Feature-enriched hyperbolic network geometry

Roya Aliakbarisani [1,2,*] M. Ángeles Serrano [1,2,3,†] and Marián Boguñá [1,2,‡]
[1]Departament de Física de la Matèria Condensada, Universitat de Barcelona, Martí i Franquès 1, E-08028 Barcelona, Spain
[2]Universitat de Barcelona Institute of Complex Systems (UBICS), Barcelona, Spain
[3]Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, E-08010 Barcelona, Spain

(Received 6 July 2023; accepted 6 April 2025; published 8 July 2025)

Graph-structured data provide an integrated description of complex systems, encompassing not only the interactions among nodes, but also the intrinsic features that characterize these nodes. These features play a fundamental role in the formation of links within the network, making them valuable for extracting meaningful information. In this paper, we present a comprehensive framework that treats features as tangible entities and establishes a bipartite graph connecting nodes and features. By assuming that nodes sharing similarities should also share features, we introduce a hyperbolic geometric space where both nodes and features coexist, shaping the structure of both the node network and the bipartite network of nodes and features. Through this framework, we can identify correlations between nodes and features in real data and generate synthetic datasets that mimic the topological properties of their connectivity patterns. Notably, node features are at the core of deep learning techniques, such as graph convolutional neural networks (GCNs), offering great utility in downstream tasks. Therefore, our approach provides insights into the inner workings of GCNs by revealing the intricate structure of the data.

DOI: 10.1103/PhysRevResearch.7.033036

I. INTRODUCTION

The formation of links in complex networks has been a central topic in network science research over the past two decades. Understanding the mechanisms that drive interactions between individual elements is crucial for unraveling the behavior of the entire system. A key approach to this challenge is network geometry [1], which embeds networks in metric spaces to capture their underlying topological properties [2–4].

While traditional mechanisms of link formation, such as preferential attachment [5], focused solely on topological structure, they often overlooked the role of node features, which are also essential to the link formation process. With the increasing availability of graph-structured data—networks enriched with annotated features—it has become possible to analyze how features influence network structure, leading to insights into community detection [6–10] and percolation properties [11]. A critical open question remains: How can we effectively model not just network topology but also the influence of features on link formation?

This challenge is particularly relevant in the context of graph convolutional networks (GCNs), which have become essential tools for analyzing graph-structured data. GCNs extend classical convolutional neural networks (CNNs) to irregular, graph-like data, allowing them to capture both local and global structures in a graph by aggregating information from neighboring nodes [12–26]. These capabilities make GCNs effective for tasks such as node classification, link prediction, and graph classification. However, like many machine learning models, GCNs have been criticized for their lack of interpretability [27].

An implicit assumption in GCNs is that connected nodes are likely to share similar features. This assumption underscores the need to first understand the structure of the underlying data before improving model explainability. Without a clear understanding of the correlations between network topology and node features, it remains difficult to explain why certain predictions are made by GCNs. Recent work in the field of GCNs has explored some methodologies to enhance their explainability. For instance, attribution methods enhance the interpretability of GCNs by highlighting which parts of the input data are most influential in the decision-making process of the model [28]. Nevertheless, a deeper understanding of the underlying data is essential for advancing the explainability of these models.

In this paper, we introduce a novel framework based on hyperbolic geometry to model real-world graph-structured data. Hyperbolic geometry naturally supports scale-free networks and effectively embeds high-dimensional hierarchical structures into lower-dimensional space. This geometry aligns well with the hierarchical, clustered, scale-free, and small-world characteristics often observed in real-world networks, offering robust modeling and interpretation. Our approach has two critical contributions. First, we consider features as real entities

*Contact author: roya_aliakbarisani@ub.edu
†Contact author: marian.serrano@ub.edu
‡Contact author: marian.boguna@ub.edu

that define a bipartite graph of nodes connected to features which, for real systems, shows a complex topological organization. This builds on the understanding that node features are crucial in determining network structure [29,30]. Second, we propose a hyperbolic embedding where both nodes and features coexist in the same geometric space. In this framework, nodes are similar if they share features, and features are similar if they share nodes. This framework allows us to capture the correlations between nodes and features in real data and to generate synthetic datasets that accurately replicate the topological properties of real networks. The proposed model marks a preliminary step in a promising research direction, establishing the foundation for a future feature-aware embedding method in hyperbolic space, which could improve the interpretability of GNNs.

II. RESULTS

A typical graph-structured dataset consists of a set of N_n nodes forming a complex network $\mathcal{G}_n$ and a set of N_f features associated with the same set of nodes. The features are usually binarized, so the set of features for a given node i is represented as a vector $\vec{f}_i \in \{0,1\}^{N_f}$ with entries of zero or one, indicating the presence or absence of a particular feature. For example, the Cora dataset is a standard benchmark used in GCN studies. It is defined by a citation network among scientific publications—or nodes—and each publication is characterized by a vector, where the entries indicate the presence or absence of specific words—or features—from a unique dictionary.

A. Detecting correlations between nodes and features

As mentioned earlier, for GCNs to be effective, there must be a correlation between the features of nodes and the underlying graph topology $\mathcal{G}_n$. Therefore, it is essential, not only to prove that, indeed, this correlation is present in real-world datasets, but also that it has a geometric nature.

As a preliminary step to measure the correlation between topology and features, we can compute the cosine similarity of binary feature vectors assigned to each node in $\mathcal{G}_n$. Specifically, we calculate the average cosine similarity between connected pairs of nodes in $\mathcal{G}_n$ and compare it to the average cosine similarity of an equally sized set of node pairs that are unconnected in $\mathcal{G}_n$, which is equivalent to the set of connected pairs in its complementary graph $\overline{\mathcal{G}}_n$. The complementary graph $\overline{\mathcal{G}}_n$ is generated using the same set of vertices as $\mathcal{G}_n$, with two nodes being connected if and only if they are unconnected in $\mathcal{G}_n$. As a result, the correlation between the nodes' features and the network structure is given by

$$\text{corr} = \frac{\overline{CS}(\mathcal{G}_n) - \overline{CS}(\overline{\mathcal{G}}_n)}{\overline{CS}(\overline{\mathcal{G}}_n)}, \tag{1}$$

where $\overline{CS}(\mathcal{G}_n)$ represents the average cosine similarity in $\mathcal{G}_n$ and $\overline{CS}(\overline{\mathcal{G}}_n)$ represents the same quantity for randomly selected sets of connected node pairs in the complementary graph $\overline{\mathcal{G}}_n$.

