Latent geometry of bipartite networks

Maksim Kitsak,[1] Fragkiskos Papadopoulos,[2] and Dmitri Krioukov[3]

[1]*Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA*

[2]*Department of Electrical Engineering, Computer Engineering and Informatics, Cyprus University of Technology,*
33 Saripolou Street, 3036 Limassol, Cyprus

[3]*Department of Physics, Department of Mathematics, Department of Electrical and Computer Engineering,*
Northeastern University, Boston, Massachusetts 02115, USA

Despite the abundance of bipartite networked systems, their organizing principles are less studied compared to unipartite networks. Bipartite networks are often analyzed after projecting them onto one of the two sets of nodes. As a result of the projection, nodes of the same set are linked together if they have at least one neighbor in common in the bipartite network. Even though these projections allow one to study bipartite networks using tools developed for unipartite networks, one-mode projections lead to significant loss of information and artificial inflation of the projected network with fully connected subgraphs. Here we pursue a different approach for analyzing bipartite systems that is based on the observation that such systems have a latent metric structure: network nodes are points in a latent metric space, while connections are more likely to form between nodes separated by shorter distances. This approach has been developed for unipartite networks, and relatively little is known about its applicability to bipartite systems. Here, we fully analyze a simple latent-geometric model of bipartite networks and show that this model explains the peculiar structural properties of many real bipartite systems, including the distributions of common neighbors and bipartite clustering. We also analyze the geometric information loss in one-mode projections in this model and propose an efficient method to infer the latent pairwise distances between nodes. Uncovering the latent geometry underlying real bipartite networks can find applications in diverse domains, ranging from constructing efficient recommender systems to understanding cell metabolism.

I. INTRODUCTION AND MOTIVATION

Many real-world networks have a bipartite structure: Nodes can be separated into two disjoint sets and links exist only between nodes of different sets. Real bipartite networks are often characterized by two common properties: (i) heterogeneity in distributions of node degrees and (ii) a large number of common neighbors shared between pairs of nodes. The heterogeneity of degree distributions has been studied extensively in both traditional (unipartite) and bipartite networks and comes as no surprise. In many bipartite systems heterogeneous degree distributions can be approximated by power laws, $P(k) \sim k^{-\gamma}$, which are observed for at least one of the two sets of nodes [1–6]. The second property, on the other hand, has not been studied extensively and deserves a thorough investigation.

In Fig. 1(a) we show the distribution $P(m)$ of the number m of common neighbors between nodes in several real bipartite systems: the actor-film network derived from the International Movie Database (IMDb) [7], condensed matter (Condmat) and high-energy physics (HEP) collaboration networks derived from the arXiv [8], Wikipedia [9], and the network of metabolic reactions [10] (see Appendix A). For each of these networks, we calculated the distribution of the number of common neighbors shared between pairs of nodes, $P(m)$. We see that the number of common neighbors in these networks is power-law distributed,

$$P(m) \sim m^{-\tau}, \quad \tau > 2, \qquad (1)$$

so significant fractions of node pairs share many common neighbors. Similar observations have been made for other bipartite systems. For instance, the probability of two insect species pollinating m different kinds of flowers in common has been shown to follow a truncated power law [11]. Similarly, a fat-tail distribution of the number of shared requests between two users has been observed in peer-to-peer networks [12].

