/m$. Therefore, excluding the root of the plane tree, we can set

$$R_1(y) = \frac{F_0(y)}{z} = q_0 + \frac{\mathrm{Li}_\alpha(y)}{\zeta(\alpha-1)}. \tag{31}$$

This is indeed a straightforward consequence of Eq. (10), which states that the probability distributions p_i and q_i coincide for all nodes except the root, regardless of the in-degree distribution of the vertices.

C. The sizes of out-components for a plane tree

Let us now examine the distribution of the sizes of the out-components for our directed graph. We denote by $\hat{\psi}_t$ and $\hat{\rho}_t$ the distributions of the total number of vertices reachable by a randomly chosen node and edge, respectively. These represent the probabilities that a node belongs to, and that an edge leads to, an out-component containing exactly t vertices in total. The corresponding generating functions are

$$\hat{H}_0(x) = \sum_{t=0}^{\infty} \hat{\psi}_t x^t \quad \text{and} \quad \hat{H}_1(x) = \sum_{t=0}^{\infty} \hat{\rho}_t x^t. \tag{32}$$

It has been proven [67] that the generating function for the out-component size distribution obeys the following equation:

$$\hat{H}_0(x) = x \sum_{u=0}^{\infty} \hat{p}_u [\hat{H}_1(x)]^u = x\hat{F}_0[\hat{H}_1(x)], \tag{33}$$

where the generating function $\hat{H}_1(x)$ satisfies the self-consistency condition:

$$\hat{H}_1(x) = x \sum_{u=0}^{\infty} \hat{q}_u [\hat{H}_1(x)]^u = x\hat{R}_1[\hat{H}_1(x)], \tag{34}$$

with the definitions provided in Eqs. (18) and (20). The particularly simple form of the generating functions for the out-degree distributions of the target vertices of a randomly chosen edge allows us to solve the system of Eqs. (33) and (34), which governs the sizes of the out-components for an uncorrelated plane tree. Specifically, by substituting Eq. (27) into Eq. (34), and solving for $\hat{H}_1$, we obtain

$$\hat{H}_1(x) = \frac{i_r x}{m - (m - i_r)x}. \tag{35}$$

Then, by substituting Eqs. (25) and (35) into Eq. (33), we obtain

$$\hat{H}_0(x) = \frac{x}{n} + \frac{m}{n} \frac{i_r x^2}{m - (m - i_r)x}. \tag{36}$$

The size distribution of the out-components is then given by

$$\hat{\psi}_t = \frac{1}{t!} \frac{d^t \hat{H}_0(x)}{dx^t}\bigg|_{x=0}. \tag{37}$$

Using current symbolic manipulation programs to evaluate the tth derivative of Eq. (36) leads to

$$\hat{\psi}_t = \frac{i_r}{n} \left(\frac{m - i_r}{m} \right)^{t-2}, \tag{38}$$

for each $t \geqslant 2$. The initial values are $\hat{\psi}_0 = 0$, because every component has at least one vertex, and $\hat{\psi}_1 = 1/n$, since the probability that a vertex belongs to a cluster of size 1 is the same as the probability of finding a vertex with zero out-going edges.

The same result can also be derived by considering the probability that a randomly selected edge originates from a node with in-degree i. As discussed in Sec. II C, the structure of the plane tree implies that the probability of a randomly chosen vertex belonging to an out-component of size t is equivalent to the probability of finding a neighbor who is $d = t-1$ steps away from the root node. The number of the first neighbors of the root is i_r, and the mean in-degree of the source nodes for an edge is given by the first equality in Eq. (11). This represents the average number of vertices two steps away from the root via a particular first neighbor. Consequently, the average number of the second neighbors from which the root node can be reached is

$$N_2 = i_r \langle i \rangle_q. \tag{39}$$

According to Ref. [3], for uncorrelated networks, the average number of edges entering any vertex at a distance $d-1$ from the root is also $\langle i \rangle_q$. Hence, the average number of neighbors at distance d from the root node is

$$N_d = N_{d-1} \langle i \rangle_q, \tag{40}$$

where $B_d = N_d/N_{d-1}$ is the branching ratio of the configuration model. Finally, by recursion and making use of the second equality in Eq. (11), we get

$$N_d = i_r \langle i \rangle_q^{d-1} = i_r \left(\frac{m - i_r}{m} \right)^{d-1}, \tag{41}$$

for any $d \geqslant 1$, which is the same as Eq. (38).

It is important to note that Eq. (40) indicates that the branching ratio in the configuration model remains constant for uncorrelated networks, and corresponds to the mean in-degree of the source nodes for a randomly selected edge. This result aligns with predictions from branching process theory [57]. In fact, a plane tree can be viewed as a perfect i-ary tree [68], with randomization introduced into the elementary branching process. Starting from the root node, the tree expands by having each vertex branch into i distinct neighbors at each generation, with each neighbor assigned a branching probability p_i. Consequently, random branching trees can be considered as plane trees, where the in-degree distribution of nodes is determined by the branching probability distribution. For a scale-free plane tree, this probability is assumed to follow the form given by Eq. (13).

Borrowing the approach outlined in Ref. [69], we define the branching ratio based on the in-degree distribution of the source nodes of a given edge. Specifically, it is calculated as the average number of offspring generated by vertices at a distance $d-1$ from the root. To express this, let us introduce the conditional probability $q_{i|d-1} = e_{i|d-1}/e_{d-1}$, where $e_{i|d-1}$ denotes the number of edges originating from nodes at level $d-1$ with in-degree i, and e_{d-1} represents the total number of edges emanating from nodes at that level. The branching ratio is then given by

$$B_d = \sum_i i q_{i|d-1} = \frac{\sum_i i e_{i|d-1}}{e_{d-1}} = \frac{N_d}{N_{d-1}}, \tag{42}$$

for any $d \geqslant 2$, and $B_1 = i_r$, as specified by the initial condition. This formulation allows the branching ratio to vary with the tree level, enabling growth analysis at different levels of the tree. However, for uncorrelated branching processes, the branching ratio remains independent of d for the random tree [69]. To gain further insight, we adopt the standard formalism from branching process theory. Consider the subtree of a plane tree formed by excluding the leaves at the maximum level,

effectively neglecting the growth interruption due to the finite system size. Let e_i represent the number of edges leaving all nodes with in-degree i, and e denote the total number of edges in this pruned subtree. The average branching ratio, which represents the average number of children per node, is given by

$$\langle i\rangle_q = \sum_i i q_i = \frac{1}{e}\sum_i i e_i = \frac{1}{e}\sum_i i \sum_d e_{i|d-1}$$
$$= \frac{1}{e}\sum_d \sum_i i e_{i|d-1} = \frac{1}{e}\sum_d N_d, \tag{43}$$

where $q_i = e_i/e$, and the identity $\sum_i i e_{i|d-1} = N_d$ is used. The level ranges from 2 to the height of the plane tree, denoted by Δ. For uncorrelated branching trees, where $B_d = B$, Eq. (43) reduces to

$$\langle i\rangle_q = B, \tag{44}$$

which corresponds to Eq. (40), except that the in-degree distribution of the source nodes for a randomly chosen edge now refers to the subtree with a total of $e = m - N_\Delta$ edges [70].

Let us now explore this point in greater detail. Both the random graph model for uncorrelated networks with arbitrary degree distributions and the theory of uncorrelated branching processes predict that the average branching ratio is equal to the mean in-degree of source nodes for a given edge [as expressed in Eqs. (40) or (44), depending on whether the average is taken over a total of m or e edges, respectively]. However, this result does not align with our empirical observations. Specifically, in the data we analyzed, the number of vertices that are d steps away from the root of plane trees does not follow the exponential function described in Eq. (41). Instead, the probability distribution of the tree levels closely fits a Gamma distribution (see Sec. II C). As a result, the branching ratio adheres to the following scaling law:

$$\frac{N_d}{N_{d-1}} \sim \left(\frac{d}{d-1}\right)^{\gamma-1}, \tag{45}$$

where γ is the scale parameter of the Gamma distribution. The inset in Fig. 9 displays a log-log plot of the branching ratio for the Agri River as a function of $d/(d-1)$, showing an increasing power-law dependence.

