complexities

MDPI

Article

A Multi-Criteria Framework for Comparative Topological Regime Characterization in Complex Networks

Fabiane de Fatima Carvalho [1,*], Ivan Bergier [1], Silvia Maria Fonseca Silveira Massruhá [2] and Jayme Garcia Arnal Barbedo [1]

[1] Embrapa Digital Agriculture, Embrapa, 209 André Tosello Avenue, Campinas 13083-886, SP, Brazil; ivan.bergier@embrapa.br (I.B.); jayme.barbedo@embrapa.br (J.G.A.B.)
[2] Embrapa, W3 Norte Avenue-Asa Norte, Brasilia 70770-901, DF, Brazil; silvia.massruha@embrapa.br
* Correspondence: fabicfc90@gmail.com

Abstract

Collaboration networks are frequently studied as empirical instances of complex social systems, yet standardized methodological frameworks for consistently identifying heterogeneous mesoscopic structural regimes remain limited. This study proposes an integrated multi-criteria classification framework and demonstrates its application to the structural characterization of communities extracted from a large-scale scientific collaboration network. The framework combines community detection, classical network metrics, statistical modeling of weighted degree tails, small-world diagnostics, information-entropy measures, and fractal analysis based on the Song–Havlin–Makse box-covering renormalization framework. As an empirical application, the methodology is applied to the giant coauthorship component of Embrapa's scientific production (1974–2024), derived from the Brazilian Agricultural Research Database (BDPA), comprising 60,636 nodes. The weighted Louvain algorithm partitions the network into 25 major communities, which are evaluated through an integrated classification protocol combining the Akaike Information Criterion model selection, Kolmogorov–Smirnov goodness-of-fit tests, small-worldness diagnostics, and fractal scaling analysis. The proposed framework identifies three network families, namely Barabási–Albert (BA-like)/scale-free small-world, scale-free fractal (non-BA) and small-world (non-scale-free), while explicitly distinguishing supported and ambiguous classifications according to the overall consistency of the statistical and structural evidence. The results demonstrate that distinct mesoscopic structural regimes coexist within the same connected collaboration system, highlighting the usefulness of the proposed reproducible multi-criteria framework for comparative topological characterization across complex collaboration networks.

Keywords: scale-free small-world networks; small-world (non-scale-free) networks; scale-free fractal networks

Check for updates

Academic Editors: Alexandre G. Evsukoff and Yilun Shang

Received: 28 May 2026
Revised: 31 July 2026
Accepted: 25 August 2026
Published: 11 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Scientific collaboration networks constitute an important framework for understanding organization, knowledge diffusion, and systemic resilience in science [1–3]. Empirical studies consistently report heavy-tailed degree distributions, small-world organization, and degree assortativity in such systems [4,5]. Despite the abundance of descriptive analyses, however, methodological consistency in identifying and comparatively characterizing mesoscopic structural regimes across subnetworks remains limited. In particular, the relationship between small-world organization and fractal self-similarity remains debated, both

conceptually and in terms of the statistical criteria used for their identification [6,7]. Recent analyses of the Brazilian Agricultural Research Corporation (Embrapa) reinforce these structural patterns in an applied institutional context [8]; however, a systematic, reproducible, and integrated multi-criteria framework for comparative subnetwork characterization is still lacking.

Community detection has been widely investigated in complex networks, including unipartite and bipartite representations, with approaches ranging from direct algorithms for bipartite graphs [9] to modularity-based methods applied to projected networks. While these techniques effectively identify mesoscopic partitions, less attention has been devoted to the structural characterization of the resulting communities. In many empirical studies, communities are reported primarily as partitions, without evaluating whether they correspond to distinct topological organizations or structural regimes.

Complex network theory further suggests that large-scale coordination and robustness may emerge from repeated local interactions among agents. Iterative decisions involving cooperation, imitation, brokerage, and bounded rational search can generate recurrent structural patterns such as heavy-tailed weighted degree distributions, clustered organization, shortcut-induced small-world structure, or disassortative mixing, even in the absence of centralized planning. These recurrent structural patterns have been widely discussed in the literature as possible outcomes of decentralized interaction dynamics in social and rural institutional systems [10–12]. However, relatively few studies translate these theoretical insights into explicit statistical criteria capable of distinguishing structural regimes within a single giant connected component.

From a network perspective, the emergence of a giant connected component can be interpreted as a phase-transition phenomenon in which link density exceeds a critical threshold. Importantly, such components are rarely homogeneous: multiple structural regimes may coexist within the same large-scale connected structure, reflecting thematic, institutional, spatial, or incentive-driven constraints. Evidence of this coexistence, including hub-dominated regions, clustered small-world areas, and sparsely connected fractal-like zones, has been reported in cooperative and organizational knowledge networks [12–14]. Nevertheless, the absence of standardized analytical pipelines limits cross-study comparability and weakens reproducibility in comparative structural characterization.

This gap motivates the present study. The central research question addressed here is whether distinct and statistically identifiable structural regimes can coexist within a single giant connected component of a scientific collaboration network and how such regimes can be comparatively characterized using reproducible statistical criteria. To answer this question, we develop and demonstrate a statistically grounded and algorithmically explicit framework for comparative cluster-level structural characterization. Rather than treating community detection as an end in itself, we use it as an intermediate step in an integrated multi-criteria classification framework combining weighted degree-tail modeling, information-entropy measures of degree–degree correlations, small-world diagnostics, and fractal analysis based on the Song–Havlin–Makse box-covering renormalization framework. The complementary statistical and structural diagnostics are integrated through an explicit classification protocol that assigns network families while distinguishing supported and ambiguous classifications according to the overall consistency of the available evidence. The resulting categories should therefore be interpreted as comparative topological descriptions derived from statistical diagnostics, rather than definitive evidence of specific underlying generative mechanisms.

Applying this framework to the giant coauthorship component of Embrapa's scientific production derived from the BDPA, we identify the coexistence of distinct network families, including BA-like/scale-free small-world, scale-free fractal (non-BA) and small-world

(non-scale-free) organizations within the same connected collaboration system. Rather than assigning deterministic labels to individual communities, the proposed framework integrates complementary statistical and structural evidence to identify network families while explicitly representing the confidence associated with each classification. These findings indicate that large scientific collaboration networks may display substantial mesoscopic structural heterogeneity even after percolation, challenging interpretations based solely on global averages. By providing explicit decision rules and a reproducible integrated classification protocol, this study contributes a transferable methodological framework for the comparative topological characterization of heterogeneous collaboration networks.

2. Materials and Methods

The integrated multi-criteria classification framework adopted in this study follows a structured procedure designed to characterize the mesoscopic topological organization of empirical subnetworks in a statistically robust and reproducible manner. The framework consists of the following sequential stages:

First, a weighted coauthorship network is constructed from bibliographic metadata, and its giant connected component is extracted to ensure global connectivity and statistical stability.

Second, the weighted network is partitioned into communities using the Louvain community detection algorithm, allowing the analysis to focus on mesoscopic subnetworks that may exhibit distinct structural organizations.

Third, classical weighted structural metrics are computed for each community, including density, clustering coefficient, assortativity, average shortest-path length, and complementary topological descriptors. These metrics provide baseline information on connectivity, hierarchy, and local cohesion.

Fourth, small-world behavior is assessed by comparing the empirical clustering coefficient and average path length with Erdős–Rényi random graph baselines, yielding the small-worldness index σ, which quantifies the coexistence of strong local clustering and efficient global connectivity.

Fifth, the weighted degree-tail distribution of each community is analyzed through maximum-likelihood estimation. The lower cutoff $x_{\min}$ is determined by Kolmogorov–Smirnov minimization, goodness-of-fit is evaluated by bootstrap resampling, and competing power-law, exponential, and log-normal models are compared using Akaike's Information Criterion (AIC).

Sixth, fractal scaling is evaluated using the Song–Havlin–Makse box-covering renormalization framework. The relationship between the number of covering boxes and box diameter is modeled and compared using AIC-based model selection to determine whether the observed scaling is more consistent with a power-law model or an exponential decay.

Seventh, information-entropy measures of degree–degree correlations, including mutual information and conditional entropy, are computed to complement the structural characterization by quantifying the statistical dependence between connected node degrees.

Finally, all statistical and structural diagnostics are integrated through an explicit integrated classification protocol. Rather than relying on a single metric, the protocol combines complementary evidence from weighted degree-tail modeling, small-world diagnostics, fractal analysis, and information-theoretic measures to assign each community to a network family (BA-like/scale-free small-world, scale-free fractal (non-BA) and small-world (non-scale-free), or undefined) while explicitly distinguishing supported and ambiguous classifications according to the overall consistency of the available evidence.

This integrated framework ensures that the final classification emerges from the convergence of multiple complementary statistical and structural diagnostics, providing

a reproducible and transferable methodology for the comparative characterization of heterogeneous mesoscopic regimes in complex collaboration networks.

2.1. Network and Giant Component Construction

The empirical dataset was obtained from the Brazilian Agricultural Research Database (BDPA), which indexes the scientific production of Embrapa researchers. Bibliographic metadata were retrieved directly from BDPA through its native JSON export interface and subsequently processed using custom Python routines in Python 3.12.3, using the NumPy 1.26.4, Pandas 2.1.4, NetworkX 2.8.8, and Matplotlib 3.6.3 libraries. The exported records include publication identifiers, publication year, author lists, institutional affiliations, and additional bibliographic metadata. Only peer-reviewed journal articles published between 1974 and 2024 were considered in this study.

Prior to network construction, duplicate publication records were removed, and author names were normalized to reduce textual inconsistencies arising from spelling variations, abbreviations, and formatting differences. Since BDPA does not provide persistent author identifiers (e.g., ORCID) for all records, a rule-based normalization and heuristic canonicalization procedure was applied before constructing the collaboration network. The procedure standardized trivial orthographic variations by removing leading Portuguese prepositions, eliminating unnecessary whitespace, and harmonizing common suffix abbreviations (e.g., jr" to junior"). Candidate author variants were subsequently identified using a conservative rule-based heuristic to reduce textual inconsistencies while avoiding unnecessary merges whenever potential ambiguities were detected. The resulting canonical author table was further subjected to quality-control procedures, including the detection of malformed or truncated names, consistency checks between intermediate network representations, and verification that isolated authors were preserved during network construction.

The coauthorship network was modeled as an undirected weighted graph, in which each node represents an author. For every publication containing multiple authors, all possible pairs of coauthors were connected, generating a complete clique associated with that publication. Whenever the same pair of authors collaborated in multiple publications, the corresponding edge weight was incremented according to the cumulative number of joint publications throughout the observation period. Thus, edge weights quantify the collaboration intensity between pairs of researchers.

After preprocessing, the largest connected component (giant component), containing 60,636 authors, was extracted using the connected-components algorithm implemented in the NetworkX library. Restricting the analysis to the giant component ensures that all subsequent structural analyses are performed on a connected network, avoiding artifacts associated with isolated nodes and disconnected subgraphs.

Communities were identified using the Louvain algorithm [15], executed on the weighted network using collaboration frequencies as edge weights. The resulting partition contains 25 communities. The Louvain algorithm identifies communities by maximizing the weighted network modularity,

$$Q = \frac{1}{2W} \sum_{i,j} \left(w_{ij} - \frac{s_i s_j}{2W} \right) \delta(c_i, c_j), \tag{1}$$

where w_{ij} denotes the weight of the edge connecting nodes i and j, $s_i = \sum_j w_{ij}$ is the weighted degree (node strength) of node i, $W = \frac{1}{2} \sum_{i,j} w_{ij}$ is the total weight of the network, c_i and c_j denote the communities to which nodes i and j belong, and $\delta(c_i, c_j)$ is the Kronecker delta function, which equals 1 when both nodes belong to the same community and 0 otherwise. High values of Q indicate a strong community structure, where the observed

collaboration intensity within communities is significantly higher than expected under the weighted null model.

Throughout this study, the term *degree* refers to the weighted degree (node strength), defined as the sum of the weights of all edges incident on a node, unless otherwise stated. Consequently, all degree-distribution analyses were performed using the weighted degree. Community detection was also carried out on the weighted network, whereas structural descriptors based primarily on network topology, including the clustering coefficient, assortativity, and average shortest-path length, were computed using their standard graph-theoretic definitions.

2.2. Per-Cluster Structural Metrics

For each cluster we compute several structural quantities that capture its internal connectivity patterns:

- Density quantifies how many edges exist relative to the maximum possible number of edges in a fully connected graph,

$$\rho = \frac{2m}{n(n-1)}, \tag{2}$$

  where m is the number of edges and n is the number of nodes in the cluster. Values of ρ close to 1 indicate highly connected subgraphs.
- Average clustering coefficient [4] measures the tendency of nodes to form triangles,

$$C = \frac{1}{n}\sum_i \frac{2e_i}{k_i(k_i-1)}, \tag{3}$$

  where e_i is the number of edges among the neighbors of node i, and k_i is the degree of node i. The term $\frac{2e_i}{k_i(k_i-1)}$ represents the local clustering of node i, while C is the average over all nodes.
- Degree assortativity [5] captures degree correlations between connected nodes,

$$r = \frac{\sum_{jk} jk\,(e_{jk} - q_j q_k)}{\sigma_q^2}, \tag{4}$$

  where e_{jk} is the joint probability that an edge connects nodes of degrees j and k, q_k is the degree distribution of the nodes at the ends of edges, and σ_q^2 is the variance of q_k. Positive r indicates that nodes tend to connect with others of similar degree (assortative mixing), while negative r reflects disassortative networks.
- Average path length represents the typical separation between pairs of nodes,

$$L = \frac{1}{n(n-1)}\sum_{i\neq j} d(i,j), \tag{5}$$

  where $d(i,j)$ is the length of the shortest path between nodes i and j. For large clusters, L is estimated by random sampling of node pairs.

To evaluate whether clusters exhibit small-world properties, we compare the empirical clustering coefficient and path length to those of equivalent Erdős–Rényi random graphs (C_{rand} and L_{rand}), and compute the small-worldness index [16]:

$$\sigma = \frac{C/C_{\text{rand}}}{L/L_{\text{rand}}}, \tag{6}$$

where values of $\sigma > 1$ indicate small-world behavior, combining high local clustering with short global distances.

2.3. Degree-Tail Modeling, $x_{\min}$ Determination, and Model Selection

We assess whether cluster-level degree distributions exhibit power-law tails by combining maximum-likelihood estimation (MLE), Kolmogorov–Smirnov (KS) minimization to select the lower cutoff $x_{\min}$, a bootstrap p-value for goodness-of-fit, and Akaike's Information Criterion (AIC) for model comparison, following [17–19].

Data and notation: Let $\{k_i\}_{i=1}^{N}$ be the observed degrees in a given cluster. For a candidate lower cutoff $x_{\min}$, define the tail sample $\{k_i : k_i \geq x_{\min}\}$ of size n.