To better understand the mechanisms leading to the abundance of common neighbors in bipartite systems, we first ask if the observed fat-tail distributions of $P(m)$ are the consequence of heterogeneous degree distributions $P(k)$. To answer this question, we randomly rewire our real bipartite networks by preserving the degrees of individual nodes (see Appendix B). We find that $P(m)$ in the randomized networks exhibits very fast decays, such that the maximum number of common neighbors between nodes is very small, see Fig. 1(a). This result suggests that the heterogeneity of $P(m)$ is not caused by the heterogeneity of $P(k)$. Second, we also check if the heterogeneous shape of $P(m)$ is driven by all pairs of nodes in the network or by a handful of high-degree nodes. To this end, we focus on node pairs with a large number of common neighbors. We create a heatmap by counting pairs of HEP authors sharing at least $m = 10$ publications, and whose publication record sizes, i.e., degrees, are k_1 and k_2. As seen from Fig. 1(b), author pairs with $m \geqslant 10$ common publications do not necessarily consist of authors that have published a large number of papers, as one would expect from a random collaboration pattern. On the contrary, we see that the majority of author pairs with at least 10 common publications involve authors who barely published over 10 publications each. This observation is not specific to the HEP collaboration network: We checked that in all considered networks both small- and large-degree nodes can have a large number of common neighbors.
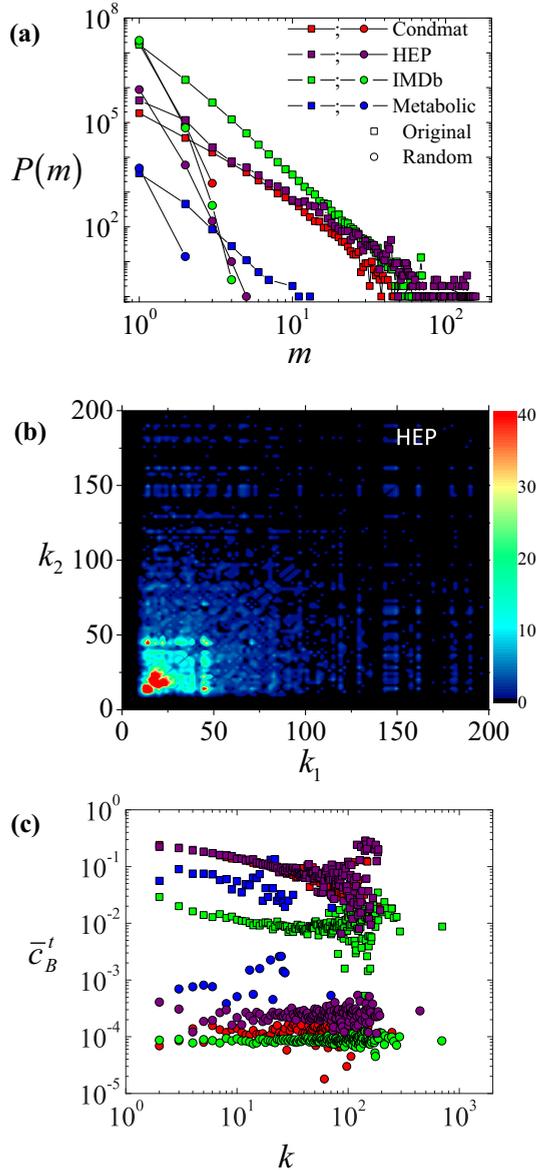
FIG. 1. Number of common neighbors and clustering in real bipartite networks and in their degree-preserving randomized counterparts. (a) The distribution of the number of common neighbors, $P(m)$, for pairs of top nodes in the Condmat and HEP collaboration networks, the IMDb network, and the metabolic network. The results for the original real networks are shown with squares, whereas those for their degree-preserving randomized counterparts are shown with circles. By the *top* nodes in each network we refer respectively to authors, actors, and metabolites, whereas manuscripts, films, and reactions are referred to as *bottom* nodes (Appendix A). (b) Number of HEP author pairs having $m \geqslant 10$ common publications as a function of their total number of publications k_1 and k_2. (c), Average bipartite clustering coefficient of top nodes, $\bar{c}_B^t$, as a function of their degree k for the same real networks (squares) as in panel (a) and for their degree-preserving randomized counterparts (circles). The bipartite clustering coefficient is significantly larger in the original real networks compared to their randomized counterparts.

The heterogeneity in the observed number of common neighbors implies the existence of a large number of 4-loops
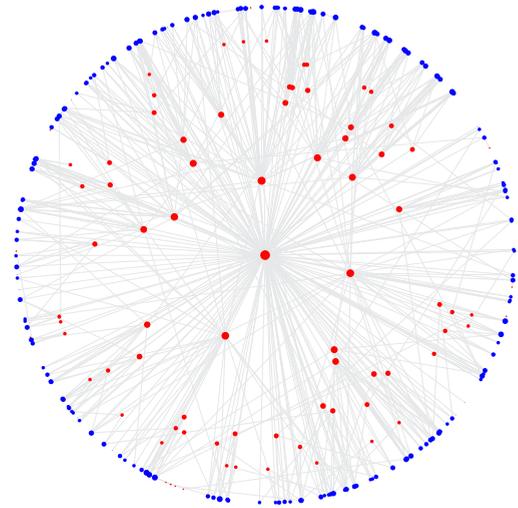

FIG. 2. Schematic illustration of a synthetic bipartite network in latent geometric space. The network consists of $N = 100$ top (red) nodes and $M = 200$ bottom (blue) nodes, connected as described in Sec. III. The nodes are placed on a disk of radius 1. Their angular coordinates are random on $[0, 2\pi]$. The radial coordinates of all bottom nodes are $r_j = 1$, $\forall j$, so they all are at the disk edge, while the radial coordinates of top nodes are $r_i = 1 - \frac{\ln k_i}{\ln k_{max}}$, where k_i is the degree of node i and k_{max} is the largest node degree of the top domain. Node sizes are proportional to the logarithm of node degrees.

in real bipartite networks. Indeed, a pair of nodes a, b sharing a large number of common neighbors will have a large number of 4-loops passing through them, that is, loops of the form $a \rightarrow c \rightarrow b \rightarrow d \rightarrow a$. Supporting this observation, we also find that real bipartite systems are characterized by strong bipartite clustering, which quantifies the density of 4-loops in the network, see the definition in Sec. III G. Bipartite clustering is typically several orders of magnitude larger in the original networks compared to their degree-preserving randomized counterparts, Fig. 1(c). We also note that similar clustering-related heterogeneity has also been observed in unipartite networks. Many real unipartite networks have been shown to exhibit power-law distributions of edge multiplicity, defined as the number of triangles shared by edges [13].

A. Latent geometry of bipartite networks

Here we show that the observed common properties of real bipartite networks can be explained by the existence of latent geometric spaces underlying these networks. That is, we assume that nodes in bipartite networks are points in some geometric space underlying the system. The latent coordinates of nodes in the space abstract node attributes, while latent distances between nodes play the role of a generalized similarity measure: The more similar the two nodes, the smaller the latent distance between them, and the higher the probability that the nodes are connected (Fig. 2). To illustrate, consider the IMDb network for instance, where actors are linked to films they starred in. Clearly, connections in this network are not random. Both actors and films can be characterized by numerous attributes, so connections are the result of a certain mutual match between these attributes.