Table I presents the correlations in the real dataset, averaged over 100 different randomly selected sets of unconnected node pairs in $\mathcal{G}_n$, along with their standard deviations. These results indicate a clear correlation between the nodes and their associated features.

TABLE I. Parameters of the bipartite network $\mathcal{G}_{n,f}$ for the analyzed datasets. Parameter β for $\mathcal{G}_n$ is directly inferred by Mercator and corr represents the correlation between nodes' features and the network structure.

Network	N_n	N_f	$\langle k_n \rangle$	$\langle k_f \rangle$	β_b	β	corr
Cora	2708	1432	18.17	34.37	0.9	1.6	2.02 ± 0.05
Facebook	12374	3720	7.54	25.09	2.0	1.7	8.86 ± 0.09
Citeseer	3264	3703	31.75	27.98	0.9	1.6	3.32 ± 0.06
Chameleon	2277	3132	21.55	15.66	1.0	1.6	2.53 ± 0.03

B. Modeling the network: The $\mathbb{S}^1/\mathbb{H}^2$ model

To fully characterize such complex graph-structured data, we must first understand the complex network $\mathcal{G}_n$ that defines the relationships between nodes. In CNNs applied to images, for instance, this network is defined by the nearest neighbors in a two-dimensional grid of pixels. However, in complex graph-structured data, the relationships between nodes are better described by a complex network with intricate topological properties. Our research over the last decade has shown that complex networks, such as the ones of interest in this context, can be accurately characterized using geometric random graph models [1]. Random graph models help in generating synthetic networks that mimic the structural properties of real-world networks, enabling researchers to study and understand these properties in a controlled manner [31]. Moreover, they can serve as null models by randomizing networks to have some of the same properties as real networks, thereby statistically assessing the existence or magnitude of properties of interest in empirical networks [32,33].

In geometric random graph models, nodes are positioned in a metric space, and the probability of connection between nodes depends on their distances in this space. This approach has led to the emergence of network geometry as a field, providing a comprehensive understanding of real complex networks. Geometric models in a latent hyperbolic metric space have proven effective in generating networks with realistic topological properties, including heterogeneous degree distributions [2,3,34], clustering [3,34–36], small-worldness [37–39], percolation [40,41], spectral properties [42], and self-similarity [2]. They have also been extended to encompass growing networks [4], weighted networks [43], bipartite networks [44,45], complementarity driven networks [46], multilayer networks [47,48], networks with community structure [49–51], and serve as the basis for defining a renormalization group for complex networks [52,53]. In this case, this approach is particularly interesting as it naturally introduces the concept of an underlying similarity space, allowing the unambiguous quantification of similarity between nodes.

To describe the network between nodes $\mathcal{G}_n$, we employ the $\mathbb{S}^1$ model, also known as the geometric soft configuration model [2,3,54,55]. The $\mathbb{S}^1$ model plays a crucial role in network geometry by mapping nodes onto a one-dimensional circle in a way that naturally encodes their similarity

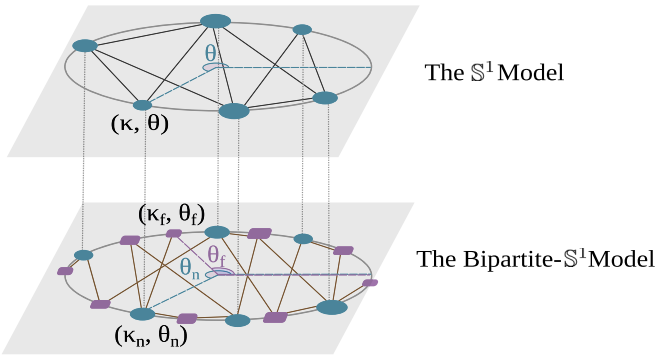

FIG. 1. Illustration of the bipartite model used to describe graph structured data. The sketch at the top depicts the generation of a network using the $\mathbb{S}^1$ model. The bottom part illustrates the generation of the bipartite network between nodes (green circles), keeping the same angular coordinates, and features (rounded purple squares).

regardless of their popularity. Its elegance lies in capturing the hidden geometric constraints that shape how nodes connect, providing a low-dimensional yet powerful representation of complex networks. By assigning angular positions to reflect how similar or related nodes are, the $\mathbb{S}^1$ model reveals underlying community structures that might otherwise remain obscured. This model not only offers deeper theoretical insights into the geometric nature of network connectivity but also facilitates practical applications, such as enhanced link prediction and community detection, to name just a few.

Specifically, in the $\mathbb{S}^1$ model, each node is assigned two hidden variables (κ, θ), where κ determines its expected degree, quantifying its popularity within the network, and θ encodes similarity for all attributes of the node (apart from its degree) representing its position on a one-dimensional sphere of radius $R = N_n/2\pi$. This sphere represents the abstraction of the similarity space where nodes are placed [56]. The connection probability between two nodes with hidden variables (κ, θ) and (κ', θ') is defined as follows:

$$p(\kappa, \kappa', \Delta\theta) = \frac{1}{1+\chi^\beta} \quad \text{with} \quad \chi \equiv \frac{R\Delta\theta}{\mu\kappa\kappa'}, \tag{2}$$

where $\Delta\theta = \pi - |\pi - |\theta - \theta'||$ represents the angular separation between the nodes, $\mu = \frac{\beta}{2\pi\langle k\rangle}\sin\frac{\pi}{\beta}$ is a parameter that fixes the average degree $\langle k\rangle$ (see the top part of Fig. 1), and $\beta > 1$ [57] is the inverse of the temperature of the graph ensemble and determines the level of clustering in the network. Note that clustering refers here to transitive relations represented as triangles in the network's topology. The hidden variables of the nodes can either be generated from an arbitrary probability density $\rho(\kappa, \theta)$ if the goal is to create synthetic networks (see Algorithm 1) or can be inferred from a real network by maximizing the likelihood of the model to reproduce the desired real network. In this work, we use the latter approach through the embedding tool called MERCATOR [58].