The estimated value of γ, obtained by fitting B_d versus $d/(d-1)$, is approximately 8.52, which is in close agreement with the observed Gamma scale parameter of around 8.06. Similar trends are observed across all the examined basins (see Table I). Additionally, for small d, the ratio $d/(d-1)$ decays roughly as a power-law with scaling exponent ϑ, allowing Eq. (45) to be rewritten as

$$B_d \sim d^{-\tau}, \tag{46}$$

where $\tau = \vartheta(\gamma-1)$. The blue line in Fig. 9 illustrates the power-law decay of the branching ratio for the Agri River, plotted on a bilogarithmic scale. The estimated scaling exponent τ is approximately 4.19. As shown in Fig. 9, this scaling behavior is consistently observed across all basin studied, with exponent values reported in Table I.

In contrast to the Horton-Strahler scheme [46], which results in a constant bifurcation ratio, the branching ratio in the line graph of real river networks exhibits a power-law decay. This behavior uncovers a scaling law in fluvial systems, which effectively captures the growth rate of tributary rivers under suitable renormalization of the drainage network.

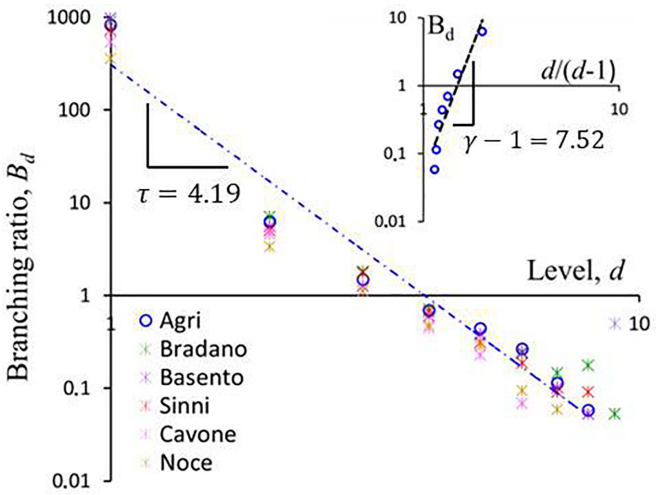

FIG. 9. Empirical evidence of the scaling law governing the decay of the branching ratio as a function of the plane tree level. The blue circles represent the experimental values of the branching ratio for the Agri River, while the blue dashed straight line fits the data on a log-log scale. The same scaling behavior is consistently observed across all other basins examined (see legend). Inset: A log-log plot of the branching ratio as a function of the ratio $d/(d-1)$ for the Agri River. The resulting scaling exponent is in good agreement with the scale parameter of observed Gamma distribution of topological distances (see Table I).

It is worth noting that the scaling law in Eq. (46) is not predicted by a first-order random graph model, as it directly arises from the Gamma distribution for topological distances in the plane tree. On the other hand, correlations between the in-degree of neighboring vertices in real river networks (which we will discuss in Sec. IV) imply that the number of incoming edges at the endpoints of a given edge is not independent. Consequently, the mean in-degree of the source nodes of a given edge does not govern the average number of vertices at a distance d from the root node, as in the case of uncorrelated networks, where $\langle i\rangle_q$ is central to Eq. (41). Therefore, we argue that the observed Gamma distribution for the plane tree levels may be a direct consequence of the correlation structure present in real-world drainage networks. In Sec. IV to address the limitations of the random graph model for uncorrelated plane trees, we will introduce a second-order approximation that refines Eq. (42). This approach enables the interpretation of the branching ratio as a specific correlation measure between the number of vertices on two consecutive levels, while Eq. (46) witnesses the presence of these correlations.

D. The sizes of in-components for a plane tree

In this section, we analyze the distribution of the in-component sizes for uncorrelated plane trees, which represents the probability of finding a plane subtree with s vertices, i.e., the distribution of the total contributing area of the basin

drained by the corresponding tributary tree (see Sec. II B). Let ψ_s and ρ_s denote the distributions of the total number of vertices from which a randomly chosen node and edge can be reached, respectively. These distributions also provide the probabilities that a node and an edge, respectively, are the root and the edge outgoing from the root of a plane subtree with s vertices. The corresponding generating functions are given by

$$H_0(y) = \sum_{s=0}^{\infty} \psi_s y^s \quad \text{and } H_1(y) = \sum_{s=0}^{\infty} \rho_s y^s. \tag{47}$$

As proven in Ref. [3], a pair of equations similar to Eqs. (33) and (34) governs the sizes of the in-components:

$$H_0(y) = y \sum_{i=0}^{\infty} p_i [H_1(y)]^i = y F_0[H_1(y)] \tag{48}$$

and

$$H_1(y) = y \sum_{i=0}^{\infty} q_i [H_1(y)]^i = y R_1[H_1(y)], \tag{49}$$

where we have used the definitions provided in Eqs. (21) and (23).

Given the system of Eqs. (48) and (49), the Lagrange inversion theorem [30] enables us to derive an explicit expression for the in-component size distribution in a network with arbitrary in-degree distribution. In general, for any two formal power series related by Eq. (49), the Lagrange inversion formula states that the sth coefficient in the formal power series $F_0[H_1(y)]$ is given by

$$[y^s] F_0[H_1(y)] = \frac{1}{s} [y^{s-1}] F_0'(y) [R_1(y)]^s, \tag{50}$$

where the shorthand notation $[y^s]G(y)$ refers to the coefficient of y^s in the Taylor expansion of $G(y)$ at the point $y = 0$, i.e., the inverse operation to the generating function transform. Using Eq. (48), the l.h.s. of Eq. (50) reads as $[y^s]F_0[H_1(y)] = [y^{s+1}]H_0(y) = \psi_{s+1}$. On the other hand, using the first equality in Eq. (31) and rearranging, the r.h.s. of Eq. (50) becomes $s^{-1}[y^{s-1}]F_0'(y)[R_1(y)]^s = z^{-s}(s+1)^{-1}[y^s][F_0(y)]^{s+1}$. From this, we get

$$\psi_{s+1} = \frac{1}{z^s} \frac{1}{(s+1)!} \left. \frac{d^s [F_0(y)]^{s+1}}{dy^s} \right|_{y=0}. \tag{51}$$

For undirected graphs, a similar equation was derived in Ref. [71] using the same technique, with a slightly different formulation presented earlier in Ref. [64]. The differences between Eq. (51) and the results in Refs. [64,71] arise from the fact that we are dealing with a plane tree, whose topological structure is encapsulated in Eq. (31).

In principle, Eq. (51) allows for the calculation of the probability that a randomly selected vertex belongs to an in-component of size $s+1$ by taking the sth derivative of the generating function for the in-degree distribution of nodes. In a geomorphological context, Eq. (51) indicates that the only information required to analytically recover the distribution of drainage areas in tributary basins is encapsulated in the length distribution of tributary rivers. This suggests that the aggregation mechanism of total contributing areas is governed by the hierarchical development of the channel network around tributary rivers, with the probability distribution of their lengths serving as the fundamental distribution, thereby offering a potential combinatorial foundation for Hack's law.

Various methods for solving equations like Eq. (51) have been proposed in the literature. For instance, using symbolic manipulation programs, Newman derived a closed-form expression for ψ_s in the case of networks with Poisson, exponential, and Yule degree distributions [64]. Additionally, asymptotic power-law distributions for in-component sizes were obtained in Ref. [71] by exploiting the multiplication property of generating functions and appealing to a generalized form of the central limit theorem.