These attributes include genres, geographic locations of films, film release dates, etc. Similar parallels can be drawn for other systems. For instance, connections between authors and manuscripts in collaboration networks are driven by many factors, including the expertise of the authors, their geographic location, their methodology, and so on. Collectively these attributes define similarity distances between nodes in a latent space.

The large numbers of common neighbors and strong clustering observed in real bipartite systems are then a reflection of the triangle inequality in the latent space. Consider, for example, two actors A_1 and A_2 and a film F mapped to three points in a metric space. The triangle inequality in the space prescribes that $d(A_2,F) < d(A_1,A_2) + d(A_1,F)$, where $d(X,Y)$ denotes the distance between nodes X,Y in the space. If distances $d(A_1,A_2)$ and $d(A_1,F)$ are small, then $d(A_2,F)$ is also small, so both actors A_1,A_2 are likely to costar in film F; that is, F is likely to be a common neighbor of A_1 and A_2. As the number of common films between two actors A_1,A_2 increases, so does bipartite clustering, that is, the number of loops of the form $A_1 \rightarrow F_1 \rightarrow A_2 \rightarrow F_2 \rightarrow A_1$. We stress the importance of the metric property in a latent space: If latent distances do not satisfy the triangle inequality, then bipartite networks built using these distances do not have many common neighbors and strong bipartite clustering (Appendix C).

B. Organization of the manuscript

To further support our geometric assumption we organize the rest of the manuscript as follows.

We begin with the review of related work in Sec. II. In Sec. III we conduct a detailed analysis of a model of synthetic bipartite networks constructed in the latent space. We show that the model generates bipartite networks with either heterogeneous or homogeneous degree distributions in a given node domain and with power-law distributions of the number of common neighbors and strong bipartite clustering. In Sec. IV, we investigate how the latent geometry of bipartite systems is transformed under one-mode projections and prove that one-mode projections cannot fully preserve latent geometry. However, we also show that under certain conditions latent geometry can be preserved approximately. In Sec. V, we propose a procedure for efficient estimation of the latent pairwise distances between pairs of nodes. Final remarks are in Sec. VI.

II. RELATED WORK

Bipartite networks have been successfully used to model a large array of complex systems including collaboration networks [14,15], metabolic reactions [10,16,17], peer-to-peer networks [12,18], and recommender systems [19–22]. Bipartite networks can be represented as hypergraphs, generalizations of graphs where a single edge can connect multiple nodes [23]. Hypergraphs, in their turn, are further generalizable to multipartite hypergraphs, where hyperedges may connect several nodes of different type. Recently, tripartite hypergraphs have been proposed to model tagged social networks, also known as folksonomies [24–28].

The concept of latent space has been initially introduced to model homophily and similarity in social networks [29,30]. Lately, latent space models are attracting great interest in many diverse fields including sociology [31–33], statistical physics [34,35], and computer science [34–36]. Another closely related research area is that on random geometric graphs, well studied in mathematics and engineering [37–39], particularly due to its relevance to wireless networks [39].

Two models of random bipartite geometric graphs have been proposed recently. The first model is the AB random geometric graph (AB RGG) [40], defined as the two sets of points scattered as two independent Poisson processes in Euclidean space with connections between points from different sets established if distance between them is less then the threshold distance. AB RGGs are motivated by wireless networks where transmission and reception of a signal occurs at different frequencies [41]. Thus, of primary interest in AB RGGs are the connectivity and percolation properties [40,42].

The second model is inspired by the hidden variable formalism [43,44]: Network nodes map to points in the latent space and connections between them are drawn with probabilities depending on distances between the nodes in the underlying space. It was shown in Refs. [45–47] that if latent geometry is hyperbolic, then random geometric graphs in it reproduce common structural and dynamical properties of unipartite networks—scale-free degree distribution, strong clustering, community structure, and large-scale growth dynamics. Equivalent to hyperbolic random graphs are random graphs in Euclidean space with power-law distributed hidden variables [45]. This model has been recently generalized to bipartite networks in Ref. [48], where it was called the $\mathbb{S}^1 \times \mathbb{S}^1$ model and utilized to study cell metabolism.

III. TOPOLOGICAL PROPERTIES OF BIPARTITE NETWORKS AS REFLECTIONS OF LATENT GEOMETRY

In this section we conduct a detailed analysis of the $\mathbb{S}^1 \times \mathbb{S}^1$ model of a bipartite network in the simplest compact latent space, circle $\mathbb{S}^1$.

A. Definitions and the $\mathbb{S}^1 \times \mathbb{S}^1$ model

We refer to the two groups of nodes in a bipartite network as top and bottom nodes, and denote their number by N and M, respectively. Within the modeling framework we consider, network nodes map to points in a latent geometric space, and, as a result, every node of the network is characterized by its coordinates in this space. Both top and bottom nodes belong to the same space and, thus, distances are defined between all pairs of nodes. Yet, to generate bipartite networks, connections are allowed only between nodes of different domains. To achieve heterogeneity in node degrees and to allow some nodes to connect over large distances, every node is also assigned a hidden variable. To distinguish top and bottom nodes, we denote these hidden variables as $\{\kappa_i\}$ and $\{\lambda_j\}$, respectively.