Interestingly, the $\mathbb{S}^1$ model is isomorphic to a purely geometric model in the hyperbolic plane, the $\mathbb{H}^2$ model [3]. By mapping the expected degree of a given node to a radial coordinate as

$$r = R_{\mathbb{H}^2} - 2\ln\frac{\kappa}{\kappa_0} \quad \text{with} \quad R_{\mathbb{H}^2} = 2\ln\frac{2R}{\mu\kappa_0^2}, \tag{3}$$

and $\kappa \geqslant \kappa_0$, the connection probability Eq. (2) becomes

$$p(x) = \frac{1}{1+e^{\frac{\beta}{2}(x-R_{\mathbb{H}^2})}} \quad \text{with} \quad x = r + r' + 2\ln\frac{\Delta\theta}{2}, \tag{4}$$

and where x is a very good approximation of the hyperbolic distance between two points at radial coordinates r and r' and separated by an angular distance $\Delta\theta$. That is, pairs of nodes are connected with a probability that depends solely on their hyperbolic distance, simultaneously capturing both the popularity and similarity dimensions. Thanks to this equivalence, we can use either version of the model depending on the particular application. The $\mathbb{S}^1$ model is more convenient for performing analytic calculations, whereas the $\mathbb{H}^2$ model is more suited for tasks like greedy routing [59] and for visualization purposes.

One key advantage of using the $\mathbb{H}^2$ model is that it provides a continuous, two-dimensional hyperbolic space in which both the popularity (radial coordinate) and similarity (angular coordinate) of nodes can be represented naturally. This fully geometric model often yields a more flexible description of real-world networks for tasks like visualization and navigation. Additionally, by allowing the radial dimension to reflect node "popularity" or centrality, the $\mathbb{H}^2$ model captures hierarchy and scale in a purely geometric way.

C. Geometric model of nodes and features: The bipartite-$\mathbb{S}^1/\mathbb{H}^2$ model

The key aspect of our approach is to view the set of nodes and their features as a bipartite graph $\mathcal{G}_{n,f}$. While bipartite graphs have been previously considered in the computer science and network science literature [44,45,60–62], to the best of our knowledge, although a few models considered nodes and features as elements of a bipartite graph [63–65], their topological properties have not been measured. In this representation, each node has a degree k_n that indicates the number of distinct features it possesses, while each feature has a degree k_f that represents the number of connected nodes. The top row of Fig. 2 displays the complementary cumulative distribution function of node and feature degrees for the Cora and Facebook datasets. Across all the datasets we examined, we observed a consistent pattern characterized by a homogeneous distribution of node degrees and a heterogeneous distribution of feature degrees in the bipartite graph $\mathcal{G}_{n,f}$. The insets in these plots also reveal weak correlations between the degrees k_n and k_f of connected pairs for the Cora dataset and negative correlations in the case of the Facebook dataset.

Our objective is to develop a model for this bipartite graph that is correlated with the node network $\mathcal{G}_n$. To achieve this, we propose a geometric model called the bipartite-$\mathbb{S}^1$ model [44,45], where the similarity space is shared between $\mathcal{G}_n$ and $\mathcal{G}_{n,f}$, as suggested by the empirical correlation found in Table I. In this model, each node is assigned two hidden variables (κ_n, θ_n), where κ_n represents its expected degree in the bipartite graph, and the angular coordinate corresponds to that of $\mathcal{G}_n$, i.e., $\theta_n = \theta$. Similarly, features are equipped

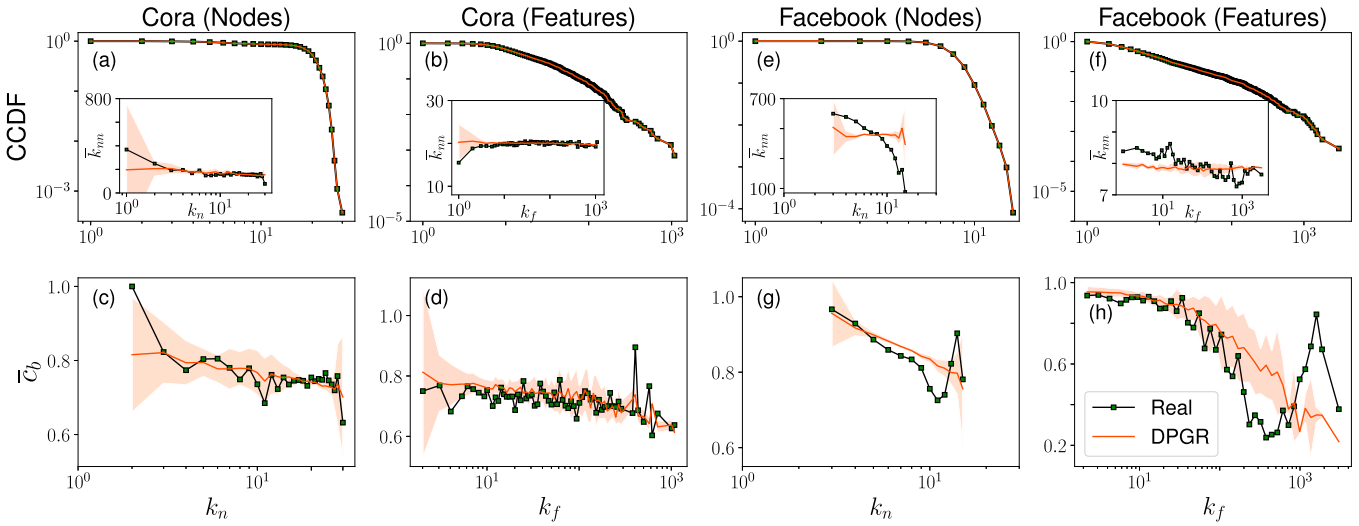

FIG. 2. Topological properties of graph-structured data. Topological properties of $\mathcal{G}_{n,f}$ for the Cora and Facebook datasets (symbols) and their synthetic counterparts generated by the bipartite-$\mathbb{S}^1$ model with the DPGR algorithm in Eq. (6) (red solid lines). The top row (a), (b), (e), and (f) shows the complementary cumulative distribution functions of nodes and features degrees, whereas the insets in these plots show the average nearest-neighbors degree functions $\bar{k}_{nn}$. These functions are defined, for features, as the average degree of nodes in $\mathcal{G}_{n,f}$, k_n, that are neighbors of features of degree k_f, and similarly for nodes. The bottom row (c), (d), (g), and (h) shows the bipartite clustering spectrum of nodes and features as a function of nodes and features degrees, respectively. The orange shaded area represents two-σ intervals of the ensemble. Exponential binning is applied in the computation of $\bar{k}_{nn}$ and $\bar{c}_b$ for the features.