In this study, leaving aside the task of calculating the complete distribution of in-component sizes, we take advantage of the closed-form expression available for average properties of clusters to derive useful relationships for practical applications, such as those encountered in geomorphological studies. Closely following the derivation in Ref. [3], the mean size of the set of vertices from which a randomly chosen node can be reached is given by $\langle s \rangle_{\psi} = H_0'(1) = 1 + z H_1'(1)$. The expression for $H_1'(1)$ is obtained by differentiating and rearranging Eq. (49), yielding

$$\langle s \rangle_{\psi} = 1 + \frac{z}{1 - R_1'(1)}, \tag{52}$$

where $R_1'(1) = \langle i \rangle_q$ is the mean in-degree of the source nodes of a randomly chosen edge, according to Eq. (23). By applying the same technique to Eqs. (33) and (34), one obtains the analogous relation:

$$\langle t \rangle_{\hat{\psi}} = 1 + \frac{z}{1 - \hat{R}_1'(1)}, \tag{53}$$

where $\hat{R}_1'(1) = \langle u \rangle_{\hat{q}}$ is the average out-degree of the target nodes of a randomly selected edge, according to Eq. (20). Therefore, by comparing Eqs. (52) and (53) and using Eq. (11), one arrives at the relation (3) derived in Sec. II C through a different reasoning.

Now, a theoretical relationship between the scaling exponents of the distributions of in-component sizes and the in-degree of nodes can be derived from Eq. (52). Specifically, by using the second equality in Eq. (11) and the relations $\psi_s = \Phi_a^*$ and $p_i = \Pi_l^*$, with $a = 2s-1$ and $l = i+1$, Eq. (52) can be rewritten as

$$\frac{1}{2}(\langle a \rangle_{\Phi^*} + 1) = 1 + (\langle l \rangle_{\Pi^*} - 1)\frac{m}{i_r}. \tag{54}$$

Then, based on the experimental evidence discussed in Sec. II C, we can impose that the mean drainage area in tributary basins and the mean length of tributary rivers satisfy the expression valid for the average value of a random variable with a power-law discrete density function. This leads to the following relations: $\langle a \rangle_{\Phi^*} = \zeta(\beta-1)/\zeta(\beta)$ and $\langle l \rangle_{\Pi^*} = \zeta(\alpha-1)/\zeta(\alpha)$. After a little algebra and using a continuous approximation [72], Eq. (54) yields the following synthetic relation:

$$\beta = 2 + \frac{i_r}{2m}(\alpha - 2), \tag{55}$$

where i_r is the in-degree of the root node (i.e., the length of the main river in the basin), and $m = n-1$ denotes the total

number of edges in the plane tree, i.e., the number of sources in the channel network, minus one. As shown in Table I, Eq. (55) provides an excellent fit to the real data. It is also worth noting that classical geomorphological analysis yields a relationship similar to that in Eq. (55), specifically $\beta = 2-1/\alpha$ [40], which was derived using fractal arguments and by invoking Hack's law.

Two noteworthy observations can be drawn from Eq. (55). First, based on an uncorrelated random graph model, Eq. (55) accounts for the large-scale organization of river networks, as encoded in the combinatorial structure of the plane tree described by Eqs. (48) and (49), while also adhering to the observed topological structure, which is characterized by scale-free distributions of both tributary river lengths and drainage areas in tributary basins. While simplifying to an average-parameter form of Eq. (51), Eq. (55) again reflects the influence of elongation statistics on aggregation statistics.

Second, the ratio on the r.h.s. of Eq. (55) approaches zero in the large n limit, since $i_r \ll m$. Consequently, in this limit, Eq. (55) predicts $\beta = 2$ for any fluvial system, regardless of the scaling exponent α of the in-degree distribution. This asymptotic value for β suggests that the aggregation patterns of total contributing areas are an inevitable consequence of the combinatorial structure of the line graph of river networks. Under a coarse-grained description of the system, and assuming that both in-component size and node degree distributions follow power-law decays, a value of $\beta = 2$ is expected, irrespective of the detailed topology of connections in the channel network. This may support the hypothesis of a universal scaling exponent for the in-component size distribution in growing random networks [18], as well as in supercritical stochastic trees, whether branching [17,19] or planted [20].

We conclude this section with a technical note. The giant in-component in a directed network is defined as the largest connected in-component whose size is of the same order as the entire graph, n. The out-component is defined similarly. Below the phase transition, in the regime where no giant components exist, each cluster of vertices is finite in size and contains no cycles. As a result, a plane tree does not undergo a phase transition, as it consists of a single, finite, loopless, connected component. Indeed, Eq. (52) diverges only when $R_1'(1) = 1$, while in our case, we find $R_1'(1) = (m - i_r)/m < 1$. For $i_r = 0$, the plane tree degenerates to a single node. In all other cases, the condition $R_1'(1) = 1$ is asymptotically approached in the limit of large system size, where each node in an infinite plane tree trivially acts as the root of another large plane tree. Finally, the symmetry of Eq. (11) ensures that the same condition governs the phase transition for the giant out-component. These observations make our experimental findings consistent with the theoretical predictions in Ref. [64], whereby, below the phase transition, a network with power-law distributed degrees is also expected to exhibit a power-law tail in the component size distribution.

IV. THE CORRELATION PATTERNS IN COMPLEX RIVER NETWORKS

Although the degree distribution is crucial for characterizing the local topology of graphs, many real-world networks display a more intricate global structure, as evidenced by the nontrivial correlations between the degrees of adjacent vertices [5–11]. Therefore, a comprehensive understanding of connection patterns in complex networks requires specifying how vertices are paired with each other based on their degree classes or hierarchical levels.

A. Length-length correlation structure in tributary rivers

Correlation analysis has a long-standing tradition in geomorphology. For instance, long-range spatial correlations in the width function of river networks have been identified through the examination of the power-law decay in its power spectrum [58].

In this section, we investigate the correlation properties of tributary river lengths using the line graph representation of natural drainage networks, which maps lengths of tributary rivers onto the in-degrees of nodes, thereby applying the complex network formalism. In general, for oriented graphs, there are four possible ways to define a correlation measure between the degrees of neighboring vertices, depending on whether the in-degree or out-degree of the vertices at the ends of an edge is considered [5]. Given that the complex topology of plane trees is revealed by examining the distribution of incoming edges at a node, we search for correlations between the in-degrees, denoted as h and k, of the target and source vertices of an edge, respectively. This can be accomplished using the nonsymmetric edge matrix, here denoted as e_{hk}, where the entries count the number of directed edges leaving a vertex with in-degree k and pointing to a vertex with h incoming edges [21–23]. The joint degree-degree distribution, given by $g_{hk} = e_{hk}/m$, then represents the fraction of edges in the plane tree connecting vertices with in-degree k and h, out of a total of m edges.

From a geomorphological perspective, the joint distribution g_{hk} represents the probability that a tributary river of length $k+1$ flows into another of length $h+1$, thereby describing how rivers connect to each other in terms of their lengths. Marginal distributions follow from the summation rules:

$$\sum_k g_{hk} = r_h = \sum_u h\frac{f_{uh}}{z} \tag{56}$$

and

$$\sum_h g_{hk} = q_k = \sum_u u\frac{f_{uk}}{z}, \tag{57}$$

with the normalization constraint given by $\sum_{hk} g_{hk} = 1$. Compared to Eqs. (9) and (10), in Eqs. (56) and (57), the index i is replaced by h or k depending on whether the vertex is the target or the source node of an edge, respectively. As expected [11], unlike in undirected graphs, where the distributions of edges at the two endpoints of an edge coincide, Eq. (56) gives the fraction of edges entering all vertices of in-degree h, out of a total of m edges. This fraction represents the probability that a randomly chosen edge leads to a vertex with in-degree h. Similarly, Eq. (57) enumerates the edges leaving all vertices of in-degree k, providing the probability that a randomly selected edge originates from a vertex with k incoming edges.