2.3.1. Power-Law MLE in the Tail

For discrete degrees, the power-law probability mass above $x_{\min}$ is

$$p(k \mid \alpha, x_{\min}) = \frac{k^{-\alpha}}{\zeta(\alpha, x_{\min})}, \qquad k \in \{x_{\min}, x_{\min}+1, \dots\}, \tag{7}$$

where $\zeta(\alpha, x_{\min})$ is the Hurwitz zeta function ensuring normalization. The MLE for the exponent, using the standard 0.5 continuity correction, is

$$\hat{\alpha}(x_{\min}) \;=\; 1 + n\left[\sum_{i=1}^{n} \ln\!\left(\frac{k_i - 0.5}{x_{\min} - 0.5}\right)\right]^{-1}, \tag{8}$$

here, $n = |\{i : k_i \geq x_{\min}\}|$; k_i denotes observed degrees in the tail.

2.3.2. Empirical Versus Theoretical CCDF and the KS Distance

Define the empirical complementary CDF (CCDF) on the tail as

$$S(k) \;=\; \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(k_i \geq k), \tag{9}$$

and the fitted power-law CCDF (continuous approximation) as

$$P(k \mid \hat{\alpha}, x_{\min}) \;=\; \left(\frac{k}{x_{\min}}\right)^{1-\hat{\alpha}}, \qquad k \geq x_{\min}. \tag{10}$$

The Kolmogorov–Smirnov statistic is the maximum vertical gap between these CCDFs:

$$D(x_{\min}) \;=\; \max_{k \geq x_{\min}} \left| S(k) - P(k \mid \hat{\alpha}(x_{\min}), x_{\min}) \right|. \tag{11}$$

We choose the cutoff that minimizes this discrepancy:

$$x^{*}_{\min} \;=\; \arg\min_{x_{\min} \in \mathcal{X}} D(x_{\min}), \tag{12}$$

where $\mathcal{X}$ is the set of unique observed degrees.

2.3.3. Goodness-of-Fit via Bootstrap p-Value

To test the plausibility of the fitted tail, we generate B synthetic tails from $p(k \mid \hat{\alpha}(x^{*}_{\min}), x^{*}_{\min})$, each of size n, and compute their KS distances $D^{(b)}$. The bootstrap p-value is

$$p \;=\; \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\Big(D^{(b)} \geq D_{\text{obs}}\Big), \tag{13}$$

with $D_{\text{obs}} = D(x^*_{\min})$. Large p (e.g., $p > 0.1$) indicates that the observed discrepancy is compatible with sampling noise under the fitted model.

2.3.4. Model Selection with AIC

We compare, on the same empirical tail, three candidate models: power-law (PL), exponential (EXP), and left-truncated log-normal (LN). Let $\hat{L}_j$ be the maximized likelihood and k_j the number of free parameters of model $j \in \{\text{PL}, \text{EXP}, \text{LN}\}$. The Akaike Information Criterion is

$$\text{AIC}_j = 2k_j - 2\ln \hat{L}_j, \tag{14}$$

and the preferred tail model is the one with the smallest AIC. Relative support is evaluated by

$$\Delta_j = \text{AIC}_j - \min_{\ell} \text{AIC}_\ell. \tag{15}$$

Following Burnham and Anderson [19], $\Delta_j < 2$ indicates that competing models receive comparable statistical support, $4 \leq \Delta_j \leq 7$ indicates considerably less support, and $\Delta_j > 10$ provides strong evidence against the model relative to the best candidate.

In the present classification framework, however, clusters with $\Delta_j < 2$ are not considered to have a uniquely supported tail model. Instead, these cases are treated as statistically ambiguous and require complementary evidence from the remaining structural diagnostics before a final topological assignment is made.

For each cluster we report $(x^*_{\min}, n_{\text{tail}}, \hat{\alpha})$, KS (D, p), AIC for the PL, EXP, and LN models, together with the selected best tail. The bootstrap KS test evaluates the absolute plausibility of the fitted distribution, whereas AIC quantifies the relative statistical support among competing models.

2.4. Fractality via Box Covering

To evaluate whether the collaboration clusters exhibit self-similar organization, we performed a box-covering analysis following the renormalization framework introduced by Song, Havlin, and Makse [6] and further developed by Rozenfeld et al. [7]. Let LCC denote the largest connected component of a cluster. For a given box diameter ℓ_B, let $N_B(\ell_B)$ denote the number of covering boxes required to cover all vertices of the LCC, where each box contains only vertices whose pairwise shortest-path distance does not exceed ℓ_B.

Fractal networks are expected to satisfy the scaling relation

$$N_B(\ell_B) \sim \ell_B^{-d_B}, \tag{16}$$

where d_B denotes the box (fractal) dimension. In contrast, hub-dominated small-world networks exhibit an exponential decay,

$$N_B(\ell_B) \sim e^{-a\ell_B}, \tag{17}$$

reflecting the rapid collapse of distances during successive renormalization steps [6,7].

2.4.1. Operational Definition and Algorithmic Implementation

Computing the exact minimum box covering is an NP-hard problem [6,7]. We therefore implemented a deterministic greedy box-covering heuristic inspired by the original renormalization framework.

For each covering radius r_B, the vertices of the largest connected component are first sorted in decreasing order of degree. The highest-degree uncovered vertex is then selected as the center of a new box. All vertices located within graph distance r_B from this center are identified by a breadth-first search (BFS) and assigned to the same box. Covered vertices

are removed from the remaining search space, and the procedure is repeated until every vertex has been assigned to a box.

The corresponding box diameter is defined as

$$\ell_B = 2r_B + 1, \tag{18}$$

ensuring that any two vertices belonging to the same box remain separated by a shortest-path distance not exceeding ℓ_B.

The procedure is repeated for $r_B = 1, \ldots, 8$, corresponding to box diameters $\ell_B = 3, \ldots, 17$. For each radius, the resulting number of covering boxes $N_B(\ell_B)$ is recorded.

To reduce finite-size effects associated with the trivial regime where the entire network is covered by only one or two boxes, only observations satisfying $N_B > 2$ are retained. The largest contiguous scaling interval satisfying this condition is subsequently used for the regression analysis.

2.4.2. Model Fitting and Notation

Although both the degree-tail analysis (Section 2.3) and the present box-covering analysis employ the Akaike Information Criterion (AIC), they address different statistical questions. In Section 2.3, the AIC compares alternative probability distributions describing the weighted degree tail (power-law, exponential, and log-normal). Here, the AIC compares two competing scaling relationships between the box diameter (ℓ_B) and the number of covering boxes (N_B): a power-law model, expected for fractal networks, and an exponential model, expected for non-fractal small-world networks.

The retained observations are fitted using two competing regression models,

$$\text{PL:} \quad \log N_B = \beta_0 - d_B \log \ell_B, \tag{19}$$

$$\text{EXP:} \quad \log N_B = \gamma_0 - a\ell_B. \tag{20}$$

The power-law model is fitted in log–log coordinates, whereas the exponential model is fitted in semi-log coordinates.

For the power-law regression, we compute the coefficient of determination,

$$R^2_{\log-\log},$$

which quantifies the quality of the fractal scaling.

Model comparison is performed using the Akaike Information Criterion,

$$AIC = n \ln\left(\frac{RSS}{n}\right) + 2k,$$

where RSS is the residual sum of squares, n is the number of fitted observations, and $k = 2$ is the number of fitted parameters.

The quantity

$$\Delta AIC = AIC_{\text{EXP}} - AIC_{\text{PL}}$$

is then used to compare the competing scaling models. Positive values indicate better support for the power-law model (fractal organization), whereas negative values favor the exponential model (small-world organization).

Following the conventional interpretation of AIC differences, values satisfying $|\Delta AIC| \leq 2$ indicate comparable support for the competing models, values between 4 and 7 indicate moderate evidence, and $|\Delta AIC| \geq 10$ indicate strong evidence for the model with the lower AIC.

2.4.3. Decision Criteria

The box-covering analysis provides one component of the integrated multi-criteria classification framework proposed in this work. Fractal scaling is considered to be supported when the power-law scaling model is preferred over the exponential alternative ($\Delta AIC > 2$) and the corresponding log–log regression exhibits high goodness of fit. Conversely, exponential scaling ($\Delta AIC < -2$) supports a non-fractal small-world organization. Cases exhibiting comparable statistical support ($|\Delta AIC| \leq 2$) or poor regression quality are regarded as inconclusive and are subsequently resolved by the integrated classification protocol described in Section 2.5.

From the perspective of network renormalization, each covering box is replaced by a super-node, generating a coarse-grained representation of the original network. Fractal networks preserve their heterogeneous organization under successive renormalization steps, producing the power-law scaling described by Equation (16). In contrast, hub-dominated small-world networks rapidly lose structural heterogeneity as hubs are merged, resulting in the exponential behavior described by Equation (17) [6,7]. Consequently, the box-covering analysis provides an independent structural characterization that complements the degree-tail modeling and the small-world diagnostics within the integrated classification framework.

2.5. Integrated Classification Protocol

To reflect the overall consistency of the evidence, each network-type assignment was associated with one of two qualitative confidence levels:

-Supported: the statistical and structural diagnostics consistently converged toward the same topological regime.

-Ambiguous: competing statistical models exhibit comparable support (e.g., ΔAIC < 2), the bootstrap KS test does not provide sufficient evidence for a unique tail model, or different structural diagnostics lead to conflicting conclusions.

Here, the term "BA-like" is used exclusively as a phenomenological topological descriptor derived from the combined statistical and structural evidence rather than as proof of a specific network generative mechanism.

This integrative protocol ensures that the final classification is based on the consistent agreement among multiple independent statistical and structural diagnostics rather than on any single network metric. The corresponding decision workflow is summarized schematically in Figure 1.

In the Supplementary Material File, we provide a validation of the framework presented in Figure 1 against synthetic, benchmark networks with known structures. The pipeline was tested on canonical Barabási–Albert, Watts–Strogatz, and fractal/hierarchical networks; the classification accuracy, robustness, and sensitivity to parameter choices are reported.

2.6. Temporal Growth Analysis

To complement the structural characterization of the collaboration network, we estimated the temporal growth exponent β originally introduced by Barabási and Albert [20]. Under the preferential-attachment framework, the cumulative degree of node i evolves approximately as

$$k_i(t) \propto \left(\frac{t}{t_i}\right)^{\beta}, \tag{21}$$

where $k_i(t)$ is the cumulative weighted degree of author i at time t, and t_i denotes the year in which the author first appeared in the coauthorship network. Under the classical Barabási–

Albert model, the expected value is $\beta = 0.5$, whereas deviations from this prediction may reflect departures from idealized linear preferential attachment.

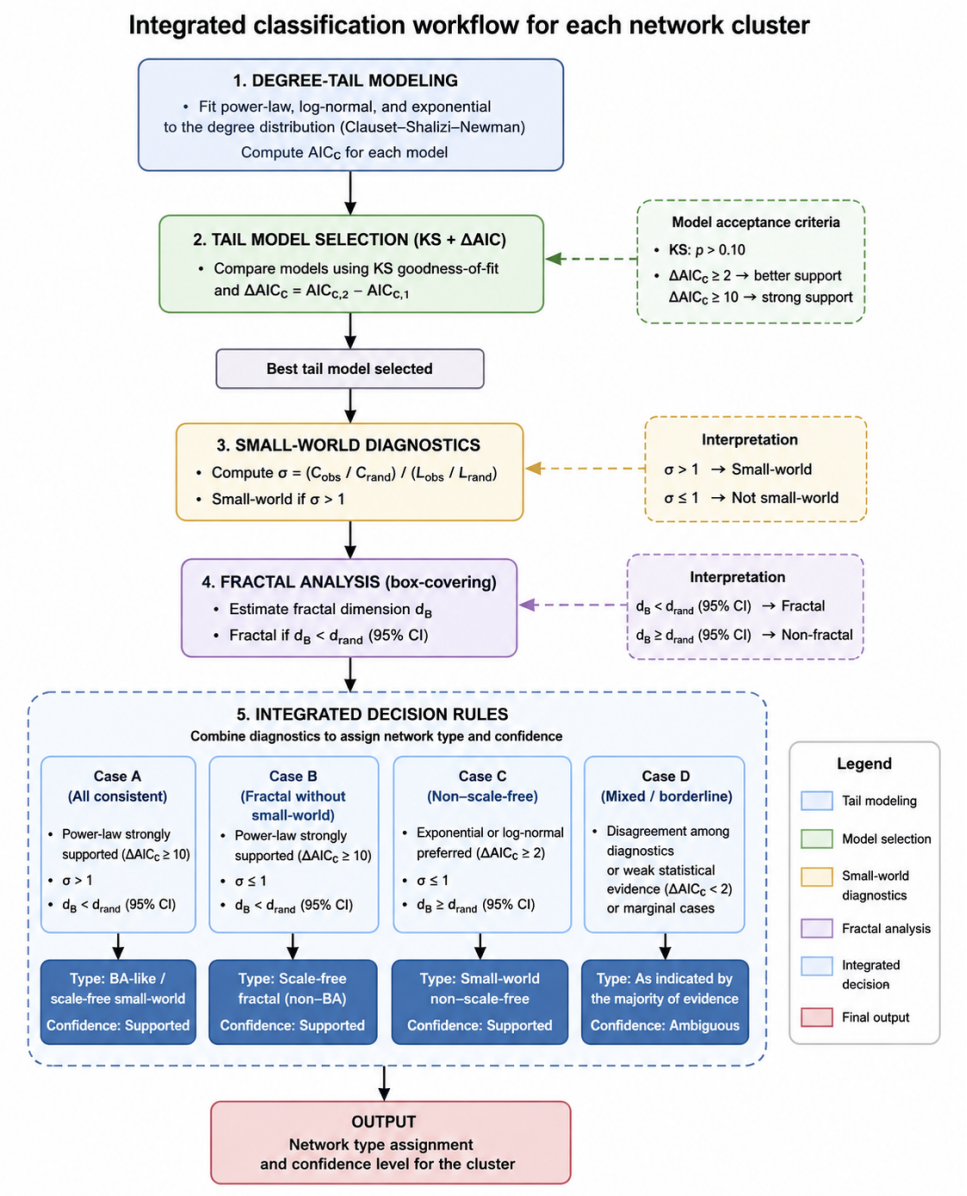


Figure 1. Integrated decision workflow for the mesoscopic classification of network clusters. The proposed framework combines complementary statistical and structural diagnostics into a unified decision protocol. The workflow begins with degree-tail modeling, followed by tail-model selection using the bootstrap KS goodness-of-fit test and ΔAIC. Small-world diagnostics and fractal box-covering analysis provide additional structural evidence. The final network type is assigned by integrating all available diagnostics, and a qualitative confidence level summarizes the overall agreement among the independent criteria. The dark-blue boxes in the integrated decision stage indicate the resulting network type and confidence level assigned under each decision case.