with two hidden variables (κ_f, θ_f), indicating their expected degrees and angular positions in the common similarity space. The probability of a connection between a node and a feature with hidden degrees κ_n and κ_f, separated by an angular distance $\Delta\theta$, is given by

$$p_b(\kappa_n, \kappa_f, \Delta\theta) = \frac{1}{1+\chi^{\beta_b}} \quad \text{with} \quad \chi \equiv \frac{R\Delta\theta}{\mu_b \kappa_n \kappa_f}, \tag{5}$$

where $\mu_b = \frac{\beta_b}{2\pi\langle k_n\rangle} \sin\frac{\pi}{\beta_b}$ is a parameter determining the average degree of nodes $\langle k_n\rangle$ and features $\langle k_f\rangle = \frac{N_n}{N_f}\langle k_n\rangle$ (the sketch in Fig. 1 illustrates the construction of the model). Similar to the $\mathbb{S}^1$ model, this choice ensures that the expected degrees of nodes and features with hidden degrees κ_n and κ_f are $\langle k_n(\kappa_n)\rangle = \kappa_n$ and $\langle k_f(\kappa_f)\rangle = \kappa_f$, respectively [44,45]. The hidden variables of nodes and features can be generated from arbitrary distributions (see Algorithm 2) or fitted to replicate the topology of a real network of interest.

In the latter case, following the approach in [66], it is also possible to define the "microcanonical" ensemble of the model, by using a degree-preserving geometric randomization (DPGR) Metropolis-Hastings algorithm. This algorithm allows us to explore different values of β_b while exactly preserving the degree sequences. Given a network and after assigning angular coordinates at random to all nodes and features, the algorithm randomly selects a pair of node-feature links $i_n - j_f$ and $l_n - m_f$, and swaps them (avoiding multiple connections) with a probability given by

$$p_{\text{swap}} = \min\left[1, \left(\frac{\Delta\theta_{i_n j_f}\Delta\theta_{l_n m_f}}{\Delta\theta_{i_n m_f}\Delta\theta_{j_f l_n}}\right)^{\beta_b}\right], \tag{6}$$

where $\Delta\theta$ is the angular separation between the corresponding pair of nodes. This algorithm maximizes the likelihood that the network is generated by the bipartite-$\mathbb{S}^1$ model, while preserving the degree sequence and the set of angular coordinates. Notice that $\beta_b = 0$ corresponds to the bipartite configuration model.

Similarly to the $\mathbb{S}^1$ model, the bipartite-$\mathbb{S}^1$ model can also be mapped to the hyperbolic plane. In this case, the mapping is given by

$$r_n = R^b_{\mathbb{H}^2} - 2\ln\frac{\kappa_n}{\kappa_{n,0}} \quad \text{and} \quad r_f = R^b_{\mathbb{H}^2} - 2\ln\frac{\kappa_f}{\kappa_{f,0}}, \tag{7}$$

where the radius of the hyperbolic disk is $R^b_{\mathbb{H}^2} = 2\ln\frac{2R}{\mu_b\kappa_{n,0}\kappa_{f,0}}$, and $\kappa_{n,0}$ and $\kappa_{f,0}$ are the minimum values of the expected nodes and features degrees, respectively. With this mapping, the connection probability between a node and a feature depends only on the hyperbolic distance between them and takes the same form as in Eq. (4).

D. Geometry-driven nodes and features correlations

To highlight the geometric nature of nodes and features correlation, we define a new unipartite network between nodes, called $\hat{\mathcal{G}}_n$, where two nodes are connected if they share a significant number of features (for technical details, refer to the Sec. II of the Supplementary Material [67]). It is important to note that the links in the networks $\mathcal{G}_n$ and $\hat{\mathcal{G}}_n$ are defined by different connection mechanisms so that, *a priori*, they could be totally unrelated. To measure any possible correlation between them, we assume that $\hat{\mathcal{G}}_n$ also follows the $\mathbb{S}^1$ model. Subsequently, the angular coordinates of nodes from $\mathcal{G}_n$ are inferred using MERCATOR [58], and these coordinates are then employed as initial estimates to infer the angular coordinates of nodes from $\hat{\mathcal{G}}_n$, again using MERCATOR. The outcomes are depicted in Fig. 3, showcasing the results obtained from the Cora and Facebook datasets (additional information about

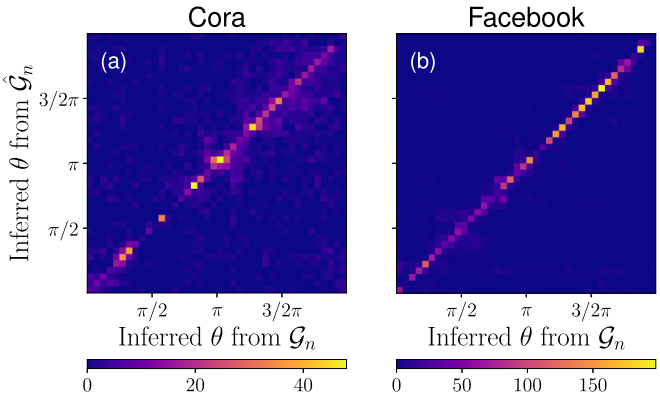

FIG. 3. Detecting correlations between the spaces of nodes and features. Heatmap of the angular coordinates of nodes inferred by Mercator from $\mathcal{G}_n$ (in the x-axis) and from $\hat{\mathcal{G}}_n$ (in the y-axis) for the (a) Cora and (b) Facebook datasets. A detailed description of these datasets is provided in the Supplementary Material (Sec. I) [67]. Color indicates the number of nodes in each pixel.

these datasets can be found in the Sec. I of the Supplementary Material [67]). The figure clearly illustrates a significant correlation between angular coordinates of nodes determined exclusively from topology and those determined exclusively from features. In contrast, randomized versions of $\hat{\mathcal{G}}_n$, which maintain the degree distribution and clustering coefficient, do not exhibit this correlation, even though they are initialized using the coordinates of $\mathcal{G}_n$ (see Sec. II in the Supplementary Material [67]). This empirical evidence strongly suggests that the similarity space of nodes and features is highly congruent.

E. Quantifying bipartite clustering in real datasets

In the $\mathbb{S}^1$ model, the parameter β governs the clustering coefficient and thus influences the relationship between the network topology and the underlying metric space. Similarly, the parameter β_b accounts for the coupling between the bipartite graph $\mathcal{G}_{n,f}$ and the underlying metric space. As both $\mathcal{G}_n$ and $\mathcal{G}_{n,f}$ are defined on the same underlying metric space, the parameters β and β_b control the correlation between them. It is therefore important to measure the value of β_b for a real dataset. To achieve this, we use the simplest possible extension of the clustering coefficient to bipartite networks, denoted as $\bar{c}_b$, as explained in Fig. 4.