For the purpose of comparing different networks, and following Ref. [5], we refer to the normalized covariance

function between the in-degrees, denoted as h_l and k_l, of the target and source nodes for a directed edge l, respectively. Known as the Pearson correlation coefficient [73], this measure is defined as

$$R = \frac{m^{-1}\sum_l (h_l - \langle h_l\rangle)(k_l - \langle k_l\rangle)}{\sigma_h \sigma_k} = \frac{\langle hk\rangle_g - \langle h\rangle_r \langle k\rangle_q}{\sigma_r \sigma_q}, \tag{58}$$

where the mean and standard deviation for h_l are $\langle h_l\rangle = \sum_l h_l/m$ and $\sigma_h = \sqrt{m^{-1}\sum_l (h_l - \langle h_l\rangle)^2}$, respectively, with sums taken over all directed edges in the plane tree. Averages for k_l are defined in a similar fashion. To obtain the second equality in Eq. (58), we moved from summation over edges to summation over degree classes. The average values are as follows: $\langle hk\rangle_g = \sum_{hk} hk g_{hk}$, which represents the mean of having in-degrees h and k at the target and source nodes, respectively; $\langle h\rangle_r = \sum_h h r_h$ and $\sigma_r^2 = \langle h^2\rangle_r - \langle h\rangle_r^2$ are the mean and variance of the in-degree of target nodes, respectively. Similarly, $\langle k\rangle_q = \sum_k k q_k$ and $\sigma_q^2 = \langle k^2\rangle_q - \langle k\rangle_q^2$ represent the mean and variance of the in-degree for source vertices.

In the absence of correlations between the in-degrees of adjacent nodes, the joint degree-degree distribution factorizes into the product of its marginal distributions, i.e., $g_{hk} = r_h q_k$, and the Pearson coefficient equals zero, indicating that edges attach to nodes independently of their degrees [10,11]. In contrast, a positive covariance function implies assortative mixing by degree, meaning that vertices tend to preferentially attach to others with similar degrees. Conversely, a negative covariance indicates disassortative behavior, where high-degree vertices are more likely to connect to low-degree ones, and vice versa [5,10,11]. Earlier works have proposed a simple classification of complex networks, where social networks are assortative [10,11], while technological networks [7–11] and biological systems [6,10,11] exhibit disassortative mixing. However, the study in Ref. [5] highlights the key role of edge direction, showing that many real-world directed networks exhibit a mixture of both positive and negative correlations, depending on the degree types at the endpoints of the edges.

While a drainage network functions as a system for transporting water and sediments, similar to other communication or transportation networks, our empirical analysis (see Table I) reveals that river networks exhibit assortative mixing by degree. This suggests a correlation structure akin to that observed in social systems, often attributed to hierarchical division into groups and communities [74]. Analogously, the positive correlations observed in real river networks arise from the hierarchical structure inherent in the corresponding line graphs. Our mapping induces positive correlations by construction: it recursively identifies tributary rivers of progressively shorter length, leading to the corresponding vertices in the plane tree being ordered such that the in-degree of each vertex exceeds that of its children. As a result, for any edge l, the condition $h_l > k_l$ holds, reflecting the ordering scheme that prevents tributary rivers from flowing into smaller rivers. The positive value of the Pearson coefficient thus directly emerges from the dispersion of data points in hk plane, where all pairs of in-degrees for target and source nodes of a given edge fall below the bisector of first and third quadrants.

On the other hand, although positive correlations are expected by construction, we emphasize that our experimental findings reveal an additional intriguing regularity in real river networks: the strength of assortativity remains nearly constant across all the basins examined, with Pearson coefficient values consistently around 0.23—a value also observed in other complex networks across various fields [5]. The assortative behavior is also reflected in the shape of the distribution of plane tree levels (see Sec. III C). The increasing regime of the width-type function near the root is consistent with the behavior of longer tributary rivers flowing into even longer ones. Conversely, farther from the root node, the decreasing trend in the distribution aligns with shorter tributary rivers feeding into others of similar length. Both of these conditions contribute to the development of a hierarchical structure that tends toward a perfect positive correlation, which is counteracted by certain shorter tributary rivers meeting longer ones, causing the Pearson correlation to deviate from 1. The fact that all basins examined exhibit a similar deviation from this maximum value suggests a comparable hierarchical organization, particularly with respect to the number of short-length rivers entering longer ones. This number is not large enough to undermine the assortative trend, so that, on average, tributary rivers are more likely to flow into others of similar length. This prevents the channel network from resembling a backbone, with smaller tributary branches feeding into the main river. Instead, the positive Pearson coefficient value confirms that fluvial systems tend to enlarge as they elongate, a behavior also captured by Hack's law [36].

However, the Pearson coefficient provides an average measure of correlation that fails to capture the hierarchical organization of vertices across the plane tree levels. The way tributary rivers organize themselves according to their stream orders calls for a more refined correlation measure, which will be introduced in the next section.

To conclude this comment, we underscore that correlations are defined here with respect to neighboring vertices in the plane tree, which do not correspond to adjacent sites in the underlying river network. Consequently, the observed assortativity cannot be directly attributed to any local geological control, as tributary rivers span the entire catchment, often crossing landscape regions with markedly different orographic and lithological characteristics.

B. The branching ratio in the configuration model as a measure of correlation

To relate the two-point correlations to the variability of the branching ratio as a function of the tree levels—both phenomena observed in real river networks—we begin by introducing the conditional distribution $q_{k|h}$. This represents the probability that the source vertex of a given edge has an in-degree of k, provided that its target node has h incoming edges [75]. By definition, it follows that $g_{hk} = r_h q_{k|h}$. In general, for correlated networks, the conditional probability explicitly depends on both k and h, while in uncorrelated networks, $q_{k|h}$ reduces to q_k, as each edge emanates from a source node with in-degree k, regardless of the in-degree h of the target node.

Now, we adapt the approach proposed in Ref. [26] to planted trees. Specifically, we define the average branching

ratio of an edge entering a target vertex with in-degree h as follows:

$$B_h = \sum_k k q_{k|h}. \tag{59}$$

This represents the average in-degree of the nearest neighbors of target vertices with in-degree h, which, in its undirected formulation, is commonly used to detect correlations in real-world networks [6–9]. When no two-point correlations exist, B_h is independent of h, and Eq. (59) immediately reduces to (40), with k in place of i. In contrast, in correlated directed networks, the mean branching ratio can be defined as

$$B = \sum_h r_h B_h = \sum_k k \sum_h r_h q_{k|h} = \sum_k k q_k = \langle k \rangle_q, \tag{60}$$

where the third equality follows from the relation $g_{hk} = r_h q_{k|h}$ and the first equality in Eq. (57). The identity chain in Eq. (60) thus recovers Eq. (40), with k in place of i. This is not just a trivial result indicating that the mean branching ratio is a network constant; the independence of B from h also implies that information about the correlation structure is lost when averaging over h. Due to this dilution, the probability $q_{k|h}$ does not capture the growth rate of nodes across the tree levels, a feature observed in real fluvial systems (see Sec. III C).

In an effort to develop a more refined correlation measure that simultaneously reflects both the statistical dependence between the in-degrees of neighboring vertices and the variability of the branching ratio as a function of topological distances, we introduce the d-conditional distribution, denoted as $q_{k|h|d-1} = e_{hk|d-1}/e_{h|d-1}$. Here $e_{hk|d-1}$ represents the number of edges that connect a source node with in-degree k at level $d-1$ to a target node with in-degree h at the previous level, while $e_{h|d-1}$ indicates the total number of edges leading to nodes at level $d-2$ with h incoming edges. Compared to the conditional probability $q_{k|h}$, which provides a distributed correlation measure by pairing adjacent vertices with a prescribed number of incoming edges throughout the entire network, the d-conditional distribution decomposes $q_{k|h}$ into subclasses by grouping adjacent vertices with specified in-degrees that reside on consecutive levels. By projecting two-point correlations onto topological distances, the d-conditional probability serves as a hierarchical correlation index that, from a geomorphological perspective, illustrates how tributary rivers flow into one another in relation to both their lengths and nesting order.