To form a bipartite network, every top-bottom pair of nodes i,j is connected with a probability r_{ij}, which depends on the distance between the nodes and their hidden variables,

$$r_{ij} = r\left(\frac{d_{ij}}{d_c(\kappa_i, \lambda_j)}\right), \qquad (2)$$

where $r(x)$ is the connection probability function, d_{ij} is the distance between the nodes in the geometric space, and $d_c(\kappa_i, \lambda_j)$ is a characteristic distance scale, allowing one to vary the importance of small distances depending on the nodes' hidden variables. Even though any integrable function $r(x)$ can, in principle, serve as the connection probability function, we use

$$r(x) = (1 + x^\beta)^{-1}, \quad \beta \in (1, \infty). \tag{3}$$

Our choice for the connection probability function is dictated by the maximum entropy principle [45] and formalizes one's intuition that similar nodes are more likely to be connected than dissimilar nodes. Indeed, $r(x)$ is a decreasing function of x, such that $r(x) \to 0$ as $x \to \infty$ and $r(x) \to 1$ as $x \to 0^+$. Further, parameter β in Eq. (3) controls the abundance of long-distance connections: The larger the β, the less preferred the longer-distance connections.

We focus on the simplest realization of the model, where top and bottom nodes are placed uniformly at random on a one-dimensional Euclidean ring $\mathbb{S}^1$ of radius R, with probability density functions (pdfs) $\rho_t(\theta) = \rho_b(\phi) = \frac{1}{2\pi}$. Given R, the densities of top and bottom nodes on the ring are $\delta_t = \frac{N}{2\pi R}$ and $\delta_b = \frac{M}{2\pi R}$. The hidden variables of the top and bottom nodes are drawn at random from pdfs $\rho_t(\kappa)$ and $\rho_b(\lambda)$. To ease notation, in the rest of the paper we drop the indices from the top and bottom hidden variable distributions, i.e., $\rho_t(\kappa) \equiv \rho(\kappa)$ and $\rho_b(\lambda) \equiv \rho(\lambda)$. To achieve heterogeneous degree distributions, we choose the characteristic scale in Eq. (2) as $d_c(\kappa_i, \lambda_j) = \mu \kappa_i \lambda_j$, which allows nodes with large hidden variables to connect over large distances with higher probability. Parameter $\mu > 0$ rescales all latent distances and controls the expected degrees of the top and bottom domains. Without loss of generality, we set $R = \frac{N}{2\pi}$, which corresponds to the unit density of top nodes, i.e., $\delta_t = 1$. We are interested in large and sparse bipartite networks, $E \propto N \propto M \gg 1$, where E is the number of edges.

Even though nodes from both domains in the model belong to the same Euclidean ring $\mathbb{S}^1$, we refer to the model as the $\mathbb{S}^1 \times \mathbb{S}^1$ model to emphasize its bipartite structure and to distinguish it from the $\mathbb{S}^1$ model for unipartite networks developed in Ref. [49]. The model is fully specified by the number of top and bottom nodes N and M, the connection probability function $r(x)$, and the pdfs $\rho(\kappa)$ and $\rho(\lambda)$. It can be summarized as follows:

(1) Sample the angular coordinates of top nodes θ_i, $i = 1, 2, \ldots, N$, uniformly at random from $[0, 2\pi]$, and their hidden variables κ_i, $i = 1, 2, \ldots, N$, from the pdf $\rho(\kappa)$;

(2) Sample the angular coordinates of bottom nodes ϕ_j, $j = 1, 2, \ldots, M$, uniformly at random from $[0, 2\pi]$, and their hidden variables λ_j, $j = 1, 2, \ldots, M$, from the pdf $\rho(\lambda)$;

(3) Connect every top-bottom node pair with probability

$$r(\kappa_i, \theta_i; \lambda_j, \phi_j) = \frac{1}{1 + \left[\frac{d(\theta_i, \phi_j)}{d_c(\kappa_i, \lambda_j)}\right]^\beta},$$
$$d(\theta_i, \phi_j) = R(\pi - |\pi - |\theta_i - \phi_j||),$$
$$d_c(\kappa_i, \lambda_j) = \mu \kappa_i \lambda_j. \tag{4}$$

The hidden variables in the $\mathbb{S}^1 \times \mathbb{S}^1$ model allow long distance connections among some nodes and are necessary

to achieve heterogeneous degree distributions. An alternative approach to achieve heterogeneity in node degrees is to consider latent spaces of nonzero curvature. Even though both approaches are fully equivalent (see Appendix D), we utilize the former approach as it is more convenient for calculations.