In bipartite networks, $\bar{c}_b$ is strongly influenced by the heterogeneity of the features' degree distribution and, for finite-sized networks, it can reach high values even in the configuration model (see Sec. III in the Supplementary Material [67]). Thus, measuring β_b by adjusting $\bar{c}_b$ can be misleading and we took a different approach.

We sorted the degrees of features in decreasing order and removed 2^l of the highest degree features from the original network $\mathcal{G}_{n,f}$, starting from the highest degree, where $l = 0, 1, 2, \ldots$. After each removal, we measured the bipartite clustering coefficient $\bar{c}_b(l)$ of the remaining network and the fluctuations in features' degrees as $\langle k_f(k_f-1)\rangle/\langle k_f\rangle$ (see Supplementary Material for details [67]). Figure 5 illustrates the behavior of the bipartite clustering for the Cora and Facebook datasets, considering values of l up to $l_{\max} = 8$. We

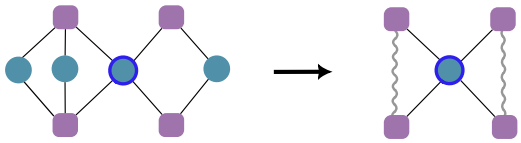

FIG. 4. Bipartite clustering. This diagram demonstrates the approach used to measure bipartite clustering. In this example, the node at the center is connected to four different features. Two features are considered connected if they share at least a common node other than the central node. The bipartite clustering of the node is then calculated by determining the fraction of connected pairs of features, following the standard definition of clustering coefficient in unipartite networks. The same definition applies to the bipartite clustering coefficient of features.

repeated this procedure for networks generated by our model with the DPGR algorithm Eq. (6) and different values of β_b. Interestingly, for real networks, the bipartite clustering coefficient $\bar{c}_b(l)$ decreases slowly as hubs are removed (going from right to left in the figure). On the other hand, in the configuration model with $\beta_b = 0$, $\bar{c}_b(l)$ decreases rapidly when the heterogeneity of the degree sequence is eliminated, even if the original network exhibits similar values to the real networks. As we increase the value of β_b, we observed that our model can accurately replicate the behavior of $\bar{c}_b(l)$, enabling us to estimate the values of β_b in real networks. Beyond the practical estimation of parameter β_b, the slow decay of clustering when removing hubs provides strong empirical evidence that the bipartite network between nodes and features is governed by an underlying similarity metric space.

Table I presents the properties of the analyzed real networks and the inferred values of β and β_b. Using these

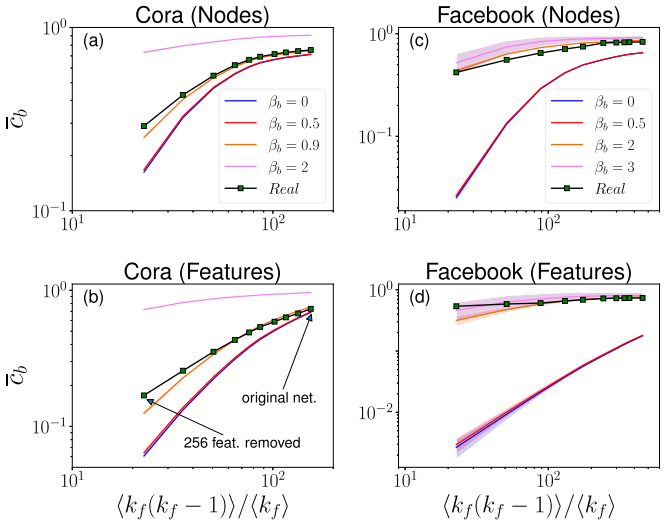

FIG. 5. Behavior of bipartite clustering upon removal of hubs. Bipartite clustering coefficient for the Cora and Facebook networks (symbols) and their surrogates generated by our model with different values of β_b (solid lines). The plots show the bipartite clustering of the networks obtained by removal of a number of the highest degree features as a function of the corresponding fluctuations of features' degrees. In all plots, solid lines represent averages over 100 synthetic networks generated by our model.

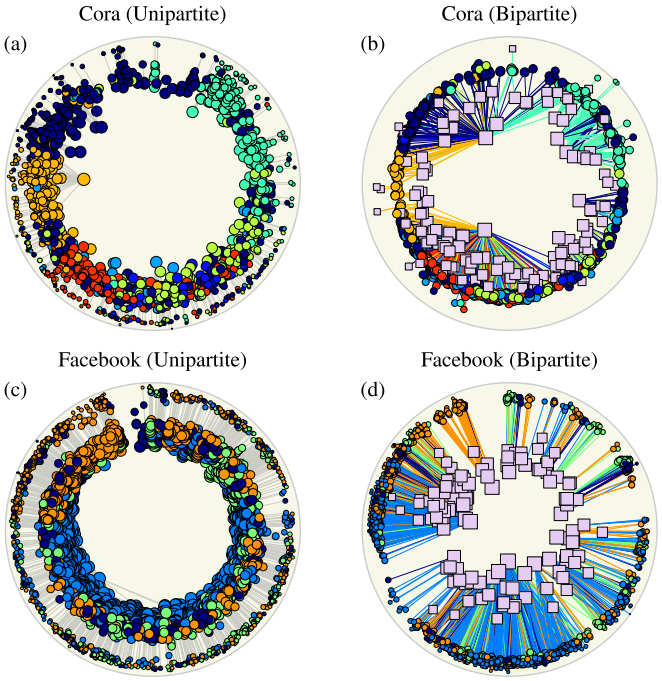

FIG. 6. Hyperbolic representation of the Cora and Facebook datasets. The left column shows hyperbolic embeddings of the unipartite network $\mathcal{G}_n$ and the right column the bipartite network $\mathcal{G}_{n,f}$ between nodes (colored circles) and features (pink squares). Nodes are colored by the labels given in the dataset. In both cases, we show edges with effective distance $\chi < 1$. Sizes of nodes and features are proportional to the logarithm of their degrees.

values, we generated network surrogates with the DPGR algorithm and compared their topological properties: degree distributions, degree-degree correlations, and bipartite clustering spectrum. Figure 2 displays the ensemble average and the two-σ intervals for all the measures. In most of the cases, the model accurately reproduced these properties. However, the model could be further improved by considering that nodes and features may not be uniformly distributed in the similarity space, but instead defining geometric communities, as discussed in [49,50]. As discussed in [58], such a nonuniform distribution of nodes is important for reproducing degree-degree correlations similar to those observed, for instance, in the Facebook dataset.