Using the d-conditional distribution, we can define the conditional branching ratio as the average number of children generated by vertices at a distance $d-1$ from the root node, provided their outgoing edges lead to vertices with in-degree h. This is expressed as

$$B_{h|d} = \sum_k k q_{k|h|d-1} = \frac{\sum_k k e_{hk|d-1}}{e_{h|d-1}} = \frac{N_{d|h}}{N_{d-1|h}}, \tag{61}$$

where the last term represents the ratio between the number of second and first neighbors of vertices at level $d-2$ with h incoming edges. Then, denoting by e_{d-1} the total number of edges entering vertices at level $d-2$, the branching ratio is given by

$$B_d = \sum_h r_{h|d-1} B_{h|d} = \sum_h \frac{N_{d-1|h}}{N_{d-1}} \frac{N_{d|h}}{N_{d-1|h}} = \frac{\sum_h N_{d|h}}{N_{d-1}} = \frac{N_d}{N_{d-1}}, \tag{62}$$

where the conditional probability, $r_{h|d-1} = e_{h|d-1}/e_{d-1} = N_{d-1|h}/N_{d-1}$, represents the fraction of the first neighbors of vertices at level $d-2$ with h incoming edges, relative to the total number N_{d-1} of vertices at distance $d-1$ from the root. The identity chain in Eq. (62) stands for the correlation-dependent counterpart of Eq. (42). Note that Eq. (62) once again defines the branching ratio; however, unlike Eq. (60), it is derived by averaging the d-conditional distribution over both h and k, without simplifying to a network constant. Although the dependence on h disappears, the branching ratio still retains similarities to the d-conditional distribution, particularly in its dependence on topological distances. This approach can be used to describe a type of correlation, where the probability of a randomly chosen edge coming from a node at a given level is constrained by the number of vertices at the preceding level, independently of the node's in-degrees. In other words, the branching ratio may be interpreted as an indicator of correlation between the number of vertices on two consecutive levels of the plane tree. Building on these considerations, the scaling law observed in Eq. (46) suggests that real-world river networks belong to a specific subclass of correlated graphs, which we refer to as d-correlated networks. These networks exhibit hierarchical correlations, encoded in a d-dependent branching ratio, which emerge from the arrangement of elements across topological levels. From a geomorphological perspective, this reflects the hierarchical organization of the channel network around tributary rivers, accounting for how many tributaries of given order flow into rivers of lower order.

V. DISCUSSION

Taking advantage of the fact that a plane tree can be viewed as a perfect i-ary tree with random variations in the branching process at every node, this section reviews studies on stochastic branching trees to present our results from a perspective slightly different from that of the configuration model [3,54]. In what follows, we first classify river networks in the context of branching process theory [57], and then offer an alternative explanation for the scaling exponent observed in the exceedance distribution of drainage areas in tributary basins within the same framework.

As discussed in Sec. III C, the growth rate of vertices in a plane tree is governed by the average number of children generated by vertices at a specified distance from the root, i.e., the branching ratio, as given by Eq. (42). In the absence of correlations between the in-degrees of adjacent vertices, the branching ratio becomes a network constant, corresponding to the average in-degree of the source nodes of a randomly chosen edge, as expressed in Eqs. (43) and (44). Based on this parameter, random trees are classified as subcritical, critical, and supercritical, depending on whether the average branching ratio is less than, equal to, or greater than 1, respectively [57]. Specifically, a critical branching tree is one in which

each generation of offspring, on average, is identical to the preceding one, indicating that branches neither disappear nor proliferate. In contrast, in the subcritical regime, the number of children tends to zero as the system size increases. Finally, supercritical trees experience exponential growth as the branching process progresses. Analytic arguments in Ref. [19] illustrate that the size of subtrees in a supercritical random branching tree follows a power-law distribution with a scaling exponent of 2. According to the classification outlined above, the branching ratio for the configuration model is given by Eq. (40), with $\langle i \rangle_q < 1$, leading to the exponential decay of the number of vertices at distance d from the root, as expressed in Eq. (41). This suggests that the predictions derived from the uncorrelated random graph model employed in this study indicate that our directed network originates from a subcritical branching process occurring in an uncorrelated manner.

On the other hand, due to the assortative mixing observed in the six rivers examined, the branching ratio shows significant fluctuations across different tree levels. As a result, the mean in-degree of source nodes for an edge does not match the average branching ratio, as predicted by Eq. (44) for uncorrelated branching trees. This indicates that our plane tree does not fit into the simple classification of branching processes outlined above. However, to assign an appropriate growth regime to correlated river networks, one can estimate the average branching ratio as $\langle B_d \rangle = \Delta^{-1} \sum_d N_d/N_{d-1}$, where d ranges from 1 to the maximum level Δ of the plane tree. As reported in Table I, all observed values of $\langle B_d \rangle$ are significantly greater than 1, primarily due to the large number of incoming edges to the root node. At the coarse-grained description provided by the average parameter $\langle B_d \rangle$, the d-dependent growth of nodes in real-world drainage networks may be viewed as resulting from a branching process in a highly super-critical regime. Thus, the scaling exponent of the in-component size distribution, which we predicted to be 2 within the context of a random graph model [3], is consistent with the theoretical explanations provided in Ref. [19].

Regarding this result, the work by Ref. [20] is particularly relevant and is briefly discussed here for comparative purposes. Through numerical simulations, the authors found that the exceedance distribution of subtree sizes in a random binary tree decays as a power law, with a scaling exponent that increases with the average branching ratio, B. Specifically, this exponent ranges from 0.5 to 1 as B varies between 1 and 2 [76]. Starting from a perfect 2-ary tree [77], the authors allowed a fraction of vertices to randomly branch, showing that the scaling exponent decreases as the randomness in the tree topology increases. However, perfect 2-ary trees were found to be highly robust against random noise, as the decreasing trend was observed to be very slow. This result led the authors to propose the scaling exponent as a measure of how closely a random binary tree approximates a perfect 2-ary tree: the closer a random binary tree is to a completely bifurcating tree, the smaller the deviation of the scaling exponent from 1. Additionally, the same study observed that a perfect 2-ary tree displays the highest hierarchical density, meaning the highest concentration of nodes near the root. Thus, the authors concluded that the scaling exponent for the exceedance distribution of subtree sizes serves as an index of hierarchical density. In this context, the observed scaling exponent of around 0.43 suggests that a binary-tree representation of river networks, in which nodes fill the Euclidean space in a spanning manner, is far from a perfect 2-ary tree, where nodes are more densely concentrated around the outlet.

We can easily extend the results from Ref. [20] to our context. Specifically, as a generalization of the binary case, it can be proven that a perfect i-ary tree also exhibits a power-law exceedance distribution of subtree sizes, with a scaling exponent equal to 1. To this end, let $\sum_u N_u = (i^{d+1} - 1)/(i-1)$ represent the total number of vertices from the root node up to level d, where $N_u = i^u$ denotes the number of vertices at the uth level, with u ranging from 0 to d. Then let $w = \Delta - d$ indicate the topological distance from a leaf node to a given vertex. Notice that the probability of finding a vertex at a level less than or equal to d is the same as the probability of finding a vertex at a distance greater than or equal to w. Therefore, the exceedance distribution of the distance w is given by

$$\text{Prob}[v \geqslant w] \propto (i^{\Delta+1-w} - 1)/(i-1). \tag{63}$$

It is straightforward to recognize that the size of the subtrees at a distance w, denoted as s_w, satisfies the following recurrence relation: $s_w = is_{w-1} + 1$, with the initial condition $s_0 = 1$. The solution to this recurrence is