B. Basic properties

The basic topological properties of synthetic networks constructed by the $\mathbb{S}^1 \times \mathbb{S}^1$ model can be obtained in a straightforward manner. Since angular node coordinates are sampled uniformly at random from $[0, 2\pi]$, the expected degree of a top node with hidden variable κ and angular position θ, $\bar{k}(\kappa, \theta)$, is given by

$$\bar{k}(\kappa, \theta) = \frac{M}{2\pi} \int d\lambda \, \rho(\lambda) \int_0^{2\pi} d\phi \, r\left(\frac{d(\theta, \phi)}{\mu \kappa \lambda}\right). \tag{5}$$

Notice that due to the uniform angular distribution of nodes $\bar{k}(\kappa, \theta)$ is independent of θ, $\bar{k}(\kappa, \theta) = \bar{k}(\kappa)$. The evaluation of the inner integral in Eq. (5) leads to

$$\bar{k}(\kappa) = \frac{\mu M}{\pi R} \int d\lambda \, \lambda \rho(\lambda) K \, {}_2F_1\left(1, \frac{1}{\beta}, 1 + \frac{1}{\beta}, -K^\beta\right), \tag{6}$$

$$K \equiv \frac{R\pi}{\mu \kappa \lambda}, \tag{7}$$

where ${}_2F_1$ is the hypergeometric function. For sufficiently large networks, the expected degrees for nodes with κ values satisfying $M \gg \kappa$ can be approximated as

$$\bar{k}(\kappa) \approx \frac{\mu M I}{\pi R} \kappa \bar{\lambda}, \tag{8}$$

where

$$I \equiv \int_0^\infty r(x) \, dx = (\pi/\beta)\csc(\pi/\beta), \tag{9}$$

and $\bar{\lambda} \equiv \int d\lambda \, \lambda \rho(\lambda)$. The expected degree of the entire top node domain, $\bar{k}$, is given by averaging $\bar{k}(\kappa)$ over all possible κ values,

$$\bar{k} = \int d\kappa \, \rho(\kappa) \bar{k}(\kappa) \approx \frac{\mu M I}{\pi R} \bar{\kappa} \bar{\lambda}, \tag{10}$$

where $\bar{\kappa} \equiv \int d\kappa \, \kappa \rho(\kappa)$. Since the model is defined symmetrically for top and bottom nodes, the expected degrees for the bottom domain can be obtained by swapping top and bottom node variables in Eqs. (8) and (10),

$$\bar{\ell}(\lambda) \approx \frac{\mu N I}{\pi R} \lambda \bar{\kappa}, \tag{11}$$

$$\bar{\ell} \approx \frac{\mu N I}{\pi R} \bar{\kappa} \bar{\lambda}. \tag{12}$$

It can be seen from Eqs. (8) and (11) that μ is a dumb parameter in the sense that for any particular value of μ, one can always rescale the hidden variables assigned to top and bottom nodes, $\{\kappa_i\}, \{\lambda_j\}$, in order to obtain desired $\{\bar{k}_i(\kappa_i)\}$ and $\{\bar{\ell}_j(\lambda_j)\}$ values. Therefore, to simplify notation we set

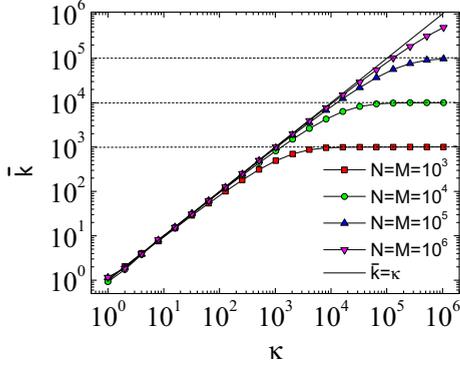$$\mu = \frac{\pi R}{M I \bar{\lambda}} = \frac{\pi R}{N I \bar{\kappa}}. \tag{13}$$

FIG. 3. Expected degree of the node, $\bar{k}$, in the $\mathbb{S}^1 \times \mathbb{S}^1$ model as a function of its hidden variable κ. The $\bar{k}(\kappa)$ curves are shown for different bottom domain sizes and $\rho(\lambda) = (\gamma_b - 1)\lambda_0^{\gamma_b-1}\lambda^{-\gamma_b}$ with $\gamma_b = 2.5$ and λ_0 corresponding to $\bar{\lambda} = 10$. Each point is the average over 100 realizations. Note that $\bar{k}(\kappa) \approx \kappa$ for $\kappa \ll M$, supporting Eqs. (8) and (14). At the same time, $\bar{k}(\kappa) \to M$ as $\kappa \to \infty$. The horizontal dashed lines show corresponding values of M.

The second equality in the above relation holds since the expected number of links in the bipartite network is $E = M\bar{\ell} = N\bar{k}$. Using Eq. (13), we can rewrite Eqs. (8)–(12) as

$$\bar{k}(\kappa) = \kappa, \tag{14}$$

$$\bar{k} = \bar{\kappa}, \tag{15}$$

$$\bar{\ell}(\lambda) = \lambda, \tag{16}$$

$$\bar{\ell} = \bar{\lambda}. \tag{17}$$

Equations (14) and (16) indicate that the hidden variables of nodes are their expected degree values in the resulting topology, see Fig. 3. Figure 3 also illustrates how the approximation in Eq. (8) becomes better for high values of κ as the size of the network increases.