For each author, the cumulative weighted degree was computed annually from the year of the first publication recorded in the BDPA database until the end of the observation period (2024). The temporal variable used in the regression was therefore the elapsed time since the author's entry into the collaboration network,

$$\tau = t - t_i + 1,$$

which avoids undefined values at the first observation and standardizes the temporal evolution of authors entering the network in different years.

The exponent β was estimated independently for each author by fitting a linear regression in logarithmic coordinates,

$$\log k_i(\tau) = \log A_i + \beta_i \log \tau, \tag{22}$$

where β_i is the estimated temporal growth exponent and A_i is a normalization constant. Only regressions with statistically significant slopes ($p < 0.05$) were retained for subsequent analyses. The resulting β estimates were then grouped according to the cluster membership obtained from the integrated structural classification protocol.

The temporal growth analysis is not part of the proposed classification framework and was not used to assign network types. Instead, the exponent β is employed exclusively as a complementary descriptive indicator that provides an independent temporal perspective on the evolution of the structurally classified clusters. Consequently, the identification of BA-like/scale-free, scale-free fractal (non-BA), and small-world (non-scale-free) regimes relies entirely on the structural diagnostics described in the previous subsections.

Some limitations should be noted. Because authors enter the collaboration network at different times, the estimated exponents may be affected by finite observation windows and survivorship bias, particularly for authors with short publication histories. Furthermore, although only statistically significant regressions ($p < 0.05$) were retained, no formal multiple-testing correction was applied. Therefore, the estimated β values should be interpreted as descriptive indicators of temporal growth rather than as definitive evidence of preferential attachment, node fitness, or any specific network generative mechanism.

3. Results

Figure 2 shows the clusters identified using the Louvain algorithm.

Table 1 summarizes the main structural metrics computed for each community identified within the Embrapa coauthorship network. These parameters describe how densely each subnetwork is connected, the extent of clustering among neighboring nodes, the average distance between any two authors, and the degree of homophily or disassortative mixing. Together, these metrics allow us to compare structural regimes across clusters and evaluate how local connectivity patterns may affect robustness, information diffusion, and collaboration efficiency.

In general, denser and highly clustered subnetworks indicate more cohesive research groups, while lower density and longer path lengths are typical of large, heterogeneous communities that bridge distinct thematic or institutional domains [1,4]. Degree assortativity further reveals whether well-connected authors preferentially collaborate with similarly connected peers or with less-connected individuals, shaping the hierarchical or more egalitarian character of the collaboration structure [21]. The data show strong heterogeneity among clusters. Larger communities (e.g., clusters 0–5) exhibit very low density (on the order of 10^{-3}) and moderate assortativity, consistent with scale-free small-world structures where a few highly connected authors act as bridges among multiple local groups. These clusters also show high average clustering coefficients (around 0.8) combined with moderate path lengths ($L \approx 4$), indicating efficient local cohesion and global reach, a hallmark of small-world organization.

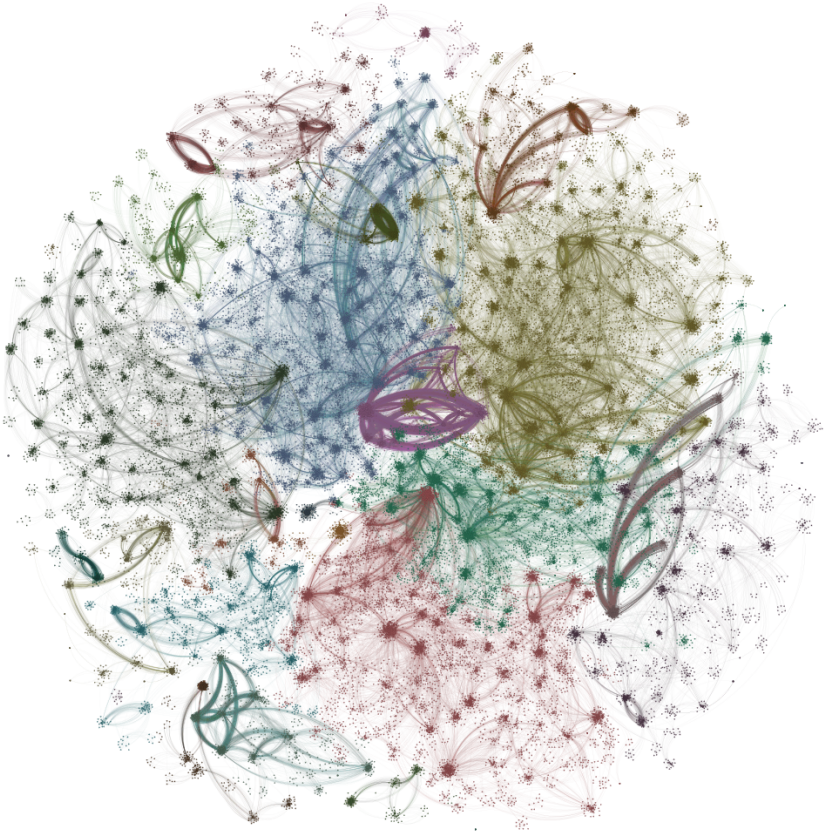


Figure 2. Community structure of the empirical coauthorship network identified using the Louvain modularity algorithm. The network was partitioned into 25 clusters, each represented by a different color. Nodes correspond to authors and edges represent coauthorship relationships.

In contrast, smaller clusters (e.g., 12–15) are characterized by high density (>0.4) and clustering (>0.9), as well as short average path lengths ($L < 2$). These metrics reveal tightly knit research groups, often corresponding to specialized or institutionally localized collaborations. Clusters with negative assortativity (e.g., 2, 6, 17, 23) suggest hub-mediated communication, where central authors connect otherwise weakly linked nodes, an architecture that enhances reachability but can reduce resilience to targeted node removal. Conversely, positive assortativity values (e.g., clusters 7, 10, 13, 19) point to more balanced collaboration patterns, where partnerships among well-connected researchers reinforce the network's structural stability and knowledge diffusion.

Overall, the observed combination of high clustering, moderate path lengths, and heterogeneous densities supports the interpretation of the Embrapa coauthorship system as a multi-scale, small-world network, simultaneously favoring local cohesion and global connectivity—an advantageous configuration for fostering interdisciplinary collaboration and long-term institutional resilience.

Table 2 presents the parameters resulting from the tail-fit analysis performed for each cluster in the Embrapa coauthorship network. The table includes, for each cluster, the cutoff value $x_{\min}$, the number of observations in the degree tail (n_{tail}), the power-law exponent α, the Kolmogorov–Smirnov distance (D) and corresponding p-value, as well as the AIC values for the three alternative models tested: power-law (PL), exponential (EXP), and log-normal (LN). The final column ("Best tail") identifies the model providing the best statistical description of the empirical weighted degree distribution according to the minimum AIC.

Table 1. Core structural metrics of the Louvain clusters identified in the Embrapa coauthorship network. The reported metrics characterize the structural organization of each cluster and constitute the basis for the subsequent small-world analysis. The columns report the number of nodes (n_{nodes}), number of edges (n_{edges}), edge density, degree assortativity coefficient, average local clustering coefficient, average shortest-path length, and global transitivity.

Cluster	n_{nodes}	n_{edges}	Density	Assortativity	Avg. Clustering	Avg. Path Length	Transitivity
0	12,127	91,421	0.001243	0.009441	0.822467	4.053900	0.319526
1	10,697	87,247	0.001525	0.059103	0.831831	4.060400	0.430386
2	8468	47,162	0.001316	−0.065824	0.834588	4.449545	0.215914
3	6960	43,406	0.001792	0.018119	0.821600	4.257426	0.378742
4	5144	39,519	0.002988	0.099192	0.850496	4.206421	0.580947
5	2461	13,832	0.004569	0.063613	0.838042	4.407874	0.527465
6	2222	339,138	0.137440	−0.067363	0.875751	2.020707	0.658145
7	1800	19,315	0.011929	0.422028	0.887630	3.789194	0.753500
8	1267	21,709	0.027068	0.249995	0.890120	3.250949	0.778782
9	1246	17,379	0.022406	0.039641	0.884913	3.059358	0.672972
10	1202	18,894	0.026176	0.395142	0.880505	3.449460	0.784321
11	1105	48,145	0.078931	0.492656	0.911597	2.660535	0.889654
12	921	173,151	0.408703	0.258058	0.930000	1.695322	0.881355
13	875	13,122	0.034317	0.619113	0.889112	3.477604	0.881345
14	818	259,194	0.775675	0.294743	0.991521	1.235688	0.992624
15	818	192,501	0.576086	0.498531	0.976851	1.477431	0.997980
16	740	12,536	0.045847	0.509017	0.888424	3.729386	0.898428
17	662	54,167	0.247574	−0.024269	0.885950	2.029910	0.798851
18	423	15,319	0.171636	0.876584	0.939893	2.654835	0.982056
19	234	6230	0.228532	0.894463	0.918984	2.476872	0.981866
20	230	5546	0.210594	0.571527	0.927852	2.278717	0.940652
21	168	5193	0.370188	0.206337	0.942436	1.737454	0.890268
22	21	73	0.347619	0.285401	0.916484	1.329670	0.816667
23	21	61	0.290476	−0.216832	0.903252	1.852381	0.654412
24	6	15	1.000000	–	1.000000	1.000000	1.000000

Table 2. Tail-fitting parameters for the weighted degree distributions of the Louvain clusters. Reported quantities include $x_{\min}$, n_{tail}, α, the KS statistic (D), bootstrap p-value, the AIC values for the PL, EXP, and LN models, and the corresponding best tail (minimum AIC).

Cluster	$x_{\min}$	n_{tail}	α	KS-D	KS-p	AIC_{PL}	AIC_{EXP}	AIC_{LN}	Best Tail
0	123	110	3.6405	0.047413	0.950000	1149.485	1156.240	1195.100	power-law
1	155	52	4.4787	0.048938	0.991667	530.424	529.315	546.860	exponential
2	34	335	2.8762	0.041791	0.975000	2960.235	3065.528	3127.696	power-law
3	106	59	4.4514	0.059379	0.975000	557.739	557.860	580.241	power-law
4	92	54	4.2576	0.065138	0.925000	503.373	503.698	524.307	power-law
5	11	633	2.3436	0.099526	0.891667	4813.238	4910.433	5010.413	power-law
6	535	419	4.1264	0.060153	0.075000	5416.584	5403.641	5531.764	exponential
7	9	1038	2.0009	0.084855	0.983333	8593.068	8742.630	8807.665	power-law
8	4	1124	1.5971	0.111960	1.000000	9989.909	10,198.685	10,010.183	power-law
9	5	1060	1.7509	0.102312	1.000000	8741.309	9129.814	8968.789	power-law
10	5	958	1.6520	0.099437	0.975000	8558.166	8658.262	8586.572	power-law
11	2	1080	1.3080	0.232758	0.000000	12,595.997	11,811.692	11,902.459	exponential
12	640	74	16.4846	0.137773	0.100000	710.242	709.422	735.919	exponential
13	5	690	1.6665	0.098119	0.991667	6087.890	6174.493	6123.641	power-law
14	722	88	48.0227	0.097780	0.516667	662.346	662.440	709.302	power-law
15	620	51	98.7751	0.329421	0.000000	293.360	294.461	360.920	power-law
16	3	686	1.5340	0.180071	0.175000	6061.543	6188.433	6086.875	power-law
17	222	266	4.1347	0.226914	0.000000	2968.912	2923.256	2909.483	log-normal
18	3	419	1.4023	0.202547	0.000000	4453.742	4401.524	4347.028	log-normal
19	3	227	1.4208	0.254640	0.000000	2344.048	2248.478	2263.653	exponential
20	40	162	3.2021	0.335176	0.000000	1408.483	1389.301	1432.777	exponential
21	2	167	1.3114	0.345027	0.000000	1933.524	1704.543	1765.500	exponential
22	–	0	–	–	–	–	–	–	undefined
23	–	0	–	–	–	–	–	–	undefined
24	–	0	–	–	–	–	–	–	undefined

The final network-type classification summarized in Table 3 was obtained by applying the integrated multi-criteria protocol described in Section 2.5. This protocol combines

tail-model selection based on the weighted degree distribution, the bootstrap KS goodness-of-fit test, the small-worldness index (σ), and the box-covering analysis to identify coherent mesoscopic structural regimes. The reported confidence level reflects the overall agreement among these independent statistical and structural diagnostics.

Table 3 summarizes the quantities entering the decision protocol together with the resulting network-type assignment and its associated confidence level.

Table 3. Summary of the statistical and structural diagnostics incorporated into the integrated classification protocol. The final network-type assignment is accompanied by a qualitative confidence level reflecting the overall agreement among the statistical and structural diagnostics. Columns correspond to the quantities entering the decision protocol: ΔAIC obtained from the weighted degree-tail fitting, p_{KS} (power-law plausibility), σ (small-worldness), and the fractal dimension d_B estimated from the box-covering analysis. Values of d_B are reported only for clusters exhibiting supported fractal scaling.

Cluster	Best Tail	ΔAIC	p_{KS}	σ	d_B	Network Type	Confidence
0	power-law	6.755	0.950	238.989	3.645	BA-like/scale-free small-world	Supported
1	exponential	1.108	0.992	260.596	3.962	Small-world non-scale-free	Ambiguous
2	power-law	105.293	0.975	156.725	3.552	BA-like/scale-free small-world	Supported
3	power-law	0.120	0.975	191.069	3.560	BA-like/scale-free small-world	Ambiguous
4	power-law	0.325	0.925	159.211	2.972	BA-like/scale-free small-world	Ambiguous
5	power-law	97.195	0.892	89.495	2.784	Scale-free fractal (non-BA)	Supported
6	exponential	12.943	0.075	4.413	–	Small-world non-scale-free	Ambiguous
7	power-law	149.562	0.983	45.837	2.704	Scale-free fractal (non-BA)	Supported
8	power-law	20.273	1.000	21.027	–	BA-like/scale-free small-world	Supported
9	power-law	227.480	1.000	24.590	–	BA-like/scale-free small-world	Supported
10	power-law	28.406	0.975	20.926	2.649	Scale-free fractal (non-BA)	Supported
11	exponential	90.767	0.000	8.137	–	Small-world non-scale-free	Ambiguous
12	exponential	0.820	0.100	2.024	–	Small-world non-scale-free	Ambiguous
13	power-law	35.751	0.992	16.996	–	BA-like/scale-free small-world	Supported
14	power-law	0.094	0.517	1.268	–	BA-like/scale-free small-world	Ambiguous
15	power-law	1.101	0.000	1.670	–	BA-like/scale-free small-world	Ambiguous
16	power-law	25.332	0.175	11.228	2.042	BA-like/scale-free small-world	Supported
17	log-normal	13.773	0.000	2.783	–	Small-world non-scale-free	Ambiguous
18	log-normal	54.496	0.000	3.945	–	Small-world non-scale-free	Ambiguous
19	exponential	15.175	0.000	3.072	–	Small-world non-scale-free	Ambiguous
20	exponential	19.182	0.000	3.516	–	Small-world non-scale-free	Ambiguous
21	exponential	60.958	0.000	2.256	–	Small-world non-scale-free	Ambiguous
22	–	–	–	3.099	–	Undefined	–
23	–	–	–	2.368	–	Undefined	–
24	–	–	–	1.000	–	Undefined	–

Although the integrated classification protocol is based on the combined interpretation of multiple statistical and structural diagnostics, it is useful to report the fitted parameters associated with each candidate tail model. Table 4 summarizes the estimated parameters obtained during the tail-fitting procedure for every cluster, including the power-law exponent (α), the parameters of the log-normal distribution ($\ln \mu$ and $\ln \sigma$), and the exponential decay rate (λ_{exp}). These quantities provide a complete description of the fitted statistical models and facilitate future comparisons with other empirical collaboration networks.