$$s_w = \sum_{v=0}^{w} i^v = (i^{w+1} - 1)/(i-1). \tag{64}$$

Finally, by solving Eq. (64) for w and substituting into Eq. (63), the exceedance distribution of the subtree sizes is obtained:

$$\text{Prob}[\sigma \geqslant s] \propto \{i^{\Delta+2}[(i-1)s + 1]^{-1} - 1\}/(i-1), \tag{65}$$

which, in the limit of large size s, simplifies to

$$\text{Prob}[\sigma \geqslant s] \sim s^{-1}. \tag{66}$$

Therefore, based on the results in Ref. [20], the scaling exponent of the exceedance distribution for the in-component sizes in a perfect i-ary tree with some randomness is expected to deviate from 1, depending on the fraction of vertices allowed to branch randomly. In Sec. III D we demonstrated that a plane tree with a power-law exceedance distribution for the subtree sizes exhibits a scaling exponent equal to 1. This suggests that the line graph of a river network closely approximates a perfect i-ary tree. Based on this conjecture, the observed differences in scaling exponents between the two representations of river networks may arise from the distinct organization of their elements. While the $2n-1$ nodes (excluding the root) in a planted tree occupy Euclidean space in a spanning manner, leading to low density near the basin's closure, a river network organizes into an oversaturated hierarchical structure when mapped onto a plane tree. In this structure, the n vertices effectively fill a genuine topological space around the tributary rivers of the entire basin. Additionally, this arrangement of vertices results in a branching ratio significantly greater than 1 near the root node of the plane tree, which is consistent with the assortative character of the river networks analyzed here. As outlined in the previous section, the observed positive correlations between the in-degrees of adjacent vertices—combined with the constraint that the root is the node with the highest number of

incoming edges—indicates that vertices closer to the root generate far more children than they would in an uncorrelated network, where edges are placed randomly, independent of in-degrees.

Finally, it is noteworthy that, in line with the asymptotic power-law in Eq. (66), a perfect i-ary tree inherently lacks a cutoff in the exceedance probability of drainage areas. Thus, in our empirical data, the absence of a large-scale cutoff in the power-law decay of the exceedance distribution for in-component sizes could also be attributed to the resemblance of the line graph of river networks to a perfect i-ary tree.

VI. CONCLUSIONS

Building on the premise that modern graph theory may offer a natural and robust framework for the exact mathematical treatment of natural drainage networks, this paper put forward a proposal for applying the complex network formalism to fluvial systems, thereby extending their analytical scope beyond traditional geomorphological methods. While the mathematical aspects of our study fit comfortably within the context of branching process theory, our approach was based on a model of uncorrelated, directed random graphs with arbitrary node degree distributions, integrating two levels of representation for the same structure. At the primary level, we adopted the planted-tree description of river networks, which provides an immediate view of their planar configuration in the Euclidean plane, where nodes may or may not bifurcate. Then, through a line graph transformation, we introduced a mapping that projects the river network into a purely topological space, where it emerges as a plane tree with a nonhomogeneous connectivity structure. This transformation results in a hierarchy of nested subbasins, offering a clear interpretation of hierarchical relationships as inherently embedded in the topological distances between vertices in the plane tree.

The line graph transformation we introduced naturally induces a physically based selection of fluvial random variables that correspond to well-defined topological elements in the line graph, thereby leading to sample distributions from the entire populations of upstream lengths and drainage areas. Specifically, under our mapping, the probability distributions of tributary river lengths and drainage areas in tributary basins correspond to the in-degree distribution of nodes and the in-component size distribution in the plane tree, respectively. Both distributions were found to decay as power laws without any large-scale cutoff, with the measured scaling exponents being consistent with those observed in various complex networks across different fields—thereby revealing a family of scaling laws in fluvial systems.

Moreover, we observed that the distribution of topological distances in the line graph, which correspond to the distribution of d-order tributary rivers in the planted tree, fits the Gamma distribution very well. This finding uncovers a scaling phenomenon in river networks, where the growth rate of vertices decays according to a power law with the levels of the plane tree, reflecting a hierarchical correlation between the average number of vertices located at consecutive levels. Furthermore, we detected significant positive correlations between the in-degrees of adjacent vertices across all analyzed networks, describing how tributary rivers join in relation to their lengths, consistently with the behavior observed in the width-type function.

On the theoretical side, by applying the generating function formalism to the in-component problem in the limit of large system size, the scaling exponent for the probability distribution of drainage areas in tributary basins was predicted to be exactly 2, independent of the scaling exponent for tributary river lengths. This theoretical prediction closely aligns with the observed scaling exponents for the in-component size distribution across all examined catchments, suggesting that river networks may belong to the same universality class as other real-world complex networks. Additionally, simple branching process arguments indicated that a scaling exponent of 2 for the sizes of subtrees is consistent with a plane tree resulting from a supercritical random process that densely fills topological space.

In summary, the line graph representation of natural drainage networks offers a high level of abstraction that enhances conventional fluvial analyses, providing deeper insights into the understanding of river topology at a fundamental level. This approach has the potential to pave the way for a more general and innovative formulation of geomorphological problems.

ACKNOWLEDGMENTS

This work was partially funded by the Italian Ministry of University and Research through the PON "Ricerca e Innovazione" 2014–2020-FSE- REACT-EU, Action IV.4, "Dottorati e contratti di ricerca su tematiche dell'innovazione." The CINID (Italian Interuniversity Consortium for Hydrology) funded the open access license. The manuscript was polished with assistance from ChatGPT.

[1] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. **74**, 47 (2002); S. N. Dorogovtsev and J. F. F. Mendes, Evolution of networks, Adv. Phys. **51**, 1079 (2002); M. E. J. Newman, The structure and function of complex networks, SIAM Rev. **45**, 167 (2003); *Networks: An Introduction*, 2nd ed. (Oxford University Press, Oxford, 2018).

[2] B. Bollobás, *Random Graphs* (Academic Press, London, 1985).

[3] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Random graphs with arbitrary degree distribution and their applications, Phys. Rev. E **64**, 026118 (2001).

[4] A.-L. Barabási and R. Albert, Emergence of scaling on random networks, Science **286**, 509 (1999); L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Classes of small-world networks, Proc. Natl. Acad. Sci. USA **97**, 11149 (2000); I. Voitalov, P. van der Hoorn, R. van der Hofstad, and D.

Krioukov, Scale-free networks well done, Phys. Rev. Res. **1**, 033034 (2019).

[5] J. G. Forster, D. V. Foster, P. Grassberger, and M. Paczuski, Edge direction and the structure of networks, Proc. Natl. Acad. Sci. USA **107**, 10815 (2010).

[6] S. Maslov and K. Sneppen, Specificity and stability in topology of protein networks, Science **296**, 910 (2002).

[7] S. Maslov, K. Sneppen, and A. Zaliznyak, Detection of topological patterns in complex networks: Correlation profile of the Internet, Physica A **333**, 529 (2004).

[8] R. Pastor-Satorras, R. Vázquez, and A. Vespignani, Dynamical and correlation properties of the Internet, Phys. Rev. Lett. **87**, 258701 (2001).

[9] A. Vazquez, R. Pastor-Satorras, and A. Vespignani, Large-scale topological and dynamical properties of the Internet, Phys. Rev. E **65**, 066130 (2002).

[10] M. E. J. Newman, Assortative mixing in networks, Phys. Rev. Lett. **89**, 208701 (2002).

[11] M. E. J. Newman, Mixing patterns in networks, Phys. Rev. E **67**, 026126 (2003).

[12] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Giant strongly connected component of directed networks, Phys. Rev. E **64**, 025101(R) (2001); N. Schwartz, R. Cohen, D. ben-Avraham, A.-L. Barabási, and S. Havlin, Percolation in directed scale-free networks, *ibid.* **66**, 015104(R) (2002); I. Kryven, Emergence of the giant weak-component in directed random graphs with arbitrary degree distributions, *ibid.* **94**, 012315 (2016).