C. Degree distributions

To compute the degree distributions of the top and bottom nodes we consider the propagators $g(k|\kappa,\theta)$ and $g(\ell|\lambda,\phi)$. The propagator $g(k|\kappa,\theta)$ ($g(\ell|\lambda,\phi)$) is defined as the probability that a top (bottom) node with hidden variables κ,θ (λ,ϕ) forms exactly k (ℓ) connections to bottom (top) nodes. Since the angular coordinates of nodes are uniformly distributed, the propagators do not depend on the node angles θ and ϕ, and in the case of sparse bipartite networks they can be approximated by the Poisson distribution [50], that is,

$$g(k|\kappa,\theta) = g(k|\kappa) \approx e^{-\kappa}[\kappa]^k/k!, \tag{18}$$

$$g(\ell|\lambda,\phi) = g(\ell|\lambda) \approx e^{-\lambda}[\lambda]^\ell/\ell!. \tag{19}$$

The degree distributions of the top and bottom domains are obtained by averaging the corresponding propagators over the possible values of the hidden variables,

$$P(k) = \int d\kappa \, \rho(\kappa)g(k|\kappa), \tag{20}$$

$$P(\ell) = \int d\lambda \, \rho(\lambda)g(\ell|\lambda). \tag{21}$$

It can be seen from Eqs. (18)–(21) that $P(k)$ and $P(\ell)$ are independent of one another—they only depend on $\rho(\kappa)$ and $\rho(\lambda)$, respectively. Furthermore, the Poissonian character of the propagators $g(k|\kappa)$ and $g(\ell|\lambda)$ indicates that the resulting degree values of nodes in both domains are narrowly distributed around their hidden variables. This means that the functional forms of $P(k)$ and $P(\ell)$ will be similar to those of $\rho(\kappa)$ and $\rho(\lambda)$, allowing one to construct different degree distributions by engineering proper pdfs of hidden variables $\rho(\kappa)$ and $\rho(\lambda)$. Even though real bipartite systems are characterized by different degree distributions, of our primary interest are scale-free and Poissonian distributions, which we discuss in Sec. III E below.

D. Degree-degree correlations

Degree-degree correlations can be quantified using the average nearest neighbor degree, defined as the average degree of all neighbors of nodes with given degree k [51]. It is straightforward to verify that the $\mathbb{S}^1 \times \mathbb{S}^1$ model is characterized by random degree-degree correlations due to the uniform placement of nodes on $\mathbb{S}^1$:

$$\bar{\ell}_{nn}(k) = \frac{\overline{\ell^2}}{\overline{\ell}}, \tag{22}$$

$$\bar{k}_{nn}(\ell) = \frac{\overline{k^2}}{\overline{k}}. \tag{23}$$

Indeed, the connection probability between two nodes with fixed hidden variables κ_i and λ_j is proportional to the product of these hidden variables:

$$r(\kappa_i,\lambda_j) = \frac{1}{4\pi^2}\iint d\theta_i \, d\phi_j \, r(\kappa_i,\theta_i;\lambda_j\phi_j) \propto \lambda_i\kappa_j. \tag{24}$$

Then, since hidden variables are equal to expected node degrees, this result can be regarded as the soft equivalent of $p(k_i,\ell_j) \propto k_i\ell_j$ in uncorrelated bipartite networks, where $p(k_i,\ell_j)$ is the probability that randomly chosen nodes with degrees k_i and ℓ_j are connected. The rigorous proof of Eqs. (22) and (23) can be obtained following the hidden variable formalism for bipartite networks [50].

E. Categories of bipartite networks

Based on the degree distributions of the top and bottom domains, real bipartite networks often fall into two categories. The first category corresponds to networks with scale-free degree distribution in both top and bottom domains (sf-sf). The second category corresponds to networks with scale-free degree distribution in one domain and Poisson degree distribution in the other domain (sf-ps). Among the real networks that we consider, IMDb, Wikipedia, and the HEP collaboration network fall into the first category. The Condmat collaboration network and the network of metabolic reactions fall into the second category (see Appendix A).

Since $P(k)$ and $\rho(\kappa)$ are expected to be of similar functional form, one can see that a scale-free degree distribution $P(k) \sim k^{-\gamma}$ can be obtained by using the continuous power-law distribution of κ on $[\kappa_0,\infty)$

$$\rho(\kappa) = (\gamma - 1)\kappa_0^{\gamma-1}\kappa^{-\gamma} \tag{25}$$

with the desired ("target") value of exponent γ of degree distribution $P(k)$. Indeed, it follows from Eqs. (18) and (20) that in this case

$$P(k) \approx (\gamma - 1)\kappa_0^{\gamma-1} \frac{\Gamma[k - \gamma + 1, \kappa_0]}{\Gamma[k+1]} \sim k^{-\gamma}. \quad (26)$$

Parameter $\kappa_0 > 0$ is the smallest κ value, i.e., the expected minimum node degree, which also controls the mean value of κ,

$$\bar{\kappa} = \kappa_0(\gamma - 1)/(\gamma - 2), \quad (27)$$

and the expected averaged degree $\bar{k} = \bar{\kappa}$ (15).

The Poisson degree distribution can be obtained by choosing $\rho(\kappa) = \delta(\kappa - \bar{\kappa})$, where $\delta(x)$ is the Dirac δ function and $\bar{\kappa}$ is the expected degree of the domain. The degree distribution $P(k)$ in this case is given by

$$P(k) = g(k|\bar{\kappa}) = e^{-\bar{\kappa}}[\bar{\kappa}]^k/k!. \quad (28)$$

Due to the symmetry of the model, the degree distribution of the bottom domain can be obtained similarly through the proper choice of $\rho(\lambda)$. The independence of $P(k)$ and $P(\ell)$ allows one to construct bipartite networks with an arbitrary combination of degree distributions for the top and bottom domains.