Reporting the fitted parameters also provides additional insight into the diversity of degree-tail behaviors observed across the collaboration network. While the integrated classification relies on multiple complementary diagnostics, the estimated model parameters quantify the characteristics of the candidate distributions themselves, allowing differences in scaling exponents, dispersion, and exponential decay rates to be directly compared among clusters. Such information may be useful in future studies investigating the evolution of collaboration patterns or benchmarking alternative network growth models.

The final network-type assignment is not determined by these parameters alone but results from the integrated multi-criteria protocol described in Section 2 and summarized in Table 3.

Table 4. Estimated parameters of the candidate tail models for each cluster. For completeness, the table reports the parameters associated with the power-law (α), log-normal ($\ln \mu$, $\ln \sigma$), and exponential (λ_{exp}) models together with the corresponding best-fitting tail model. The final network-type assignment and confidence level follow the integrated classification protocol summarized in Table 3.

Cluster	Best Tail	n_{tail}	α	$\ln \mu$	$\ln \sigma$	λ_{exp}	Network Type	Confidence
0	power-law	110	3.641	5.190	0.355	0.0143	BA-like/scale-free small-world	Supported
1	exponential	52	4.479	5.330	0.250	0.0170	Small-world non-scale-free	Ambiguous
2	power-law	335	2.876	4.054	0.530	0.0281	BA-like/scale-free small-world	Supported
3	power-law	59	4.451	4.952	0.264	0.0245	BA-like/scale-free small-world	Ambiguous
4	power-law	54	4.258	4.827	0.281	0.0261	BA-like/scale-free small-world	Ambiguous
5	power-law	633	2.344	3.122	0.639	0.0563	Scale-free fractal (non-BA)	Supported
6	exponential	419	4.126	6.602	0.274	0.0040	Small-world non-scale-free	Ambiguous
7	power-law	1038	2.001	3.167	0.798	0.0403	Scale-free fractal (non-BA)	Supported
8	power-law	1124	1.597	2.973	1.161	0.0291	BA-like/scale-free small-world	Supported
9	power-law	1060	1.751	2.880	1.052	0.0367	BA-like/scale-free small-world	Supported
10	power-law	958	1.652	3.076	1.079	0.0297	Scale-free fractal (non-BA)	Supported
11	exponential	1080	1.308	3.694	1.524	0.0110	Small-world non-scale-free	Ambiguous
12	exponential	74	16.485	6.526	0.058	0.0230	Small-world non-scale-free	Ambiguous
13	power-law	690	1.667	3.043	1.069	0.0310	BA-like/scale-free small-world	Supported
14	power-law	88	48.023	6.603	0.022	0.0638	BA-like/scale-free small-world	Ambiguous
15	power-law	51	98.775	6.440	0.018	0.1550	BA-like/scale-free small-world	Ambiguous
16	power-law	686	1.534	2.848	1.297	0.0299	BA-like/scale-free small-world	Supported
17	log-normal	266	4.135	5.721	0.198	0.0112	Small-world non-scale-free	Ambiguous
18	log-normal	419	1.402	3.442	1.461	0.0143	Small-world non-scale-free	Ambiguous
19	exponential	227	1.421	3.333	1.314	0.0190	Small-world non-scale-free	Ambiguous
20	exponential	162	3.202	4.139	0.354	0.0380	Small-world non-scale-free	Ambiguous
21	exponential	167	1.311	3.652	1.240	0.0170	Small-world non-scale-free	Ambiguous
22	–	–	–	–	–	–	Undefined	–
23	–	–	–	–	–	–	Undefined	–
24	–	–	–	–	–	–	Undefined	–

3.1. Supported BA-like/Scale-Free Small-World Clusters

According to the integrated classification protocol (Tables 3 and 4), clusters 000, 002, 008, 009, 013, and 016 received supported BA-like/scale-free small-world classifications. These clusters consistently combine statistically supported heavy-tailed weighted degree distributions, pronounced small-world organization ($\sigma > 1$), and the absence of fractal scaling according to the box-covering analysis. Clusters 22, 23, and 24 were excluded from the classification because they contain too few nodes to provide reliable statistical estimates (Table 1).

Figure 3 illustrates the topology of the supported BA-like/scale-free small-world clusters. The network layouts were generated using the OpenOrd algorithm [21], which provided the clearest visualization of the internal modular organization and hub structure of each cluster. The adopted layout parameters were: liquid stage 25%, expansion stage 25%, cooldown stage 25%, crunch stage 10%, and simmer stage 15%; edge cut equal to 0.8; 19 computation threads; 2000 iterations; fixed time step of 0.2; and random seed—2527419387533434. This configuration emphasizes community boundaries while preserving the spatial separation of densely connected regions.

The estimated power-law exponents for clusters 000, 002, 008, 009, 013, and 016 are approximately 3.64, 2.88, 1.60, 1.75, 1.67, and 1.53, respectively. Together with the bootstrap KS test, the AIC-based tail-model selection, and the small-world diagnostics, these results support their classification as BA-like/scale-free small-world networks according to the integrated decision protocol.

The topologies shown in Figure 3 reveal heterogeneous hub-dominated structures in which a relatively small number of highly connected authors coexist with many peripheral collaborators. Such an organization is compatible with heavy-tailed weighted degree distributions and decentralized collaboration patterns commonly observed in empirical

scientific networks. Cluster 002 provides the strongest overall evidence for BA-like organization, combining a well-supported heavy-tailed distribution ($\alpha \approx 2.88$), pronounced hub dominance, and temporal growth behavior that is closest to the canonical Barabási–Albert scaling inferred from the β analysis.

It is important to emphasize that the coexistence of scale-free degree distributions and small-world organization is not unique to Barabási–Albert networks. A well-known example is the worldwide air transportation network, which simultaneously exhibits heavy-tailed connectivity, short average path lengths, high clustering, and strong modular organization [22]. In such systems, highly connected hubs are not necessarily the most central vertices with respect to shortest-path betweenness because community boundaries constrain information flow. Consequently, preferential attachment and local clustering may coexist within a modular architecture, producing hybrid scale-free small-world regimes.

Accordingly, the term *BA-like* is employed here as a phenomenological topological descriptor rather than as evidence of a specific generative mechanism. The proposed framework identifies structural compatibility with the characteristic organization of Barabási–Albert-type networks based on the combined statistical and structural diagnostics, but it does not establish that the underlying collaboration dynamics necessarily follow the Barabási–Albert growth process. Alternative mechanisms, including finite-size effects, modular organization, or node fitness heterogeneity, may produce similar topological signatures.

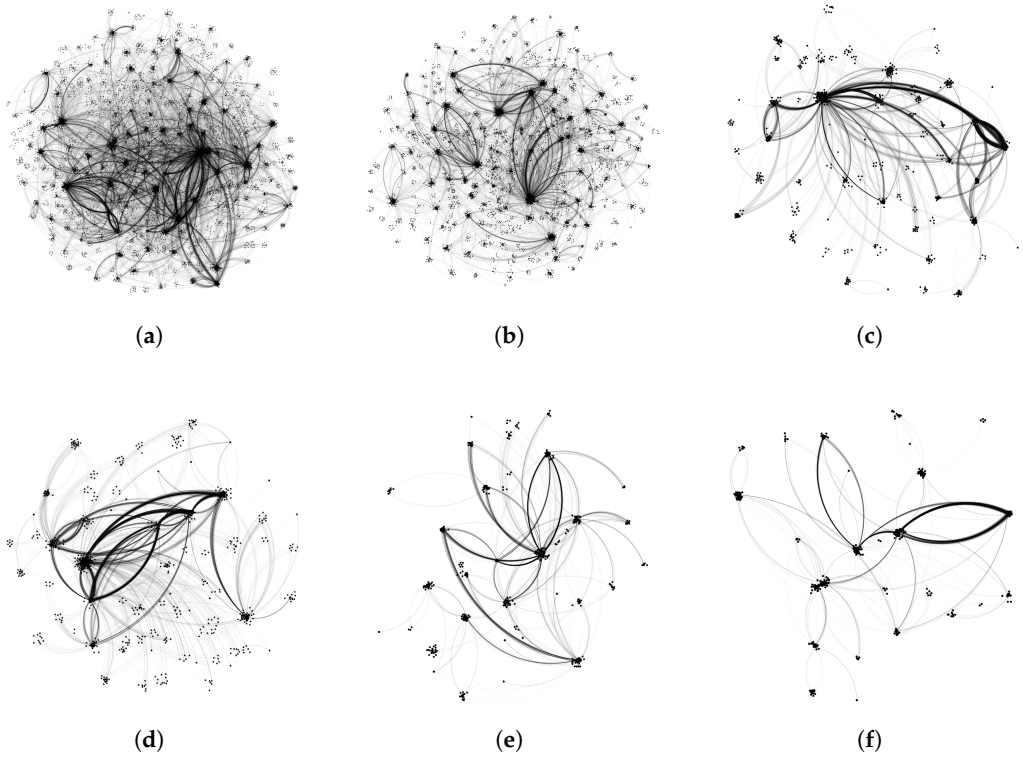


Figure 3. Representative supported BA-like/scale-free small-world clusters identified by the integrated classification protocol. These clusters exhibit consistent statistical support for heavy-tailed weighted degree distributions, pronounced small-world organization ($\sigma > 1$), and no evidence of fractal scaling according to the box-covering analysis. The term *BA-like* is used here as a phenomenological topological descriptor and does not imply that the underlying collaboration dynamics necessarily follow the Barabási–Albert generative mechanism. (**a**) Cluster 000; (**b**) Cluster 002; (**c**) Cluster 008; (**d**) Cluster 009; (**e**) Cluster 013; (**f**) Cluster 016.

Although the supported BA-like/scale-free small-world clusters share the same qualitative topological regime, they exhibit substantial variability in their scaling exponents, ranging from approximately $\alpha = 1.53$ to $\alpha = 3.64$. Such variability reflects different degrees

of hub dominance and degree heterogeneity while remaining consistent with the integrated classification protocol. Among the supported clusters, cluster 000 and cluster 002 exhibit the strongest overall agreement with the topological characteristics expected for BA-like organization, considering the combined statistical and structural diagnostics employed by the proposed framework.

The observed variability among the supported BA-like clusters suggests that preferential-attachment-like organization alone may not fully explain the structural diversity of the collaboration network. Similar deviations from idealized Barabási–Albert scaling have been widely reported in empirical complex networks and are frequently attributed to additional mechanisms such as finite-size effects, modular organization, or node fitness heterogeneity. These mechanisms may preserve the characteristic coexistence of heterogeneous connectivity and small-world organization while producing measurable differences in the degree distribution and temporal evolution of individual clusters.

This behavior has been extensively discussed in the network science literature. Bianconi and Barabási [23] proposed a fitness model in which nodes possess intrinsic abilities to compete for new links, leading to heterogeneous growth rates and, in some cases, a condensation phenomenon where a small number of highly fit nodes dominate the connectivity structure. Likewise, empirical studies of scientific collaboration networks [1,24] have reported mixed scaling regimes in which heavy-tailed degree distributions coexist with truncated power-law or log-normal behavior as a consequence of finite-size effects, temporal constraints, and heterogeneous author productivity.

To further investigate these mechanisms, we analyze the temporal growth exponent β, originally introduced in the context of preferential attachment and node fitness [20,23]. Although the proposed classification framework is based exclusively on structural diagnostics, the β exponent provides an independent temporal perspective on the growth dynamics of the supported BA-like clusters and helps assess the extent to which their evolution is compatible with preferential-attachment-like behavior. The distribution of β across the classified clusters is presented in Figure 4.

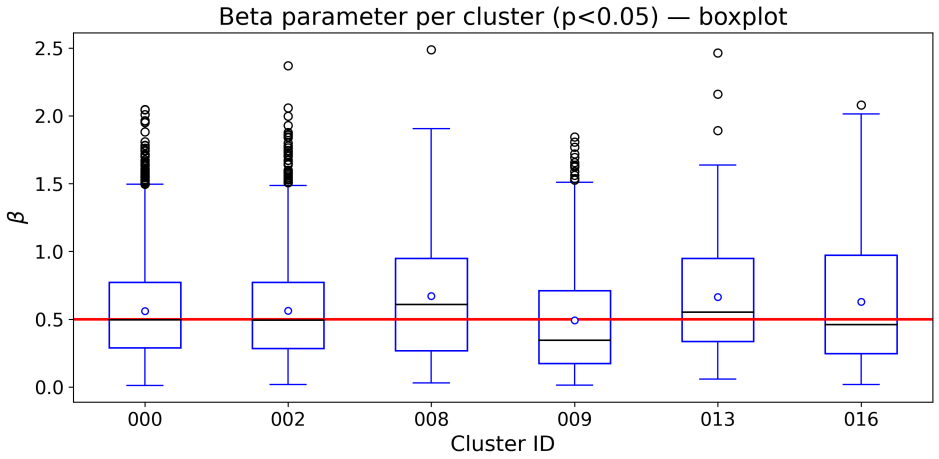


Figure 4. Distribution of the temporal growth exponent β for the supported BA-like/scale-free small-world clusters (only regressions with $p < 0.05$ were retained). The horizontal red line marks $\beta = 0.5$, corresponding to the theoretical prediction of the Barabási–Albert model. The exponent β is presented as a complementary temporal descriptor and is not used as a criterion in the integrated structural classification.