F. Hyperbolic representation of real datasets

Finally, we provide the hyperbolic representation of the Cora and Facebook datasets. First, we used MERCATOR to embed the unipartite network $\mathcal{G}_n$. The results are displayed in Figs. 6(a) and 6(c), with nodes color-coded according to the labels provided in the dataset. As expected, we observe a high degree of congruence between these labels and the angular organization of nodes, with groups of nodes with the same label distributed in the same angular sectors. Remarkably, this congruence serves as empirical evidence of the suitability of the $\mathbb{S}^1$ model for describing these networks. Notably, MERCATOR does not make use of these labels, making this congruence all the more significant. To represent the bipartite network $\mathcal{G}_{n,f}$, we maintain the angular coordinates of nodes from the embedding of $\mathcal{G}_n$ and adjust the hidden degrees of nodes and features in $\mathcal{G}_{n,f}$, κ_n, and κ_f, as well as β_b to reproduce the degree distribution of nodes and features, as explained in detail in [58]. Subsequently, while keeping the coordinates of the nodes fixed, we adjust the angular coordinates of each individual feature to maximize the likelihood that the feature is connected to its actual neighbors. The results are presented in Figs. 6(b) and 6(d). In this case, the nodes exhibit a homogeneous degree distribution, causing them to appear near the edge of the circle. Conversely, the features exhibit a heterogeneous distribution, leading to a corresponding heterogeneous distribution of radial coordinates. Interestingly, the features are closely connected to nodes within the same angular sectors. This suggests that by using this representation, it is possible to define specific sets of features associated with particular sets of nodes, thus defining bipartite communities in a similar way as was done in [68] for unipartite networks in combination with the bipartite modularity defined in [69]. It is important to note, however, that this approach represents just a preliminary step towards an embedding tool for featured-enriched networks. The comprehensive embedding will emerge from the simultaneous maximization of the likelihood of both graphs, $\mathcal{G}_n$ and $\mathcal{G}_{n,f}$, a topic we will address in a forthcoming publication.

III. DISCUSSION

To summarize, our approach represents a paradigm shift in the description of complex graph-structured data. The crucial element in our framework is to view the relationships between nodes and features as a bipartite graph influenced by the same underlying similarity space that shapes the topology of the network between nodes. In contrast to other models that treat node features as nodes [63–65], we analyze the topology of the resulting bipartite networks, revealing homogeneous degree distributions for nodes, heterogeneous distributions for features, weak correlations between the degrees of connected pairs, as well as a high level of bipartite clustering in the network confirming the presence of an underlying similarity metric space. We hypothesize that this shared similarity space, along with the strength of the coupling between networks $\mathcal{G}_n$ and $\mathcal{G}_{n,f}$ controlled by parameters β and β_b, underlies the effectiveness of GCNs. If this conjecture holds true, our formalism could provide a crucial component in addressing the black box problem [27]. This work represents a preliminary step towards a promising research direction and forms the basis for a future feature-aware embedding method in the hyperbolic space. Indeed, hyperbolic network geometry enhances the interpretability of GCNs in many ways. By using hyperbolic embeddings, it becomes possible to visualize complex, hierarchical data in a more interpretable manner. The distances and relationships in the hyperbolic space can be mapped and visualized, offering insights into the intrinsic structure of the data, thus making the relationships within the data more explicit and understandable. For example, in a hierarchical classification task, the model's decisions can be traced through the hyperbolic space, showing how different categories or concepts are related to each other. Additionally, because hyperbolic space can represent complex structures in lower dimensions, it simplifies the feature space. This

ALGORITHM 1. Algorithm to generate the network of nodes, $\mathcal{G}_n$ [55].

1: Fix the number of nodes, N_n, the level of clustering, β, and the target average degree $\langle k \rangle$.
2: Set $\mu = \frac{\beta}{2\pi\langle k \rangle} \sin\frac{\pi}{\beta}$.
3: Uniformly distribute the nodes at random along a one-dimensional sphere with radius $R = \frac{N_n}{2\pi}$ by assigning each node a parameter $\theta \in [0, 2\pi]$.
4: Assign every node a hidden degree κ sampled from an arbitrary probability density $\rho(\kappa)$ such that $\langle \kappa \rangle = \langle k \rangle$.
5: Connect each pair of nodes with a probability function given by Eq. (2).

ALGORITHM 2. Algorithm to generate the bipartite network of nodes and features, $\mathcal{G}_{n,f}$.

1: Fix the number of features, N_f, the level of clustering, β_b, and the target average degree of nodes $\langle k_n \rangle$ (then $\langle k_f \rangle = \frac{N_n}{N_f}\langle k_f \rangle$, where N_n is the number of nodes in $\mathcal{G}_n$).
2: Set $\mu_b = \frac{\beta_b}{2\pi\langle k_n \rangle} \sin\frac{\pi}{\beta_b}$.
3: Uniformly distribute the features at random along the same sphere as in $\mathcal{G}_n$ by assigning each feature a parameter $\theta_f \in [0, 2\pi]$, with nodes located at the same angular positions as in $\mathcal{G}_n$, i.e., $\theta_n = \theta$.
4: Assign every node and feature a hidden degree κ_n and κ_f, sampled from an arbitrary probability density $\rho_n(\kappa_n)$ and $\rho_f(\kappa_f)$ such that $\langle \kappa_n \rangle = \langle k_n \rangle$ and $\langle \kappa_f \rangle = \langle k_f \rangle$.
5: Connect each pair of nodes and features with a probability function given by Eq. (5).

simplification helps in understanding which features are most relevant in the model's decision-making process.

We thank Robert Jankowski for useful discussions on the measure of correlations between network topology and features.

ACKNOWLEDGMENTS

We acknowledge support from the following: Grant No. TED2021-129791B-I00 funded by MCIN/AEI/10.13039/501100011033 and by "European Union NextGenerationEU/PRTR"; Grant No. PID2022-137505NB-C22 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe"; Generalitat de Catalunya Grant No. 2021SGR00856. M.B. acknowledges the ICREA Academia award, funded by the Generalitat de Catalunya.

APPENDIX: ALGORITHMS TO GENERATE GEOMETRIC GRAPH-STRUCTURED DATA

Algorithms 1 [55] and 2 provide simple steps to generate the network of nodes, $\mathcal{G}_n$, using the $\mathbb{S}^1$ model and the bipartite network between nodes and features, $\mathcal{G}_{n,f}$, using the bipartite-$\mathbb{S}^1$ model.