For illustration, we visualize in Fig. 2 a toy sf-ps network consisting of $N = 100$ and $M = 200$ nodes. The hidden variables of the top nodes are drawn from the pdf $\rho(\kappa) = (\gamma - 1)\kappa^{-\gamma}, \gamma = 2.1, \kappa \in [1, \infty), \bar{\kappa} = (\gamma - 1)/(\gamma - 2)$, while the bottom node hidden variables are chosen as $\lambda_i = \bar{\lambda}$ for all nodes i, where $\bar{\lambda}$ satisfies $N\bar{\kappa} = M\bar{\lambda}$. The connections are drawn with probability $r(x)$ prescribed by Eq. (4) with $\beta = 1.5$.

F. Number of common neighbors

The number of common neighbors is the most basic nonbinary network-based measure of similarity between two nodes in a bipartite network—the smaller the similarity distance between two nodes, the more similar the nodes are, and the larger the number of neighbors they are expected to share. This makes the number of common neighbors a crucial measure, allowing one to estimate the similarity distance between two nodes in a bipartite network, Sec. V. Below, we analyze this measure in the $\mathbb{S}^1 \times \mathbb{S}^1$ model.

Consider two top nodes characterized by hidden variables κ_1 and κ_2 and angular coordinates θ_1 and θ_2. The probability p_{12} that these two nodes are simultaneously connected to a bottom node with hidden variable λ and angular coordinate ϕ is

$$p_{12} = r(\kappa_1, \theta_1; \lambda, \phi) r(\kappa_2, \theta_2; \lambda, \phi), \quad (29)$$

where $r(\kappa, \theta; \lambda, \phi)$ is the connection probability in Eq. (4). The expected number of common neighbors between these two top nodes, $\bar{m}(\kappa_1, \theta_1; \kappa_2, \theta_2)$, can be calculated by averaging p_{12} over all possible positions ϕ and hidden variables λ of bottom nodes,

$$\bar{m}(\kappa_1, \theta_1; \kappa_2, \theta_2)$$
$$= \frac{M}{2\pi} \iint \rho(\lambda) r(\kappa_1, \theta_1; \lambda, \phi) r(\kappa_2, \theta_2; \lambda, \phi) \, d\lambda \, d\phi. \quad (30)$$

Due to the uniform distribution of angular coordinates, $\bar{m}(\kappa_1, \theta_1; \kappa_2, \theta_2)$ depends on the angular (similarity) distance between the two top nodes, $\Delta\theta_{12} = \pi - |\pi - |\theta_1 - \theta_2||$, and not on their individual coordinates θ_1 and θ_2. That is, $\bar{m}(\kappa_1, \theta_1; \kappa_2, \theta_2) \equiv \bar{m}(\kappa_1, \kappa_2, \Delta\theta_{12})$. It is straightforward to verify (see Appendix E 1) that $\bar{m}(\kappa_1, \kappa_2, \Delta\theta_{12})$ is independent of the network size $N \propto M$. It depends only on the node hidden variables κ_1, κ_2 and the distance between the nodes $d_{12} = R\Delta\theta_{12} \propto N\Delta\theta_{12}$. It follows from Eq. (30) that for any values of κ_1, κ_2, $\bar{m}(\kappa_1, \kappa_2, \Delta\theta_{12})$ decreases as the angular distance $\Delta\theta_{12}$ increases, for both domains of sf-sf and of sf-ps bipartite networks, following a power law,

$$\bar{m}(\kappa_1, \kappa_2, \Delta\theta_{12}) \sim \Delta\theta_{12}^{-\beta}, \quad (31)$$

where the exponent $\beta > 1$ is the parameter in the connection probability function in Eq. (4), see Fig. 4(a) and Eq. (G13) in Appendix G. The conditional probability for two nodes with hidden variables κ_1, κ_2 separated by angular distance $\Delta\theta_{12}$ to have m common neighbors, is narrowly distributed around its ensemble average $\bar{m}(\kappa_1, \kappa_2, \Delta\theta_{12})$, and in the case of sparse bipartite networks can be approximated by the Poisson distribution

$$P(m|1,2) \approx e^{-\bar{m}(\kappa_1, \kappa_2, \Delta\theta_{12})}[\bar{m}(\kappa_1, \kappa_2, \Delta\theta_{12})]^m/m!, \quad (32)$$

where $P(m|1,2)$ is the shorthand notation for $P(m|\kappa_1, \kappa_2, \Delta\theta_{12})$, see Fig. 4(b) and Appendix E 2.

Finally, the unconditional distribution of the number of common neighbors, $P(m)$, is obtained by averaging $P(m|1,2)$ over all possible hidden variables κ_1, κ_2 and angular distances $\Delta\theta_{12}$,

$$P(m) = \frac{1}{\pi} \iiint P(m|1,2)\rho(\kappa_1)\rho(\kappa_2) \, d\kappa_1 \, d\kappa_2 \, d\Delta\theta_{12}. \quad (33)$$

As before, the corresponding expressions for the bottom domain nodes can be obtained by swapping the variables (κ, θ) with (λ, ϕ) and following the same analysis. The solution of the integral in Eq. (33) depends on the functional form of $P(m|1,2)$, which in turn depends on the pdfs of the hidden variables and on the value of parameter β. While in general there is no closed-form solution to Eq. (33), different closed-form solutions can be obtained for integer values of β. For instance, when $\beta = 2$ and $\rho(\kappa) \sim \kappa^{-\gamma}, \rho(\lambda) = \delta(\lambda - \bar{\lambda})$ (sf-ps networks), we can show that $P(m)$ for the top domain scales as

$$P(m) \sim m^{-\tau}, \quad (34)$$

with $\tau = 2\gamma - 3/2$ (see Appendix E 2). Our numerical experiments indicate that a similar power-law scaling of $P(m)$ also holds for a range of β values and for both domains of sf-sf networks, cf. Figs. 4(c) and 8(c)–8(h) in Appendix E 2.