The distribution of the temporal growth exponent β for the supported BA-like/scale-free small-world clusters is shown in Figure 4. Overall, the median β values are close to the canonical Barabási–Albert prediction of $\beta = 0.5$, although noticeable variability is observed

among individual clusters. This result indicates that the supported BA-like clusters exhibit temporal growth broadly compatible with preferential-attachment-like dynamics while preserving measurable differences in their rates of connectivity accumulation.

Among the supported clusters, cluster 002 exhibits the median β value closest to the theoretical Barabási–Albert prediction, in agreement with the overall structural evidence discussed previously. The remaining clusters display moderate deviations from the theoretical reference while remaining within the range commonly observed for empirical collaboration networks. These differences reinforce the view that subnetworks belonging to the same mesoscopic structural regime may nevertheless evolve through quantitatively distinct temporal dynamics.

Such variability has been widely reported in empirical collaboration networks. Bianconi and Barabási [23] proposed that node fitness may alter the probability of acquiring new links, producing growth patterns that depart from the idealized Barabási–Albert model. Likewise, empirical studies [1,24] have shown that finite-size effects, heterogeneous author productivity, temporal constraints, and modular organization may all contribute to deviations from the canonical preferential-attachment prediction while preserving heavy-tailed degree distributions and small-world organization.

The temporal analysis presented here should therefore be interpreted as complementary to the structural classification proposed in this work. While the network typology is determined exclusively from the integrated structural diagnostics described in Section 2, the exponent β provides an independent temporal characterization of the supported BA-like clusters. Consequently, the observed deviations from the theoretical value $\beta = 0.5$ should not be interpreted as definitive evidence of preferential attachment, node fitness, or any specific network generative mechanism. Rather, they provide additional evidence that structurally similar collaboration communities may exhibit different temporal growth patterns, highlighting the intrinsic heterogeneity of large-scale scientific collaboration networks.

3.2. Ambiguous BA-like/Scale-Free Small-World Cases

The integrated classification protocol identified four clusters (003, 004, 014, and 015) whose statistical and structural diagnostics did not provide sufficiently consistent evidence for a definitive BA-like/scale-free small-world assignment. Although these clusters exhibited weighted degree distributions whose best-supported tail model was the power-law model, the combined evidence from the bootstrap KS test, model selection, and structural diagnostics was insufficient to satisfy all criteria required for a supported classification.

The corresponding network topologies are shown in Figure 5. As in the previous section, the layouts were generated using the OpenOrd algorithm with identical visualization parameters to facilitate direct comparison with the supported BA-like clusters.

Clusters 003 and 004 illustrate situations in which the power-law model provides the best fit according to the AIC, yet the evidence remains statistically weak because the competing models exhibit comparable support ($\Delta\text{AIC} < 2$). Consequently, although their weighted degree distributions are compatible with heavy-tailed behavior, they cannot be distinguished robustly from alternative tail models based solely on the available statistical evidence. Their classification therefore remains ambiguous within the proposed decision protocol.

A similar situation is observed for clusters 014 and 015. While the power-law model also minimizes the AIC for these clusters, the statistical support is considerably weaker owing to both the very small ΔAIC values and the lack of consistent agreement among the complementary structural diagnostics. In addition, these clusters contain relatively small degree tails and unusually large estimated scaling exponents (Table 4), indicating that the inferred power-law behavior is unstable and should be interpreted with caution. Under

these conditions, small variations in the empirical tail may substantially alter the preferred statistical model.

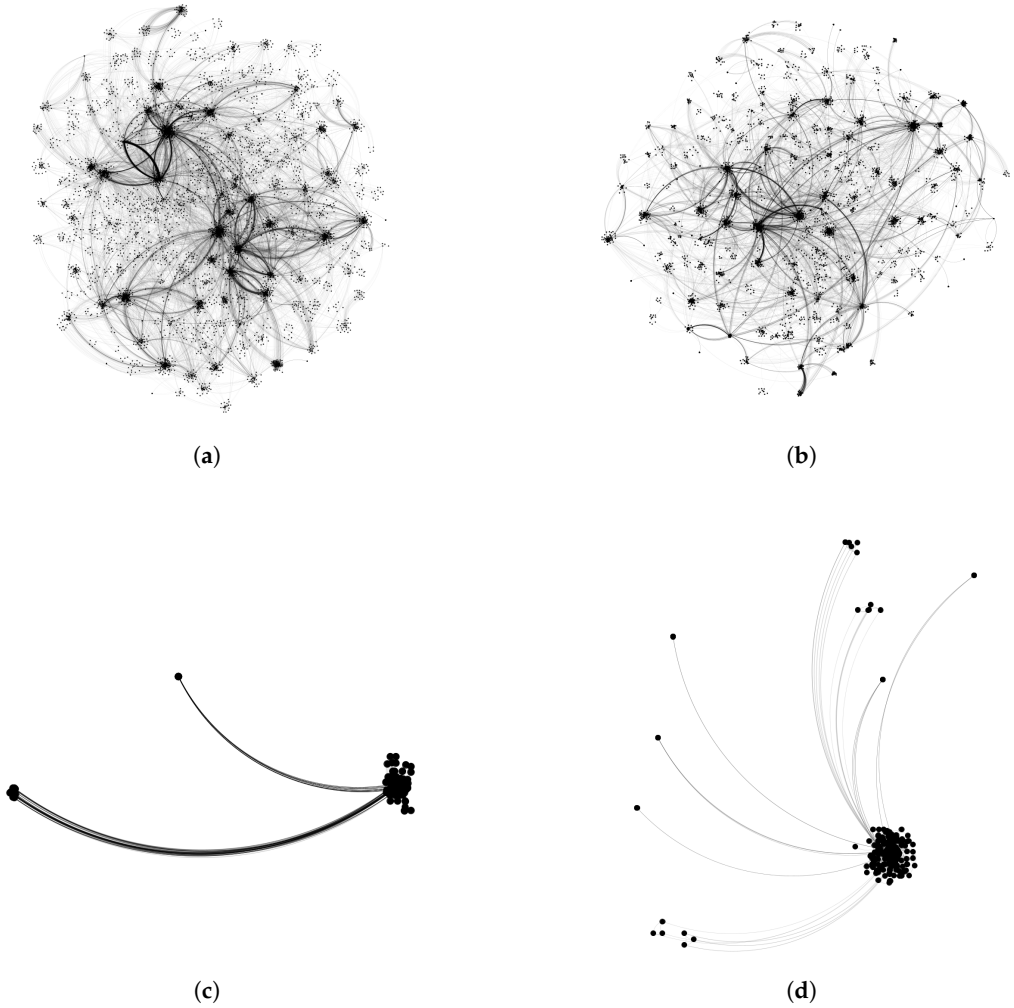


Figure 5. Representative ambiguous BA-like/scale-free small-world clusters identified by the integrated classification protocol. Although the power-law model provided the best statistical fit for these clusters, the combined evidence from the bootstrap KS test, model selection, and complementary structural diagnostics was insufficient to support a definitive BA-like/scale-free small-world classification. These examples illustrate situations in which the proposed framework explicitly preserves ambiguity rather than forcing an uncertain assignment. (**a**) Cluster 003; (**b**) Cluster 004; (**c**) Cluster 014; (**d**) Cluster 015.

For these reasons, the present framework intentionally avoids assigning definitive topological labels whenever the available evidence is inconclusive. Rather than forcing every cluster into a predefined category, the integrated protocol explicitly preserves ambiguous cases whenever the statistical or structural diagnostics fail to converge toward a unique interpretation. This conservative strategy reduces the risk of overinterpreting finite-size effects or statistical fluctuations as genuine topological regimes.

The existence of ambiguous clusters also illustrates an important characteristic of empirical collaboration networks. Real-world systems frequently exhibit transitional or mixed structural properties that cannot be adequately represented by idealized network models. Consequently, preserving an explicit ambiguous category constitutes an important feature of the proposed framework, allowing uncertain cases to be identified transparently instead of being artificially assigned to a specific topological regime.

3.3. Ambiguous Small-World (Non-Scale-Free) Cases

The integrated classification protocol identified clusters 001, 006, 011, 012, 017, 018, 019, 020, and 021 as cases that are structurally compatible with the small-world (non-scale-free) regime but whose statistical evidence is insufficient to support a definitive classification.

For all these clusters, the weighted degree distribution is better described by exponential or log-normal tails than by a statistically supported power-law model. However, some cases present limited statistical support for the preferred tail model or conflicting evidence among the complementary structural diagnostics. Consequently, according to the proposed decision protocol, these clusters are conservatively classified as ambiguous rather than being assigned a definitive small-world label.

Figure 6 illustrates the topology of these ambiguous small-world clusters. As in the previous sections, the layouts were generated using the Fruchterman–Reingold algorithm, which provides a clear visualization of the local connectivity patterns in relatively compact graphs.

Despite the absence of definitive statistical support, these clusters consistently exhibit the structural characteristics commonly associated with small-world organization. All of them display high clustering coefficients together with relatively short average path lengths, resulting in $\sigma > 1$, which indicates efficient local cohesion combined with global reachability.

The ambiguous small-world clusters nevertheless exhibit considerable structural diversity. Cluster 001 represents a large and sparse collaboration community with relatively long average path lengths, whereas clusters such as 012, 019, 020, and 021 are considerably smaller and denser, exhibiting clustering coefficients above 0.9 and average shortest-path lengths below three (Table 1). This diversity indicates that compact institutional groups and larger thematic collaboration communities may both display small-world organization despite exhibiting different connectivity patterns and degree-tail statistics.

Nevertheless, the integrated classification protocol intentionally avoids assigning definitive network types whenever the statistical evidence from degree-tail modeling remains inconclusive. Consequently, these clusters are interpreted as ambiguous small-world cases rather than confirmed representatives of a unique topological regime.

3.4. Supported Scale-Free Fractal (Non-BA) Clusters

According to the integrated classification protocol, clusters 005, 007, and 010 were consistently classified as *supported* scale-free fractal (non-BA) networks. These clusters simultaneously exhibit statistically supported heavy-tailed weighted degree distributions, fractal scaling supported by the box-covering analysis, and mutually consistent statistical and structural diagnostics according to the integrated classification protocol.

Figure 7 presents the topology of the supported fractal clusters. Identical visualization parameters were adopted for all layouts to facilitate structural comparison among the classified subnetworks. The layouts are presented solely to illustrate the qualitative structural organization of the clusters and were not used as classification criteria within the integrated classification protocol.

The three clusters exhibit the structural characteristics typically associated with fractal complex networks. Rather than being organized around one or a few dominant hubs that substantially shorten global distances, they display recursively nested modular structures connected through relatively sparse inter-module bridges. Such organization is consistent with the self-similar topology expected under the renormalization framework introduced by Song, Havlin, and Makse [6] and further developed by Rozenfeld et al. [7], in which the number of covering boxes decreases according to the power-law model with increasing

box diameter, resulting in a finite fractal dimension. Similar approaches have been widely adopted in subsequent studies of fractal complex networks [25,26].

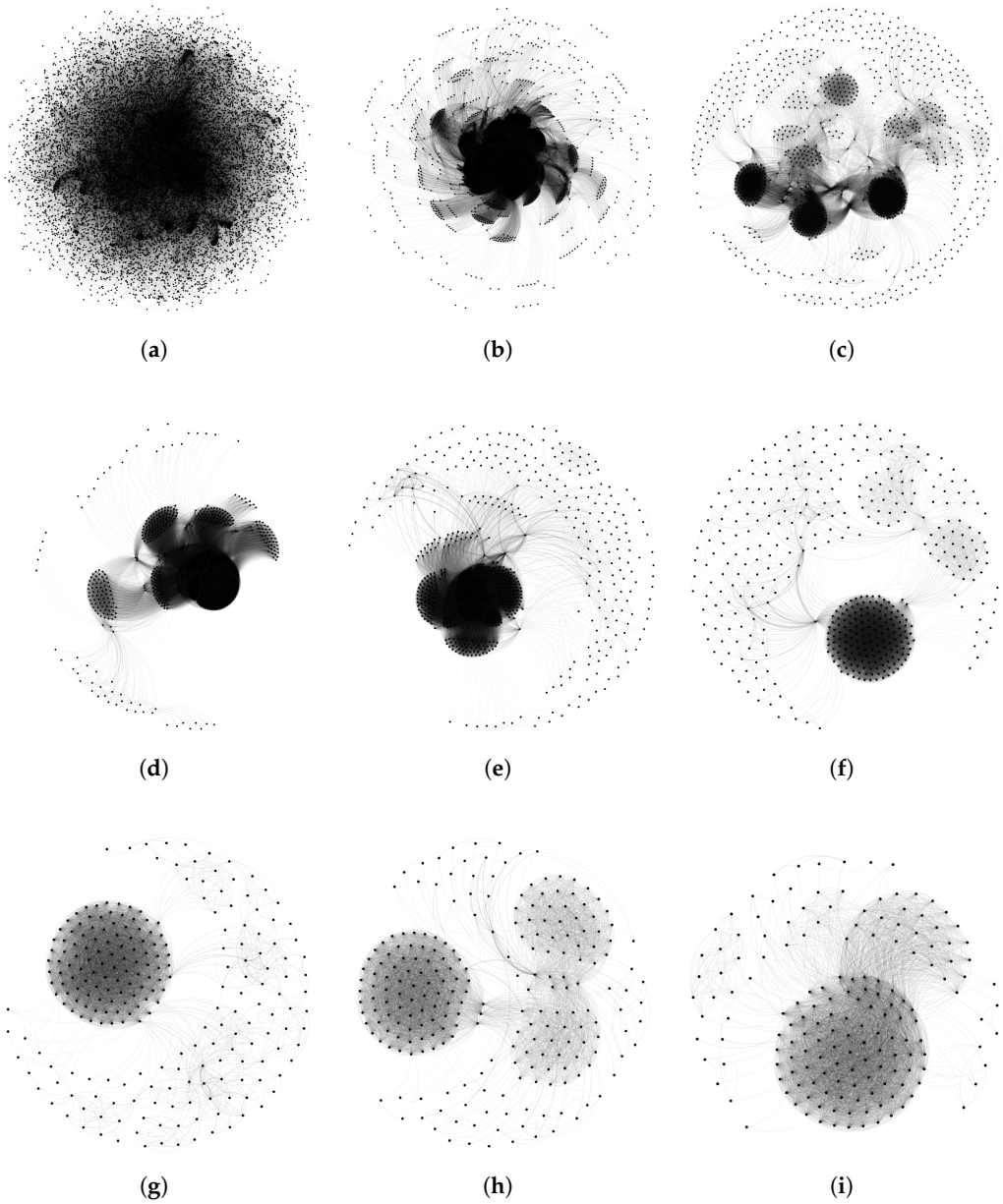
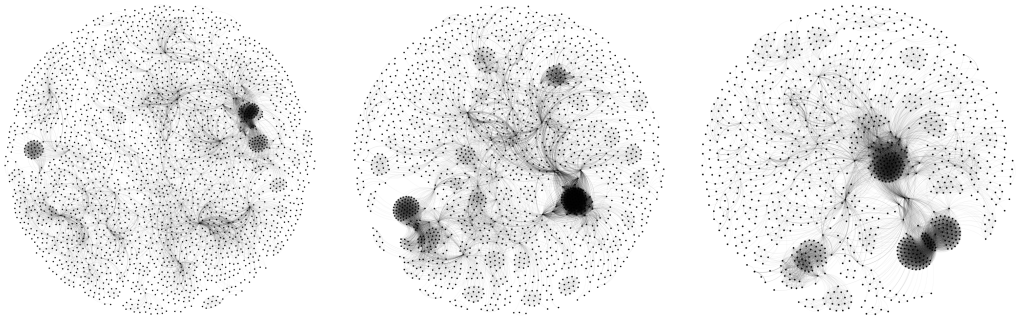


Figure 6. Representative ambiguous small-world (non-scale-free) clusters identified by the integrated classification protocol. Although these clusters exhibit the structural characteristics commonly associated with small-world organization, including high clustering coefficients and short average path lengths, the statistical evidence from degree-tail modeling and complementary structural diagnostics was insufficient to support a definitive network-type assignment. (**a**) Cluster 001; (**b**) Cluster 006; (**c**) Cluster 011; (**d**) Cluster 012; (**e**) Cluster 017; (**f**) Cluster 018; (**g**) Cluster 019; (**h**) Cluster 020; (**i**) Cluster 021.