[1] M. Boguñá, I. Bonamassa, M. D. Domenico, S. Havlin, D. Krioukov, and M. Á. Serrano, Network geometry, Nat. Rev. Phys. **3**, 114 (2021).
[2] M. Á. Serrano, D. Krioukov, and M. Boguñá, Self-similarity of complex networks and hidden metric spaces, Phys. Rev. Lett. **100**, 078701 (2008).
[3] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, Hyperbolic geometry of complex networks, Phys. Rev. E **82**, 036106 (2010).
[4] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguñá, and D. Krioukov, Popularity versus similarity in growing networks, Nature (London) **489**, 537 (2012).
[5] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, Science **286**, 509 (1999).
[6] M. E. J. Newman and A. Clauset, Structure and inference in annotated networks, Nat. Commun. **7**, 11863 (2016).
[7] L. Peel, D. B. Larremore, and A. Clauset, The ground truth about metadata and community detection in networks, Sci. Adv. **3**, e1602548 (2017).
[8] A. Bassolas, A. Holmgren, A. Marot, M. Rosvall, and V. Nicosia, Mapping nonlocal relationships between metadata and network structure with metadata-dependent encoding of random walks, Sci. Adv. **8**, eabn7558 (2022).
[9] S. Emmons and P. J. Mucha, Map equation with metadata: Varying the role of attributes in community detection, Phys. Rev. E **100**, 022301 (2019).
[10] L. M. Smith, L. Zhu, K. Lerman, and A. G. Percus, Partitioning networks with node attributes by compressing information flow, ACM Trans. Knowl. Discov. Data **11** (2016).
[11] O. Artime and M. De Domenico, Percolation on feature-enriched interconnected systems, Nat. Commun. **12**, 2478 (2021).
[12] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, Graph neural networks: A review of methods and applications, AI Open **1**, 57 (2020).
[13] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, A comprehensive survey on graph neural networks, IEEE Trans. Neural Netw. Learn. Syst. **32**, 4 (2021).
[14] M. Gori, G. Monfardini, and F. Scarselli, A new model for learning in graph domains, in *Proceedings of the 2005 IEEE International Joint Conference on Neural Networks, 2005* (IEEE, Montreal, QC, Canada, 2005), Vol. 2, pp. 729–734.
[15] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, The graph neural network model, IEEE Trans. Neural Netw **20**, 61 (2009).
[16] C. Gallicchio and A. Micheli, Graph echo state networks, in *The 2010 International Joint Conference on Neural Networks (IJCNN)* (IEEE, New York, 2010), pp. 1–8.
[17] Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel, Gated graph sequence neural networks, in *International Conference on Learning Representations, ICLR 2016* (ICLR, 2016).
[18] H. Dai, Z. Kozareva, B. Dai, A. Smola, and L. Song, Learning steady-states of iterative algorithms over graphs, in *Proceedings of the 35th International Conference on Machine Learning* (PMLR, Brookline, MA, 2018)), pp. 1106–1114.
[19] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, Spectral networks and locally connected networks on graphs, in *2nd International Conference on Learning Representations, ICLR 2014* (ICLR, 2014).

[20] M. Henaff, J. Bruna, and Y. LeCun, Deep convolutional networks on graph-structured data, arXiv:1506.05163.

[21] M. Defferrard, X. Bresson, and P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, in *Advances in Neural Information Processing Systems* (Curran, Red Hook, NY, 2016), pp. 3844–3852.

[22] T. N. Kipf and M. Welling, Semi-supervised classification with graph convolutional networks, in *International Conference on Learning Representations, ICLR 2017* (ICLR, 2017).

[23] R. Levie, F. Monti, X. Bresson, and M. M. Bronstein, CayleyNets: Graph convolutional neural networks with complex rational spectral filters, IEEE Trans. Signal Process. **67**, 97 (2019).

[24] A. Micheli, Neural network for graphs: A contextual constructive approach, IEEE Trans. Neural Netw. **20**, 498 (2009).

[25] J. Atwood and D. Towsley, Diffusion-convolutional neural networks, in *Advances in Neural Information Processing Systems* (Curran, Red Hook, NY, 2016), pp. 1993–2001.

[26] M. Niepert, M. Ahmed, and K. Kutzkov, Learning convolutional neural networks for graphs, in *Proceedings of The 33rd International Conference on Machine Learning* (PMLR, Brookline, MA, 2016), pp. 2014–2023.

[27] D. Castelvecchi, Can we open the black box of AI? Nature (London) **538**, 20 (2016).

[28] B. Sanchez-Lengeling, J. Wei, B. Lee, E. Reif, P. Wang, W. Qian, K. McCloskey, L. Colwell, and A. Wiltschko, Evaluating attribution for graph neural networks, in *Advances in Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Curran, Red Hook, NY, 2020), Vol. 33, pp. 5898–5910.

[29] J. Park and A.-L. Barabási, Distribution of node characteristics in complex networks, Proc. Natl. Acad. Sci. USA **104**, 17916 (2007).

[30] G. Bianconi, P. Pin, and M. Marsili, Assessing the relevance of node features for network structure, Proc. Natl. Acad. Sci. USA **106**, 11433 (2009).

[31] M. Drobyshevskiy and D. Turdakov, Random graph modeling: A survey of the concepts, ACM Comput. Surv. **52**, 1 (2019).

[32] Z. Wang, J. Wang, Y. Shan, S. Yu, X. Xu, Q. Xuan, and G. Chen, Null model-based data augmentation for graph classification, IEEE Trans. Netw. Sci. Eng. **11**, 1821 (2024).

[33] F. Váša and B. Mišić, Null models in network neuroscience, Nat. Rev. Neurosci. **23**, 493 (2022).

[34] L. Gugelmann, K. Panagiotou, and U. Peter, Random hyperbolic graphs: degree sequence and clustering, in *Automata, Languages, and Programming, ICALP 2012*, Lecture Notes in Computer Science (Springer, New York, 2012), Vol. 7392.

[35] E. Candellero and N. Fountoulakis, Clustering and the hyperbolic geometry of complex networks, Internet Math. **12**, 2 (2016).

[36] N. Fountoulakis, P. van der Hoorn, T. Müller, and M. Schepers, Clustering in a hyperbolic model of complex networks, Electron. J. Probab. **26**, 1 (2021).

[37] M. A. Abdullah, N. Fountoulakis, and M. Bode, Typical distances in a geometric model for complex networks, Internet Math. **1** (2017).