[13] M. Boguñá and M. A. Serrano, Generalized percolation in random directed networks, Phys. Rev. E **72**, 016106 (2005).

[14] G. Caldarelli, R. Marchetti, and L. Pietronero, The fractal properties of Internet, Eur. Phys. Lett. **52**, 386 (2000).

[15] A. Arenas, L. Danon, A. Díaz-Guilera, P. M. Gleiser, and R. Guimerà, Community analysis in social networks, Eur. Phys. J. B **38**, 373 (2004).

[16] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, Defining and identifying communities in networks, Proc. Natl. Acad. Sci. USA **101**, 2658 (2004).

[17] G. Caldarelli, C. C. Cartozo, P. De Los Rios, and V. D. P. Servedio, Widespread occurrence of the inverse square distribution in social sciences and taxonomy, Phys. Rev. E **69**, 035101(R) (2004).

[18] P. L. Krapivsky and S. Redner, Organization of growing random networks, Phys. Rev. E **63**, 066123 (2001).

[19] P. De Los Rios, Power-law size distribution of supercritical random trees, Europhys. Lett. **56**, 898 (2001).

[20] K. Paik and P. Kumar, Inevitable self-similar topology of binary trees and their diverse hierarchical density, Eur. Phys. J. B **60**, 247 (2007).

[21] Using a growing network model, a similar quantity was first introduced in Ref. [18], where the authors studied a time-dependent directed network. In the context of the present study, this corresponds to a plane tree with the root as the oldest ancestor node. Excluding the root, the authors referred to the fraction of vertices with total degree $k+1$ that attach to vertices with total degree $h+1$, out of a total of $n-1$ nodes. In Ref. [22], this definition was extended to undirected graphs, leading to the so-called matrix of connections, which is symmetric in its indices. The elements of this matrix represent the number of edges connecting a vertex of degree k to one of degree h when $h \neq k$, and twice the number of edges between vertices within the same degree class when $h = k$ (see also Ref. [23]). Finally, a related quantity was introduced in Refs. [10,11], defined in terms of the excess degree of vertices, which is one less than the vertex degree itself.

[22] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. H. Strogatz, Are randomly grown graphs really random? Phys. Rev. E **64**, 041902 (2001).

[23] M. Boguá and R. Pastor-Satorras, Epidemic spreading in correlated complex networks, Phys. Rev. E **66**, 047104 (2002); M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Epidemic spreading in complex networks with degree correlations, in *Statistical Mechanics of Complex Networks*, Lecture Notes in Physics, edited by R. Pastor-Satorras, M. Rubi, and A. Diaz-Guilera (Springer-Verlag, Berlin/Heidelberg, 2003), Vol. 625, pp. 127–147.

[24] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Network robustness and fragility: Percolation on random graphs, Phys. Rev. Lett. **85**, 5468 (2000); R. Albert, H. Jeong, and A.-L. Barabási, Error and attack tolerance of complex networks, Nature (London) **406**, 378 (2000); R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Resilience of the Internet to random breakdowns, Phys. Rev. Lett. **85**, 4626 (2000); Breakdown of the Internet under intentional attack, **86**, 3682 (2001); R. Pastor-Satorras and A. Vespignani, Epidemic spreading in scale-free networks, *ibid.* **86**, 3200 (2001); Epidemic dynamics and endemic states in complex networks, Phys. Rev. E **63**, 066117 (2001); M. E. J. Newman, Spread of epidemic disease on networks, *ibid.* **66**, 016128 (2002); L. K. Gallos, R. Cohen, P. Argyrakis, A. Bunde, and S. Havlin, Stability and topology of scale-free networks under attack and defense strategies, Phys. Rev. Lett. **94**, 188701 (2005).

[25] V. M. Eguíluz and K. Klemm, Epidemic threshold in structured scale-free networks, Phys. Rev. Lett. **89**, 108701 (2002); A. Vázquez and Y. Moreno, Resilience to damage of graphs with degree correlations, Phys. Rev. E **67**, 015101(R) (2003).

[26] A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes, Percolation on correlated networks, Phys. Rev. E **78**, 051105 (2008).

[27] C. M. Schneider, A. A. Moreira, J. S. Andrade, S. Havlin, and H. J. Herrmann, Mitigation of malicious attacks on networks, Proc. Natl. Acad. Sci. USA **108**, 3838 (2011); Tanizawa, S. Havlin, and H. E. Stanley, Robustness of onionlike correlated networks against targeted attacks, Phys. Rev. E **85**, 046109 (2012).

[28] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Critical phenomena in complex networks, Rev. Mod. Phys. **80**, 1275 (2008).

[29] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. **87**, 925 (2015).

[30] R. P. Stanley, *Enumerative Combinatorics*, Vol. 2 (Cambridge University Press, Cambridge, 1999).

[31] Upon removal of its root node, a planted binary tree reduces to a binary tree.

[32] A. Rinaldo, J. E. Banavar, and A. Maritan, Tree, networks, and hydrology, Water Resour. Res. **42**, W06D07 (2006).

[33] I. Rodríguez-Iturbe, E. Ijjász-Vasquez, R. L. Bras, and D. G. Tarboton, Power law distribution of discharge mass and energy in river basins, Water Resour. Res. **28**, 1089 (1992).

[34] A. Maritan, A. Rinaldo, R. Rigon, A. Giacometti, and I. Rodríguez-Iturbe, Scaling laws for river networks, Phys. Rev. E **53**, 1510 (1996).

[35] A. Rinaldo, R. Rigon, J. R. Banavar, A. Maritan, and I. Rodríguez-Iturbe, Evolution and selection of river networks: Statistics, dynamics, and complexity, Proc. Natl. Acad. Sci. USA **111**, 2417 (2014).

[36] R. Rigon, I. Rodríguez-Iturbe, A. Maritan, A. Giacometti, D. G. Tarboton, and A. Rinaldo, On Hack's law, Water Resour. Res. **32**, 3367 (1996).

[37] J. T. Hack, Studies of longitudinal stream profiles in Virginia and Maryland, U.S. Geol. Surv. Prof. Pap. No. 294-B (USGS Publications Warehouse, 1957), p. 45.

[38] I. Rodríguez-Iturbe and A. Rinaldo, *Fractal River Basins: Change and Self-Organization* (Cambridge University Press, Cambridge, 1997).

[39] Originally discovered by Hack trough analysis of distinct, nonoverlapping basins—specifically referring to the length of their longest stream—this relationship is now also recognized as valid with respect to main stream lengths, as well as for subbasins embedded within a larger catchment, regardless of whether they are nested or not [34,36,40].

[40] P. S. Dodds and D. H. Rothman, Unified view of scaling laws for river networks, Phys. Rev. E **59**, 4865 (1999).

[41] D. M. Gray, Interrelationships of watershed characteristics, J. Geophys. Res. **66**, 1215 (1961).

[42] G. Willgoose, R. L. Bras, and I. Rodríguez-Iturbe, A coupled channel network growth and hillslope evolution model: 1. Theory, Water Resour. Res. **27**, 1671 (1991); A. D. Howard, A detachment-limited model of drainage basin evolution, *ibid.* **30**, 2261 (1994); G. E. Tucker and R. L. Bras, Hillslope processes, drainage density, and landscape morphology, *ibid.* **34**, 2751 (1998); S. Kramer and M. Marder, Evolution of river networks, Phys. Rev. Lett. **68**, 205 (1992); R. L. Leheny and S. R. Nager, Model for the evolution of river networks, *ibid.* **71**, 1470 (1993); J. R. Banavar, F. Colaiori, A. Flammini, A. Giacometti, A. Maritan, and A. Rinaldo, Sculpting of a fractal river basin, *ibid.* **78**, 4522 (1997); E. Somfai and L. M. Sander, Scaling and river networks: A Landau theory for erosion, Phys. Rev. E **56**, R5(R) (1997).

[43] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo, Scaling, optimality and landscape evolution, J. Stat. Phys. **104**, 1 (2001).