This structural organization differs substantially from the supported BA-like/scale-free small-world clusters presented previously. Although both topological regimes exhibit statistically supported heavy-tailed weighted degree distributions, the fractal clusters preserve modular separation across multiple scales instead of concentrating connectivity into a small number of globally dominant hubs. Consequently, communication and collaboration tend to occur through hierarchically organized modules connected by relatively few bridging authors, rather than through highly centralized hub-mediated interactions.



(**a**) (**b**) (**c**)

Figure 7. Supported scale-free fractal (non-BA) clusters identified by the integrated classification protocol. These clusters combine statistically supported heavy-tailed weighted degree distributions with fractal scaling supported by the box-covering analysis, indicating self-similar modular organization rather than hub-dominated small-world topology. All three clusters satisfied the complete set of statistical and structural criteria required for a supported classification. (**a**) Cluster 005; (**b**) Cluster 007; (**c**) Cluster 010.

The estimated fractal dimensions ($d_B = 2.784$, 2.704, and 2.649 for clusters 005, 007, and 010, respectively; Table 3) provide quantitative support for this interpretation. In all three cases, the box-covering analysis favored the power-law scaling over the exponential alternative, indicating that the observed modular organization remains statistically compatible with self-similar scaling across the analyzed range of box diameters. These results provide an independent structural diagnostic that complements the degree-tail modeling, bootstrap KS goodness-of-fit test, AIC model selection, and small-world analysis incorporated into the integrated classification framework.

From the perspective of scientific collaboration, this topology suggests communities organized around relatively autonomous research groups connected through a limited number of bridging collaborations. Such an organization combines strong local cohesion with controlled communication between modules, potentially increasing robustness against localized disruptions while preserving thematic specialization. Analogous self-similar modular architectures have been reported in empirical biological, technological, and collaboration networks, where fractal organization emerges from the coexistence of heterogeneous connectivity and hierarchical modularity rather than from purely hub-dominated growth processes [7,25–27].

Unlike the BA-like and small-world cases discussed previously, no ambiguous fractal clusters were identified by the proposed decision protocol. All subnetworks exhibiting supported fractal scaling also presented mutually consistent statistical and structural diagnostics across the complementary analyses employed in this work, resulting in supported classifications. This outcome indicates that, within the Embrapa coauthorship network, the fractal regime constitutes a well-defined mesoscopic organization that is consistently identified by the integrated decision protocol. However, it is important to note that the higher thresholds adopted in this protocol may have reduced the number of networks identified as fractal, as described in the Supplementary Materials.

4. Discussion

Unless otherwise stated, the comparative analyses presented in this section summarize the structural properties of the network families identified by the integrated classification protocol. Each family includes both supported and ambiguous classifications, allowing the distributions to capture the overall structural variability associated with each topological regime. Consequently, the following boxplots are intended as descriptive summaries of

the identified network families and do not constitute additional evidence supporting the classification itself.

4.1. Assortativity Patterns

Figure 8 summarizes the distribution of degree assortativity (r) across the identified network families. The scale-free fractal family consistently exhibits positive assortativity, whereas the BA-like/scale-free small-world family displays lower and more variable values. The small-world (non-scale-free) family spans the widest assortativity range, including near-zero and slightly negative values. These observations suggest that the degree to which highly connected researchers preferentially collaborate with one another varies substantially among the identified network families and depends on the balance between hub formation, modular organization, and local collaboration patterns, in agreement with previous studies of scientific collaboration networks [5].

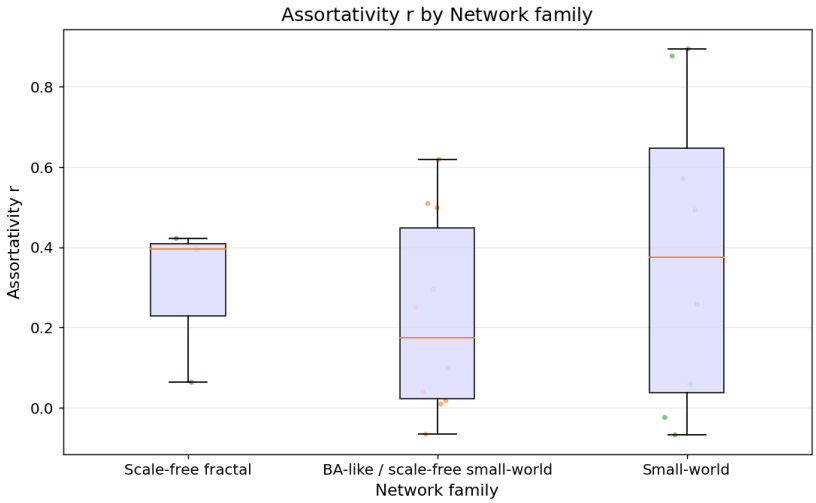


Figure 8. Distribution of degree assortativity (r) across the network families identified by the integrated classification protocol. Each family includes both supported and ambiguous classifications and is shown for descriptive comparison only. Boxes represent the interquartile range (25th–75th percentiles), horizontal lines within the boxes indicate the median, and individual points represent the assortativity values of each cluster, with different colors used to distinguish the network families.

4.2. Entropy Measures for Degree–Degree Mixing and Their Relation to Assortativity

Let r denote degree assortativity (the Pearson correlation between the degrees at the two ends of edges) [5,28]. A typical measure in networks is the Shannon entropy of the degree distribution

$$H_k = -\sum_k P(k) \log_2 P(k). \tag{23}$$

However, H_k depends only on the marginal $P(k)$ and is invariant under any degree-preserving rewiring of the graph. Since r can vary widely under degree-preserving rewirings, there is no generic monotone relationship between H_k (or its normalized variant) and r. Thus, H_k is not appropriate for studying mixing patterns [28,29].

To capture the randomness or structure that generates assortativity, we work with the joint distribution of degrees at the two ends of edges. Let

$$e_{jk} = \Pr\{\text{a uniformly random edge joins degrees } j \text{ and } k\}, \tag{24}$$

for an undirected network (hence $e_{jk} = e_{kj}$) [28]. Define the edge-end marginal

$$q_k = \sum_j e_{jk}, \qquad \sum_k q_k = 1. \tag{25}$$

From e_{jk} and q_k we build the following entropic quantities [30]:

$$H_{\text{edge}} = -\sum_{j,k} e_{jk} \log_2 e_{jk} \qquad \text{(edge joint entropy)} \tag{26}$$

$$H(q) = -\sum_k q_k \log_2 q_k \qquad \text{(edge-end entropy)} \tag{27}$$

$$H(K_2 \,|\, K_1) = -\sum_j q_j \sum_k \frac{e_{jk}}{q_j} \log_2\left(\frac{e_{jk}}{q_j}\right) \qquad \text{(conditional entropy)} \tag{28}$$

$$I(K_1; K_2) = \sum_{j,k} e_{jk} \log_2\left(\frac{e_{jk}}{q_j q_k}\right), \tag{29}$$

$$I(K_1; K_2) = H(q) - H(K_2 \,|\, K_1), \tag{30}$$

$$I(K_1; K_2) = 2H(q) - H_{\text{edge}} \qquad \text{(mutual information).} \tag{31}$$

If the network is degree–degree independent (no mixing), then $e_{jk} = q_j q_k$, so $I = 0$ and $H(K_2 \mid K_1) = H(q)$ while $r \approx 0$ [28,30]. As the strength of degree–degree dependence increases (assortative or disassortative), I increases and $H(K_2 \mid K_1)$ decreases (neighbor degrees become more predictable); the *sign* is given by r [30]. For comparability across clusters with different q, it is useful to report a normalized measure such as

$$\text{NMI} = \frac{I(K_1; K_2)}{H(q)} \in [0, 1] \quad \text{or} \quad 1 - \frac{H(K_2 \,|\, K_1)}{H(q)}. \tag{32}$$

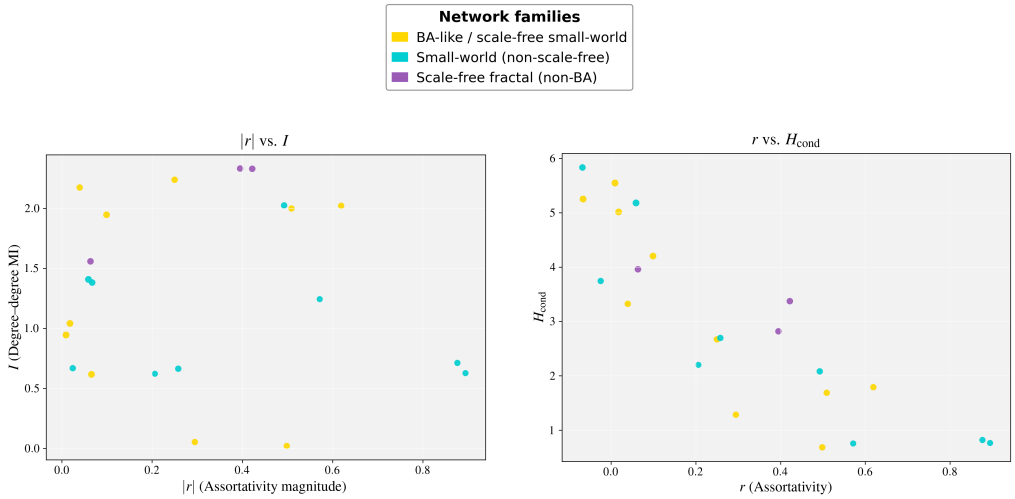
Although NMI is not the focus of the present analysis, it is defined here for completeness and may facilitate cross-study comparisons.

For each cluster we computed H_k (and its normalized version), $I(K_1; K_2)$, $H(K_2 \,|\, K_1)$, H_{edge}, and $H(q)$, and merged these quantities with the network-family assignment obtained from the integrated classification protocol. Unless otherwise stated, the analyses presented below include both supported and ambiguous classifications within each network family. The panels in Figure 9 summarize the key relationships.

Figure 9 presents two complementary two-dimensional projections of the correlation–information space. Panel (a) displays the relationship between the assortativity magnitude $|r|$ and the degree–degree mutual information I, whereas panel (b) shows assortativity r as a function of the conditional entropy of degree correlations ($H_{\text{cond}} \equiv H(K_2 \mid K_1)$). In both panels, point colors identify the network family assigned by the integrated classification protocol. As discussed previously, each family includes both supported and ambiguous classifications and is presented here as a descriptive summary of the structural variability within each topological regime.

The two panels highlight complementary aspects of degree–degree organization. Because mutual information quantifies the statistical dependence between the degrees at the two ends of an edge irrespective of whether the correlation is assortative or disassortative, whereas the assortativity coefficient r measures both the direction and strength of the linear degree correlation, no generic monotonic relationship between I and $|r|$ is expected [28,30]. This behavior is reflected in panel (a), where clusters exhibiting similar assortativity magnitudes may nevertheless present substantially different mutual-information values. Such

variability indicates that networks with comparable Pearson correlations may possess distinct joint degree distributions and, consequently, different levels of statistical dependence between neighboring vertices. Mutual information therefore captures structural information that is not fully summarized by the assortativity coefficient alone [30].



(**a**) (**b**)

Figure 9. Bivariate relationships between (**a**) assortativity magnitude ($|r|$) and mutual information (I), and (**b**) assortativity (r) and conditional entropy ($H_{\text{cond}} \equiv H(K_2 \mid K_1)$). Point colors indicate the network families identified by the integrated classification protocol. Each family includes both supported and ambiguous classifications and is presented for descriptive comparison of the information-theoretic properties associated with each topological regime. (**a**) $|r|$ vs. I; (**b**) r vs. H_{cond}.

Panel (b) reveals a more complex relationship between assortativity and conditional entropy. While highly assortative clusters typically exhibit low conditional entropy values, clusters with similar assortativity coefficients may differ substantially in their conditional entropy. For instance, clusters with assortativity around 0.5 present conditional entropies ranging from approximately 0.68 (cluster 15) to 2.08 (cluster 11). This dispersion indicates that even when linear correlations are comparable, the predictability of neighboring degrees can vary considerably, reflecting differences in the underlying joint degree distributions that are not captured by the assortativity coefficient alone. Such behavior is consistent with the information-theoretic identity

$$I(K_1; K_2) = H(q) - H(K_2|K_1),$$

which shows that the conditional entropy depends on both the mutual information and the marginal entropy of the edge-end degree distribution. As expected, I is bounded by the entropy of the edge-end degree distribution $H(q)$, with $I \leq H(K_1) = H(K_2) = H(q)$ [30]. Consequently, networks characterized by narrower edge-end degree distributions tend to attain smaller mutual-information values, whereas normalized measures such as NMI $= I/H(q)$ facilitate comparisons among networks with different degrees of heterogeneity.

The information-theoretic descriptors provide additional context for interpreting the network families identified by the integrated classification protocol. Considerable overlap is observed among the families in both projections, indicating that neither mutual information nor conditional entropy should be regarded as independent discriminative features. The scale-free fractal clusters generally occupy an intermediate region of the information space, with mutual-information values ranging from approximately 1.56 to 2.33, consistent with the hierarchically modular organization expected for self-similar networks [6,27,31,32]. In contrast, BA-like/scale-free small-world and small-world (non-scale-free) clusters occupy a

substantially broader portion of the correlation–information space, spanning both low- and high-information regimes. This dispersion indicates that similar topological classifications may still exhibit markedly different degree–degree mixing patterns, emphasizing that these information-theoretic quantities describe structural organization rather than the underlying network generation mechanism [4,5,20,28,33,34].