[38] T. Friedrich and A. Krohmer, On the diameter of hyperbolic random graphs, SIAM J. Discrete Math. **32**, 1314 (2018).

[39] T. Müller and M. Staps, The diameter of KPKVB random graphs, Adv. Appl. Probab. **51**, 358 (2019).

[40] M. Á. Serrano, D. Krioukov, and M. Boguñá, Percolation in self-similar networks, Phys. Rev. Lett. **106**, 048701 (2011).

[41] N. Fountoulakis and T. Müller, Law of large numbers for the largest component in a hyperbolic model of complex networks, Ann. Appl. Probab. **28**, 607 (2018).

[42] M. Kiwi and D. Mitsche, Spectral gap of random hyperbolic graphs and related parameters, Ann. Appl. Probab. **28**, 941 (2018).

[43] A. Allard, M. Á. Serrano, G. García-Pérez, and M. Boguñá, The geometric nature of weights in real complex networks, Nat. Commun. **8**, 14103 (2017).

[44] M. Á. Serrano, M. Boguñá, and F. Sagués, Uncovering the hidden geometry behind metabolic networks, Mol. BioSyst. **8**, 843 (2012).

[45] M. Kitsak, F. Papadopoulos, and D. Krioukov, Latent geometry of bipartite networks, Phys. Rev. E **95**, 032309 (2017).

[46] G. Budel and M. Kitsak, Complementarity in complex networks, arXiv:2003.06665.

[47] K.-K. Kleineberg, M. Boguñá, M. Á. Serrano, and F. Papadopoulos, Hidden geometric correlations in real multiplex networks, Nat. Phys. **12**, 1076 (2016).

[48] K.-K. Kleineberg, L. Buzna, F. Papadopoulos, M. Boguñá, and M. Á. Serrano, Geometric correlations mitigate the extreme vulnerability of multiplex vetworks against targeted attacks, Phys. Rev. Lett. **118**, 218301 (2017).

[49] K. Zuev, M. Boguñá, G. Bianconi, and D. Krioukov, Emergence of soft communities from geometric preferential attachment, Sci. Rep. **5**, 9421 (2015).

[50] G. García-Pérez, M. Á. Serrano, and M. Boguñá, Soft communities in similarity space, J. Stat. Phys. **173**, 775 (2018).

[51] A. Muscoloni and C. V. Cannistraci, A nonuniform popularity-similarity optimization (nPSO) model to efficiently generate realistic complex networks with communities, New J. Phys. **20**, 052002 (2018).

[52] G. García-Pérez, M. Boguñá, and M. Á. Serrano, Multiscale unfolding of real networks by geometric renormalization, Nat. Phys. **14**, 583 (2018).

[53] M. Zheng, G. García-Pérez, M. Boguñá, and M. Á. Serrano, Scaling up real networks by geometric branching growth, Proc. Natl. Acad. Sci. USA **118**, e2018994118 (2021).

[54] M. Boguñá, D. Krioukov, P. Almagro, and M. Á. Serrano, Small worlds and clustering in spatial networks, Phys. Rev. Res. **2**, 023040 (2020).

[55] M. A. Serrano and M. Boguñá, The shortest path to network geometry: A practical guide to basic models and applications, in *Elements in Structure and Dynamics of Complex Networks* (Cambridge University Press, Cambridge, England, 2022).

[56] The model can also be defined on spheres of arbitrary dimensions. However, the one-dimensional case captures the most relevant topological properties.

[57] The case $\beta < 1$ has been thoroughly analyzed in [70].

[58] G. García-Pérez, A. Allard, M. Á. Serrano, and M. Boguñá, Mercator: uncovering faithful hyperbolic embeddings of complex networks, New J. Phys. **21**, 123033 (2019).

[59] M. Boguñá, F. Papadopoulos, and D. Krioukov, Sustaining the Internet with hyperbolic mapping, Nat. Commun. **1**, 62 (2010).

[60] E. Giamphy, J.-L. Guillaume, A. Doucet, and K. Sanchis, A survey on bipartite graphs embedding, Soc. Netw. Anal. Min. **13**, 54 (2023).

[61] G. E. Hinton, Training products of experts by minimizing contrastive divergence, Neural Comput. **14**, 1771 (2002).

[62] G. E. Hinton, S. Osindero, and Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. **18**, 1527 (2006).

[63] I. Crimaldi, M. D. Vicario, G. Morrison, W. Quattrociocchi, and M. Riccaboni, Modeling networks with a growing feature-structure, Interdiscip. Inf. Sci. **23**, 127 (2017).

[64] P. Boldi, I. Crimaldi, and C. Monti, A network model characterized by a latent attribute structure with competition, Inf. Sci. **354**, 236 (2016).

[65] C. Becatti, I. Crimaldi, and F. Saracco, Collaboration and followership: A stochastic model for activities in social networks, PLoS One **14**, e0223768 (2019).

[66] M. Starnini, E. Ortiz, and M. Á. Serrano, Geometric randomization of real networks with prescribed degree sequence, New J. Phys. **21**, 053039 (2019).

[67] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.7.033036 for details on real network datasets, correlation analysis between node features and topology, bipartite clustering in the configuration model, and additional experimental results, which includes Refs. [58,66,71–74].

[68] G. García-Pérez, M. Boguñá, A. Allard, and M. A. Serrano, The hidden hyperbolic geometry of international trade: World Trade Atlas 1870–2013, Sci. Rep. **6**, 33441 (2016).

[69] M. J. Barber, Modularity and community detection in bipartite networks, Phys. Rev. E **76**, 066102 (2007).

[70] J. van der Kolk, M. Á. Serrano, and M. Boguñá, An anomalous topological phase transition in spatial random graphs, Commun. Phys. **5**, 245 (2022).

[71] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Galligher, and T. Eliassi-Rad, Collective classification in network data, AI Mag. **29**, 93 (2008).

[72] B. Rozemberczki, C. Allen, and R. Sarkar, Multi-scale attributed node embedding, J. Complex Netw. **9**, cnab014 (2021).

[73] S. P. Borgatti and D. S. Halgin, Analyzing affiliation networks, in *The SAGE Handbook of Social Network Analysis*, edited by J. Scott and P. J. Carrington (SAGE Publications, London, 2011).

[74] M. Á. Serrano, M. Boguñá, and A. Vespignani, Extracting the multiscale backbone of complex weighted networks, Proc. Natl. Acad. Sci. U.S.A. **106**, 6483 (2009).