[44] I. Rodríguez-Iturbe, A. Rinaldo, R. Rigon, R. L. Bras, and E. Ijjasz-Vasquez, Energy dissipation, runoff production and the three-dimensional structure of channel networks, Water Resour. Res. **28**, 1095 (1992); A. Rinaldo, I. Rodríguez-Iturbe, R. Rigon, R. L. Bras, E. Ijjasz-Vasquez, and A. Marani, Minimum energy and fractal structures of drainage networks, *ibid.* **28**, 2183 (1992).

[45] R. Rigon, A. Rinaldo, I. Rodríguez-Iturbe, R. L. Bras, and E. Ijjasz-Vasquez, Optimal channel networks: A framework for the study of river basin morphology, Water Resour. Res. **29**, 1635 (1993).

[46] R. E. Horton, Erosional development of streams and their drainage basins hydrophysical approach to quantitative morphology, Bull. Geol. Soc. Am. **56**, 275 (1945); A. N. Strahler, Quantitative analysis of watershed geomorphology, Eos Transactions American Geophysical Union **38**, 913 (1957).

[47] M. A. Melton, A derivation of Strahler's channel-ordering system, J. Geol. **67**, 345 (1959).

[48] R. L. Shreve, Statistical law of stream number, J. Geol. **74**, 17 (1966); Infinite topologically random channel networks, **75**, 178 (1967).

[49] B. M. Troutman and M. B. Karlinger, On the expected width function for topologically random channel networks, J. Appl. Prob. **21**, 836 (1984).

[50] O. J. Mesa and V. K. Gupta, On the main channel length-area relationship for channel networks, Water Resour. Res. **23**, 2119 (1987); M. Fiorentino, P. Claps, and V. P. Singh, An entropy-based morphological analysis of river basin networks, *ibid.* **29**, 1215 (1993); P. Claps, M. Fiorentino, and G. Oliveto, Informational entropy of fractal river networks, J. Hydrol. **187**, 145 (1996).

[51] A. Marani, R. Rigon, and A. Rinaldo, A note on fractal channel networks, Water Resour. Res. **27**, 3041 (1991).

[52] P. S. Dodds and D. H. Rothman, Geometry of river networks. I. Scaling, fluctuations and deviations, Phys. Rev. E **63**, 016115 (2000); Geometry of river networks. II. Distributions of component size and number, **63**, 016116 (2000); Geometry of river networks. III. Characterization of component connectivity, **63**, 016117 (2000).

[53] F. Harray, *Graph Theory* (Addison-Wesley, Boston, 1969).

[54] M. Molloy and B. Reed, A critical point for random graphs with a given degree sequence, Random Struc. Algor. **6**, 161 (1995); The size of the giant component of a random graph with a given degree sequence, Comb. Probab. Comput. **7**, 295 (1998).

[55] P. Erdős and A. Rényi, On random graphs, Publ. Math. **6**, 290 (1959); On the evolution of random graphs, Publ. Math. Inst. Hungar. Acad. Sci. **5**, 17 (1960); On the strength of connectedness of a random graph, Acta Math. Scientia Hungary **12**, 261 (1964).

[56] H. Gravelius, *Morphometry of Drainage Basins* (Elsevier, Amsterdam, 1914).

[57] T. E. Harris, *The Theory of Branching Processes* (Springer-Verlag, Berlin, 1963).

[58] M. Marani, A. Rinaldo, R. Rigon, and I. Rodriguez-Iturbe, Geomorphological width functions and the random cascade, Geophys. Res. Lett. **21**, 2123 (1994).

[59] B. M. Troutman and M. R. Karlinger, Unit hydrograph approximations assuming linear flow through topologically random channel networks, Water Resour. Res. **21**, 743 (1985); M. R. Karlinger and B. M. Troutman, Assessment of the instantaneous unit hydrograph derived from the theory of topologically random networks, *ibid.* **21**, 1693 (1985); V. K. Gupta and O. Mesa, Runoff generation and hydrologic response via channel network geomorphology—Recent progress and open problems, J. Hydrol. **102**, 3 (1988).

[60] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, Self-similar community structure in a network of human interactions, Phys. Rev. E **68**, 065103(R) (2003).

[61] B. Jiang and C. Claramunt, Topological analysis of urban street networks, Environ. Plan. B **31**, 151 (2004); S. Porta, P. Crucitti, and V. Latora, The network analysis of urban streets: A dual approach, Physica A **369**, 853 (2006); P. Crucitti, V. Latora, and S. Porta, Centrality in networks of urban street, Chaos **16**, 015113 (2006).

[62] The digital terrain model is available at http://rsdi.regione.basilicata.it/Catalogo/srv/ita/search?hl=ita/#_.

[63] M. Neteler and H. Mitasova, *Open Source GIS: A GRASS GIS Approach* (Springer, New York, 2008).

[64] M. E. J. Newman, Component sizes in networks with arbitrary degree distributions, Phys. Rev. E **76**, 045101(R) (2007).

[65] This explanation was first proposed in Ref. [20], where the authors noted that the sharpest cutoff occurs in a critical binary tree, in which only one node bifurcates at each topological level. In this case, the repetition of a single internal node and one leaf results in a constant increase in the subtree size.

[66] J. E. Nash, A unit hydrograph study, with particular reference to British catchments, Proc. Inst. Civ. Eng. **17**, 249 (1960).

[67] Equations (33) and (34) were derived in Ref. [3] under the key assumption that the network is locally tree-like; that is, all finite connected clusters of vertices contain no short loops of edges. By definition, this condition is trivially satisfied by any tree.

[68] A perfect i-ary tree is a specific type of plane tree where all leaves are situated at the same level, and every internal node has exactly i children.

[69] J. S. Kim, K.-I. Goh, G. Salvi, E. Oh, B. Kahng, and D. Kim, Fractality in complex networks: Critical and supercritical skeletons, Phys. Rev. E **75**, 016110 (2007).

[70] More precisely, the average branching ratio was defined in Ref. [69] as the mean in-degree of the vertices. However, for a constant branching ratio, it can be easily shown that $\langle i\rangle_p = \langle i\rangle_q$, where both averages exclude the leaf nodes at the maximum level. In this paper, to align with the standard formalism from graph theory used in Ref. [3], we define the average branching ratio in terms of the in-degree distribution of the source nodes of an edge, rather than the in-degree distribution of the vertices themselves.

[71] I. Kryven, General expression for the component size distribution in infinite configuration networks, Phys. Rev. E **95**, 052303 (2017).

[72] M. E. J. Newman, Power laws, Pareto distributions and Zipf's law, Contemp. Phys. **46**, 323 (2005).

[73] The Pearson coefficient introduced in Ref. [5] generalizes the version originally proposed in Ref. [22] for undirected networks and later modified in Ref. [10] in terms of the joint distribution of the excess degrees of the two vertices at either end of a randomly chosen edge. The first extension to directed graphs was presented in Ref. [11], where the correlation measure was defined with respect to the excess out- and in-degrees of the source and target nodes, respectively, for a given edge.

[74] M. E. J. Newman and J. Park, Why social networks are different from other types of networks, Phys. Rev. E **68**, 036122 (2003).

[75] A similar quantity was introduced in Ref. [13] in terms of the out- and in-degrees of the source and target nodes of a given edge, respectively. Referring to the notation used in the present paper, Eq. (4) in Ref. [13] can be written as $\hat{r}_k r_{h|k} = r_h \hat{r}_{k|h}$. In contrast, in our case, the detailed balance condition gives $q_k r_{h|k} = r_h q_{k|h}$. The analogous quantity for undirected graphs was discussed in Refs. [8,9,25,26], and [23].

[76] A similar result was obtained for random branching trees in Ref. [14].

[77] A perfect 2-ary tree (referred to as a perfect binary tree in Ref. [20]) is a binary tree in which all leaves are situated at the same level.