Overall, the two-dimensional projections reinforce that assortativity, mutual information, and conditional entropy provide complementary descriptions of degree–degree mixing. While assortativity summarizes the direction and linear strength of degree correlations, mutual information captures the overall statistical dependence encoded in the joint degree distribution, and conditional entropy quantifies the predictability of neighboring degrees. Rather than defining sharply separated regions for the identified topological families, these descriptors provide additional structural information that complements the statistical and topological diagnostics employed throughout this work. This interpretation is consistent with previous studies showing that different mesoscale organizations and scaling regimes often occupy partially overlapping regions of the structural phase space of complex networks, while remaining distinguishable through the combined use of multiple complementary descriptors [7,35].

4.3. Clustering and Path Length

Figures 10 and 11 summarize the distributions of the average clustering coefficient (C) and average shortest-path length (L) across the network families identified by the integrated classification protocol. Because each family includes both supported and ambiguous classifications, these comparisons are intended to illustrate overall structural tendencies rather than establish strict boundaries between topological regimes.

Average clustering coefficients remain high for all network families (Figure 10), indicating that strong local cohesiveness is a common feature of the Embrapa coauthorship clusters. This observation is consistent with the prevalence of small-world characteristics in scientific collaboration networks, where dense local neighborhoods coexist with relatively short global distances [4].

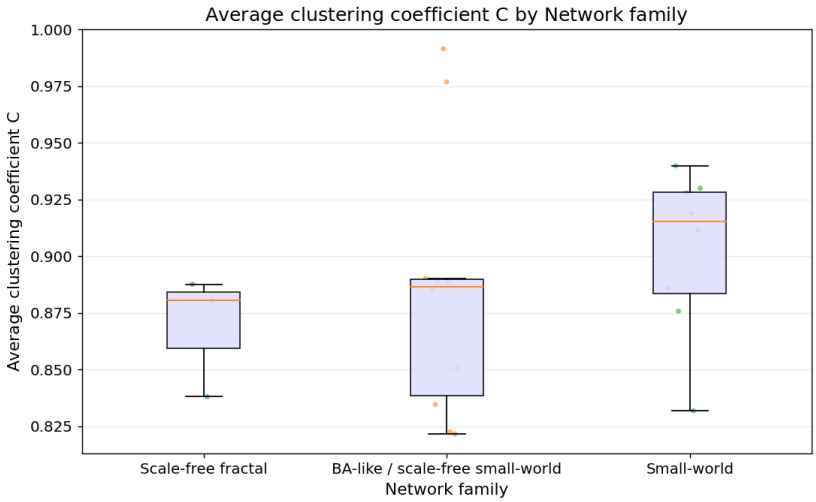


Figure 10. Average clustering coefficient (C) across the network families identified by the integrated classification protocol. Boxes represent the interquartile range (25th–75th percentiles), the horizontal line within each box indicates the median, whiskers show the extent of the data according to the standard boxplot convention, and individual points represent the values for each cluster, with horizontal jitter applied for visualization. High clustering is observed for all families, indicating that strong local cohesiveness is a common structural characteristic of the analyzed collaboration subnetworks.

Differences become more apparent when considering the average path length (Figure 11). Scale-free fractal families generally occupy the upper portion of the observed range, reflecting their more modular and hierarchically organized topology. In contrast, BA-like/scale-free small-world families tend to exhibit shorter path lengths, consistent with the presence of highly connected hubs that facilitate long-range communication across the network [6,7]. Small-world (non-scale-free) families generally present the shortest paths, although some overlap with the other families remains, reflecting the presence of ambiguous classifications within the integrated framework.

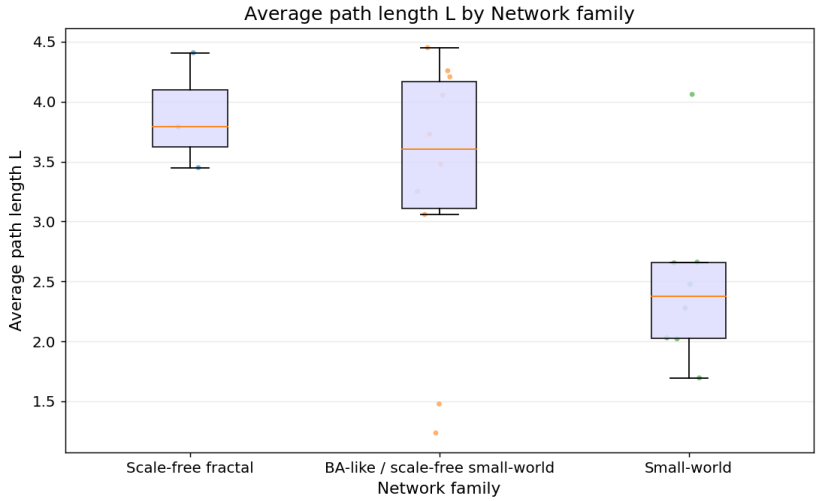


Figure 11. Average shortest-path length (L) across the network families. Boxes represent the interquartile range (25th–75th percentiles), the horizontal line within each box indicates the median, whiskers show the extent of the data according to the standard boxplot convention, and individual points represent the values for each cluster, with horizontal jitter applied for visualization. Scale-free fractal families generally exhibit longer paths, whereas BA-like/scale-free small-world families tend to show shorter paths associated with hub-mediated connectivity. Small-world (non-scale-free) families present the lowest path lengths overall, although partial overlap among families is expected due to ambiguous classifications.

4.4. Betweenness Centrality Tails

Figure 12 presents the CCDF of betweenness centrality for the three network families identified by the integrated classification protocol. Since each family includes both supported and ambiguous classifications, the median curves and interquartile ranges summarize the overall distribution of brokerage patterns within each family rather than defining strict boundaries between topological regimes.

All three families exhibit heavy-tailed behavior, confirming the existence of a relatively small number of highly central nodes. Nevertheless, differences in the extreme tail remain apparent. The BA-like/scale-free small-world family (blue) tends to exhibit a somewhat sharper decay, whereas the scale-free fractal (green) and small-world (non-scale-free) (orange) families generally maintain higher survival probabilities at large betweenness values. These trends suggest that brokerage roles tend to be distributed more broadly among intermediary nodes in the latter families, whereas the BA-like/scale-free small-world family generally concentrates global connectivity in a smaller number of highly influential hubs, consistent with preferential attachment and hub-mediated connectivity [5–7].

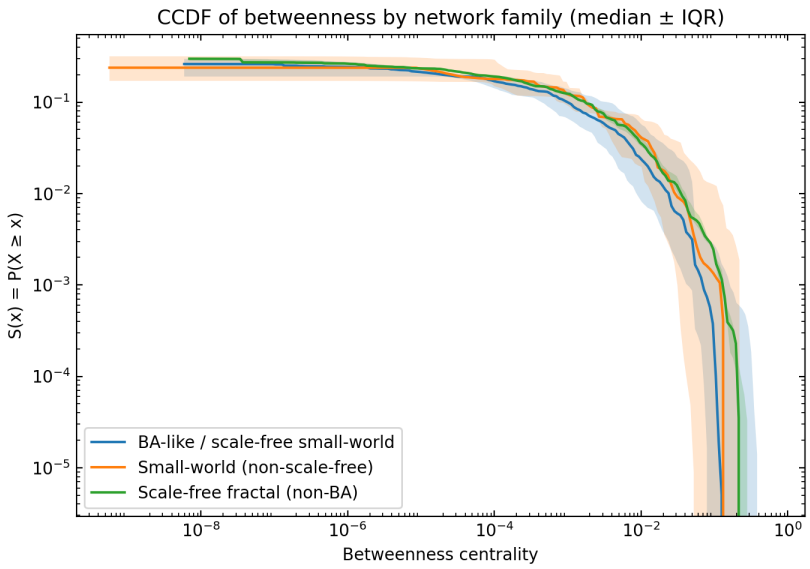


Figure 12. CCDF of betweenness centrality for the network families identified by the integrated classification protocol (median ± IQR across clusters belonging to each family). Because each family contains both supported and ambiguous classifications, the interquartile range reflects the structural variability within each family. Higher survival probabilities at large betweenness indicate a larger fraction of highly intermediary nodes.

This pattern closely parallels observations reported for the worldwide air transportation network, which has also been shown to exhibit scale-free small-world structure while displaying nontrivial centrality organization due to its modular architecture [22]. In that system, highly connected hubs are not necessarily the nodes with the highest betweenness centrality, because shortest paths are often confined within geographically or geopolitically defined communities. As a result, centrality becomes distributed across multiple regional connectors rather than concentrated exclusively in the largest hubs.

A similar mechanism may operate in the present coauthorship network. The BA-like/scale-free small-world family tends to exhibit a sharper decay in the betweenness CCDF, suggesting that global connectivity is generally concentrated in a relatively small number of hub authors, consistent with preferential-attachment-like dynamics. In contrast, the heavier tails observed for the scale-free fractal and small-world (non-scale-free) families indicate a more distributed intermediary structure, in which brokerage roles are shared among multiple authors. Such behavior is consistent with stronger modular organization and heterogeneous connectivity constraints, which redistribute shortest-path flows across several intermediary nodes rather than concentrating them in a single dominant hub. Taken together, these observations reinforce the interpretation that the hybrid scale-free small-world organization identified in this work may emerge from the interplay between preferential attachment and modular organization, rather than from a purely homogeneous Barabási–Albert generative process.

4.5. Distribution of Network Families

The integrated classification protocol partitions the 25 clusters extracted from the giant component into three principal network families–*BA-like/scale-free small-world*, *small-world (non-scale-free)* and *scale-free fractal (non-BA)*–together with a small number of undefined cases. The classification combines complementary evidence from weighted degree-tail modeling, small-world diagnostics (σ), and box-covering analysis, thereby integrating statistical and structural information rather than relying on a single topological descriptor. The spatial distribution of these network families within the giant component reveals a

clear predominance of the BA-like/scale-free small-world family (63.4% of nodes when combining supported and ambiguous classifications), followed by the small-world (non-scale-free) family (27.5%), and the scale-free fractal (non-BA) family (9.0%). Undefined clusters account for only 0.08% of the giant component.

The BA-like/scale-free small-world family dominates the giant component, whereas small-world (non-scale-free) and scale-free fractal families occupy smaller but well-defined portions of the collaboration network. This predominance is consistent with previous studies reporting that mature coauthorship networks frequently develop scale-free small-world characteristics associated with cumulative growth and preferential-attachment-like dynamics [2]. At the same time, the coexistence of multiple network families highlights the structural heterogeneity of the collaboration system, indicating that distinct mesoscopic organizations emerge within the same connected infrastructure.

This heterogeneity is further reflected in the weighted degree-tail models associated with each network family (Table 5). Power-law tails are exclusively observed among the BA-like/scale-free small-world and scale-free fractal families, whereas the small-world (non-scale-free) family is consistently associated with exponential or log-normal tails. These results reinforce that the identified network families correspond to distinct structural organizations rather than arbitrary subdivisions of the giant component, suggesting different underlying mechanisms governing collaboration patterns across the network.

Table 5. Weighted degree-tail models observed within each network family and confidence level.

Network Family	Exponential (EXP)	Log-Normal (LN)	Power-Law (PL)
Supported			
BA-like/scale-free small-world	0	0	5
Scale-free fractal (non-BA)	0	0	3
Ambiguous			
BA-like/scale-free small-world	0	0	3
Small-world (non-scale-free)	8	2	0

Entries indicate the number of clusters assigned to each weighted degree-tail model after application of the integrated classification protocol. Italicized labels indicate the confidence level of the network-family classification: "Supported" indicates agreement among the statistical and structural diagnostics, whereas "Ambiguous" indicates incomplete agreement among these diagnostics.

The distribution of network families reveals that the BA-like/scale-free small-world family accounts for the majority of the giant component, suggesting that cumulative-advantage-like dynamics remain an important organizing mechanism in the collaboration system. The smaller yet well-defined presence of scale-free fractal and small-world (non-scale-free) families, however, indicates that alternative structural organizations coexist within the same connected infrastructure, reflecting distinct collaboration patterns and connectivity constraints across different research communities.

5. Conclusions

This study introduces a multi-criteria methodological framework for identifying network families in large-scale collaboration networks. By integrating community detection, weighted degree-tail modeling, small-world diagnostics, fractal box-covering analysis, and entropy-based measures of degree–degree correlations, we propose a reproducible integrated classification protocol capable of distinguishing heterogeneous mesoscopic organizations within a single giant connected component.

The empirical application to Embrapa's scientific coauthorship network (1974–2024) demonstrates the applicability of the proposed framework to a large real-world collaboration system. The identified network families, *BA-like/scale-free small-world*, *scale-free*

fractal (non-BA), and *small-world (non-scale-free)*, illustrate how complementary statistical and structural diagnostics can be combined into a unified decision protocol rather than interpreted independently. By jointly considering weighted degree-tail selection, small-world diagnostics, fractal scaling, and information-entropy measures, the proposed methodology provides a more comprehensive characterization of mesoscopic network organization than approaches based on individual descriptors.

At the system level, the predominance of the BA-like/scale-free small-world family suggests that cumulative-advantage-like dynamics remain an important organizing mechanism in large collaboration systems. At the same time, the coexistence of scale-free fractal and small-world (non-scale-free) families within the same giant component highlights the intrinsic structural heterogeneity that single-metric analyses often overlook. In this sense, the principal contribution of this work lies less in the specific proportions observed in the Embrapa network than in demonstrating that distinct network families can be systematically identified through an integrated methodological protocol.

The incorporation of entropy-based measures further enriches the characterization by revealing complementary aspects of degree–degree organization that are not captured by degree distributions or small-world metrics alone. Together with the structural diagnostics, these information-entropy measures provide an additional layer for interpreting how hierarchical organization, modularity, and degree correlations vary across the identified network families.

Beyond the empirical case examined here, the proposed framework offers a transferable analytical methodology for investigating collaboration, organizational, and production networks. Its principal strength lies in combining statistically grounded model selection with complementary structural diagnostics within a coherent decision framework, thereby improving reproducibility, interpretability, and cross-domain comparability. Consequently, the methodology is applicable not only to scientific collaboration networks but also to broader sociotechnical systems in which cooperation, information diffusion, resilience, and coordination emerge from centralized or decentralized interaction patterns.

6. Future Research Directions

The framework proposed in this work establishes a reproducible and interpretable methodology for the comparative characterization of mesoscopic structural regimes by integrating complementary statistical and structural diagnostics within a unified decision protocol. By combining multiple independent sources of evidence rather than relying on any single network descriptor, the proposed approach provides a robust foundation for identifying heterogeneous network families while remaining sufficiently flexible to accommodate future methodological refinements and applications. Building upon this foundation, several methodological and application-oriented research directions remain open.

A natural methodological extension consists of broadening both the validation benchmark and the statistical reference models adopted by the framework. Future studies may incorporate additional classes of synthetic networks, including configuration-model networks, stochastic block models, degree-corrected community models, random geometric networks, temporal network generators, and multiplex or multilayer structures. In particular, replacing or complementing the Erdős–Rényi baseline adopted in the present work with degree-preserving configuration-model nulls would provide a more stringent assessment of small-world organization in heterogeneous networks. Such analyses would enable the computation of ensemble-based confidence intervals and z-scores for clustering, path length, and small-worldness, thereby strengthening the statistical foundation of the proposed decision protocol while preserving its interpretability and generality.

Another important direction concerns the incorporation of temporal network analysis and percolation-based order parameters to investigate the emergence, persistence, and transition of network families over time. Extending the current static framework to temporal collaboration networks may reveal critical transitions in mesoscopic organization and improve our understanding of how cumulative growth, institutional changes, and collaboration dynamics influence structural evolution. Coupling this approach with Hamiltonian formulations or energy-landscape approaches may further provide a physics-inspired interpretation of the stability and transitions between different network families.

A particularly promising avenue involves enriching the structural analysis with semantic information extracted from publication metadata. Topic embeddings, domain clustering, and semantic representations assisted by large language models could be integrated with the structural diagnostics developed here to investigate whether specific research domains are preferentially associated with particular network families. Such integration would enable a combined structural-semantic characterization capable of identifying epistemic communities and collaboration patterns that are not apparent from network topology alone.

The transferability of the proposed framework to non-academic systems also represents an important research opportunity. Applying the methodology to production systems, agricultural value chains, agricultural cooperatives, innovation ecosystems, technological districts, and regional development networks may reveal whether the same network families identified in scientific collaboration also emerge in other sociotechnical systems. Comparative analyses across these domains could evaluate whether mesoscopic structural organization exhibits common patterns and whether the identified network families possess predictive value for resilience, innovation capacity, diffusion processes, governance, and organizational performance.

Another particularly relevant direction concerns the role of modular organization in shaping centrality patterns and producing deviations from idealized generative models such as the Barabási–Albert framework. Future studies may investigate, at both macroscopic and microscopic scales, how modularity influences the distribution of betweenness centrality and the emergence of brokerage roles. At the macroscopic level, this includes quantifying the relationship between global modularity and deviations from canonical BA scaling relationships. At the microscopic level, it involves determining whether community boundaries systematically prevent highly connected nodes from becoming dominant brokers. Such analyses may help explain why networks exhibiting similar weighted degree distributions can nevertheless display substantially different patterns of centrality and information flow, providing additional insight into the structural mechanisms underlying the distinct network families identified in this work.

Finally, the current framework may be expanded by incorporating additional structural descriptors, including rich-club organization, core–periphery structure, multilayer connectivity, weighted-edge heterogeneity, motif statistics, higher-order interactions, and persistent topological descriptors. Integrating these complementary diagnostics within the existing decision protocol may further improve classification resolution while preserving the interpretability, reproducibility, and modular design of the methodology.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/complexities2030021/s1: Figure S1: Empirical CCDFs and fitted statistical tail models for Supported BA-like/scale-free small-world clusters; Figure S2: Empirical CCDFs and fitted statistical tail models for Ambiguous BA-like/scale-free small-world clusters; Figure S3: Empirical CCDFs and fitted statistical tail models for Ambiguous Small-World (non-scale-free) clusters; Figure S4: Empirical CCDFs and fitted statistical tail models for Supported scale-free fractal (non-BA) clusters.

Author Contributions: Conceptualization, F.F.C. and I.B.; methodology, F.F.C. and I.B.; software, F.F.C.; validation, F.F.C. and I.B.; formal analysis, F.F.C. and I.B.; investigation, F.F.C. and I.B.; data curation, F.F.C. and I.B.; writing—original draft preparation, F.F.C.; writing—review and editing, F.F.C., I.B., S.M.F.S.M. and J.G.A.B.; visualization, F.F.C.; supervision, I.B. and J.G.A.B.; funding acquisition, J.G.A.B. and S.M.F.S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by FAPESP, Grants 2022/09319-9 and 2024/20662-2.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The anonymized cluster data and the complete source code used in this study, including the scripts for the analyses presented in the Supplementary Materials, are publicly available in the GitLab repository (https://www.gitlab.cnptia.embrapa.br/m327882/giant-component-cluster-analysis) (access date 24 August 2026).

Acknowledgments: The authors acknowledge Luis Eduardo Gonzales and Victor Paulo Marques Simao for providing the data from BDPA. The authors are also grateful to the three anonymous reviewers for their careful reading of the manuscript and for their constructive comments and insightful suggestions, which substantially improved the quality, clarity, and robustness of this work.

Conflicts of Interest: Author Fabiane de Fatima Carvalho holds a FAPESP postdoctoral position at Embrapa Digital Agriculture. Authors Ivan Bergier and Jayme Garcia Arnal Barbedo are employed by Embrapa Digital Agriculture, a research unit of Embrapa. Author Silvia Maria Fonseca Silveira Massruhá is employed by Embrapa.

Abbreviations

LCC	largest connected component
GCC	giant connected component; same as LCC when referring to clusters
n	number of nodes in a cluster
m	number of edges in a cluster
ρ	graph density
C	average clustering coefficient
T	global transitivity (global clustering coefficient)
L	average shortest-path length
r	degree assortativity coefficient (Newman)
σ_{SW}	small-worldness index
$C_{\text{rand}}, L_{\text{rand}}$	clustering coefficient and path length of an Erdős–Rényi n, m. random graph with same
k	node degree
$P(k)$	degree distribution
CCDF	complementary cumulative distribution function, $P(X \geq x)$
$x_{\min}$	minimum degree defining the tail used in power-law fitting
n_{tail}	number of observations in the tail ($k \geq x_{\min}$)
α	power-law exponent from MLE
μ, σ	log-normal parameters (in logarithmic space)
λ_{exp}	exponential rate parameter (EXP model)
KS	Kolmogorov–Smirnov statistic for model comparison
p_{KS}	KS bootstrap p-value for power-law plausibility
$\hat{L}$	maximized likelihood of a fitted model
AIC	Akaike Information Criterion

$AIC_{PL}, AIC_{exp}, AIC_{ln}$	AIC values for power-law, exponential and log-normal fits
ΔAIC	$AIC_{exp} - AIC_{PL}$; > 0 favors fractal (PL), < 0 favors exponential (EXP)
ℓ_B	box diameter (in hops)
$N_B(\ell_B)$	minimum number of boxes of diameter ℓ_B to cover the LCC
d_B	fractal (box) dimension from $N_B(\ell_B) \sim \ell_B^{-d_B}$
a	exponential decay rate from $N_B(\ell_B) \sim e^{-a\ell_B}$
$R^2_{\text{log-log}}$	determination coefficient of the log-log regression for fractal fits
$M_B(\ell_B)$	mean box mass; average number of nodes per box in renormalization
β	growth exponent from temporal evolution of node degree, $k_i(t) \sim t^{\beta}$
fitness (η_i)	intrinsic competitiveness in the Bianconi–Barabási model
H_k	Shannon entropy of the degree distribution, $H_k = -\sum_k P(k) \log_2 P(k)$
e_{jk}	joint degree distribution at the two ends of edges
q_k	edge-end marginal distribution, $q_k = \sum_j e_{jk}$
H_{edge}	edge joint entropy
$H(K_2\|K_1)$	conditional entropy of degree mixing
$I(K_1; K_2)$	mutual information of degree correlations, $I = \sum_{jk} e_{jk} \log_2 \left(e_{jk}/(q_j q_k)\right)$
Q	modularity of a partition (Louvain)
c_i	community assignment of node i
$\delta(c_i, c_j)$	Kronecker delta indicating whether two nodes share the same community
Louvain	modularity-maximizing community detection algorithm
OpenOrd	force-directed layout optimized for large modular networks
FR	Fruchterman–Reingold force-directed layout

References

1. Newman, M.E.J. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 404–409. [CrossRef] [PubMed]
2. Barabási, A.L.; Jeong, H.; Néda, Z.; Ravasz, E.; Schubert, A.; Vicsek, T. Evolution of the social network of scientific collaborations. *Phys. A* **2002**, *311*, 590–614. [CrossRef]
3. da F. Costa, L.; Rodrigues, F.A.; Travieso, G.; Villas Boas, P.R. Characterization of complex networks: A survey of measurements. *Adv. Phys.* **2007**, *56*, 167–242. [CrossRef]
4. Watts, D.J.; Strogatz, S.H. Collective dynamics of 'small-world' networks. *Nature* **1998**, *393*, 440–442. [CrossRef] [PubMed]
5. Newman, M.E.J. Assortative mixing in networks. *Phys. Rev. Lett.* **2002**, *89*, 208701. [CrossRef] [PubMed]
6. Song, C.; Havlin, S.; Makse, H.A. Self-similarity of complex networks. *Nature* **2005**, *433*, 392–395. [CrossRef] [PubMed]
7. Rozenfeld, H.D.; Song, C.; Makse, H.A. Small-world to fractal transition in complex networks: A renormalization group approach. *Phys. Rev. Lett.* **2010**, *104*, 025701. [CrossRef] [PubMed]
8. Bergier, I.; Santos, P.M.; Oster, A.H. Scientific collaboration in a multidisciplinary organization revealed by network science. *SN Comput. Sci.* **2021**, *2*, 393. [CrossRef]
9. Wang, X.; Qin, X. Asymmetric intimacy and algorithm for detecting communities in bipartite networks. *Phys. A Stat. Mech. Its Appl.* **2016**, *462*, 569–578. [CrossRef]
10. Alimohammad, M.; Hosseini, S.J.F.; Mirdamadi, S.M.; Dehyouri, S. Collaborative networking among agricultural production cooperatives in Iran. *Heliyon* **2022**, *8*, e11846. [CrossRef] [PubMed]
11. Kraft, T.S.; Cummings, D.K.; Venkataraman, V.V.; Alami, S.; Beheim, B.; Hooper, P.; Seabright, E.; Trumble, B.C.; Stieglitz, J.; Kaplan, H.; et al. Female cooperative labour networks in hunter–gatherers and horticulturalists. *Philos. Trans. R. Soc. B Biol. Sci.* **2023**, *378*, 20210431. [CrossRef] [PubMed]

12. Codjoe, F.N.Y.; Blundo-Canto, G.; Mathe, S.; Soullier, G.; Asante, F.A.; Sarpong, D.B. Proximity and information flows in farmer networks: A comparative analysis of certified and non-certified cooperative networks in Ghana's cocoa sector. *J. Agric. Educ. Ext.* **2026**, *32*, 221–249. [CrossRef]
13. Colleran, H. Market integration reduces kin density in women's ego-networks in rural Poland. *Nat. Commun.* **2020**, *11*, 266. [CrossRef] [PubMed]
14. Zhu, Z.; Gao, C.; Zhang, Y.; Li, H.; Xu, J.; Zan, Y.; Li, Z. Cooperation and Competition among information on social networks. *Sci. Rep.* **2020**, *10*, 12160. [CrossRef] [PubMed]
15. Blondel, V.D.; Guillaume, J.L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008. [CrossRef]
16. Humphries, M.D.; Gurney, K. Network 'Small-World-ness': A Quantitative Method for Determining Canonical Network Equivalence. *PLoS ONE* **2008**, *3*, e0002051. [CrossRef] [PubMed]
17. Clauset, A.; Shalizi, C.R.; Newman, M.E.J. Power-law distributions in empirical data. *Siam Rev.* **2009**, *51*, 661–703. [CrossRef]
18. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]
19. Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed.; Springer: New York, NY, USA, 2002; p. 488. [CrossRef]
20. Barabási, A.L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512. [CrossRef] [PubMed]
21. Martin, S.; Brown, W.M.; Klavans, R.; Boyack, K.W. OpenOrd: An open-source toolbox for large graph layout. In *Proceedings of the Visualization and Data Analysis 2011*; SPIE: Washington, DC, USA 2011; p. 786806. [CrossRef]
22. Guimerà, R.; Mossa, S.; Turtschi, A.; Amaral, L.A.N. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 7794–7799. [CrossRef] [PubMed]
23. Bianconi, G.; Barabási, A.L. Competition and multiscaling in evolving networks. *Phys. Rev. Lett.* **2001**, *86*, 5632–5635. [CrossRef] [PubMed]
24. Jeong, H.; Néda, Z.; Barabási, A.L. Measuring preferential attachment in evolving networks. *EPL (Europhys. Lett.)* **2003**, *61*, 567–572. [CrossRef]
25. Wen, T.; Cheong, K.H. The fractal dimension of complex networks: A review. *Inf. Fusion* **2021**, *73*, 87–102. [CrossRef]
26. Fronczak, A.; Fronczak, P.; Samsel, M.J.; Makulski, K.; Łepek, M.; Mrowinski, M.J. Scaling theory of fractal complex networks. *Sci. Rep.* **2024**, *14*, 9079. [CrossRef] [PubMed]
27. Gallos, L.K.; Song, C.; Makse, H.A. A review of fractality and self-similarity in complex networks. *Phys. A Stat. Mech. Its Appl.* **2007**, *386*, 686–691. [CrossRef]
28. Newman, M.E.J. Mixing patterns in networks. *Phys. Rev. E* **2003**, *67*, 026126. [CrossRef] [PubMed]
29. Anand, K.; Bianconi, G.; Severini, S. Shannon and von Neumann entropy of random networks with heterogeneous expected degree. *Phys. Rev. E* **2011**, *83*, 036109. [CrossRef] [PubMed]
30. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.
31. Gallos, L.K.; Song, C.; Makse, H.A. Scaling of degree correlations and its influence on diffusion in scale-free networks. *Phys. Rev. Lett.* **2008**, *100*, 248701. [CrossRef] [PubMed]
32. Song, C.; Havlin, S.; Makse, H.A. Origins of fractality in the growth of complex networks. *Nat. Phys.* **2006**, *2*, 275–281. [CrossRef]
33. Ravasz, E.; Somera, A.L.; Mongru, D.A.; Oltvai, Z.N.; Barabási, A.L. Hierarchical organization of modularity in metabolic networks. *Science* **2002**, *297*, 1551–1555. [CrossRef] [PubMed]
34. Albert, R.; Barabási, A.L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2002**, *74*, 47–97. [CrossRef]
35. Wei, Z.W.; Wang, B.H. Emergence of fractal scaling in complex networks. *Phys. Rev. E* **2016**, *94*, 032309. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.