The power-law scaling of $P(m)$ means that a large number of node pairs have many common neighbors, and, therefore, many 4-loops passing through them, which as explained in Sec. I implies strong bipartite clustering. We focus on bipartite clustering below.

G. Bipartite clustering

To quantify bipartite clustering, we consider the bipartite clustering coefficient $c_B(i)$ introduced by Zhang et al. [52],
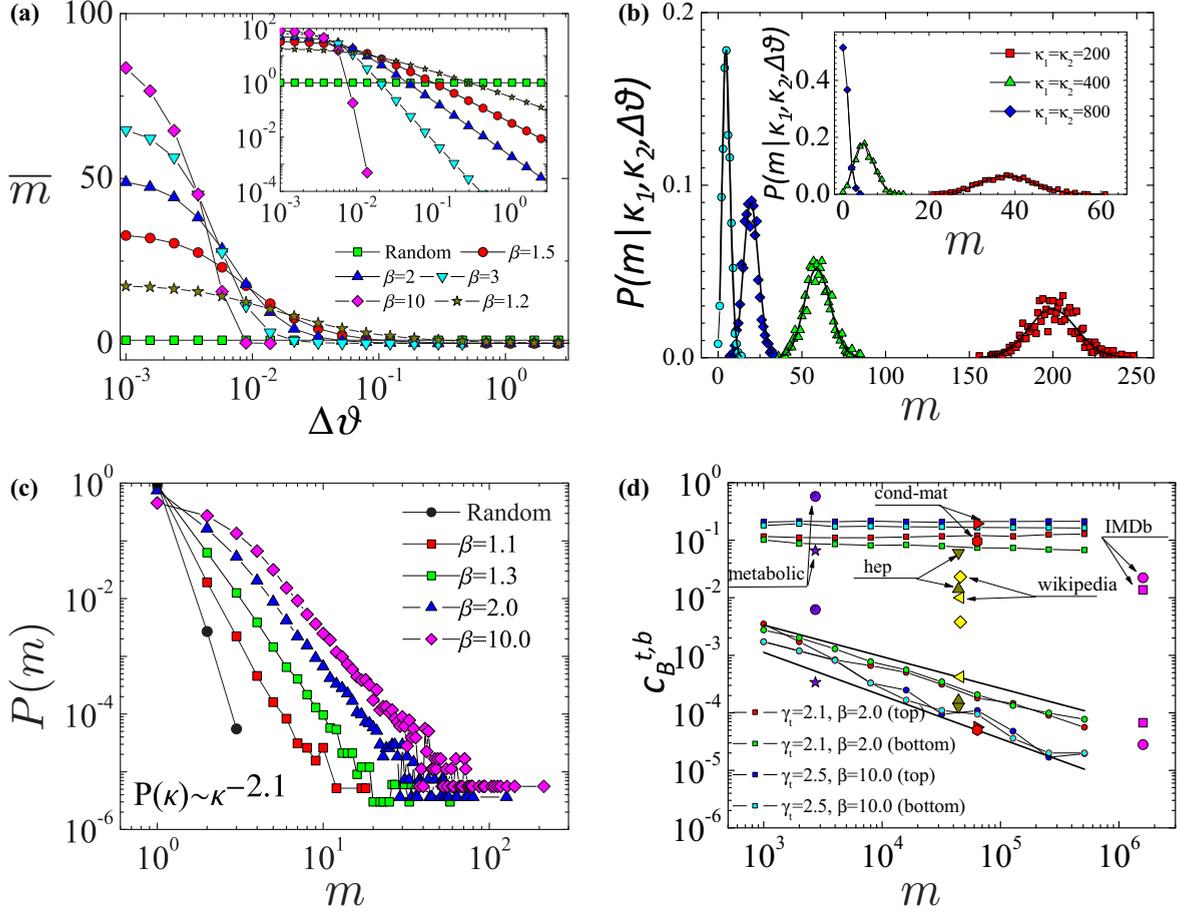
FIG. 4. Number of common neighbors and bipartite clustering in the $\mathbb{S}^1 \times \mathbb{S}^1$ model. Plots (a)–(c) correspond to sf-ps bipartite networks with $N = M = 10^5$ nodes, $\rho(\kappa) \sim \kappa^{-2.1}$ with $\overline{\kappa} = 11.0$, and $\rho(\lambda) = \delta(\lambda - 1)$. The results for the degree-preserving randomized counterparts of the networks are marked as Random. (a) The average number of common neighbors $\overline{m}$ as a function of the angular distance $\Delta\theta$ between two top nodes with $\kappa_1 = \kappa_2 = 100$, i.e., $\overline{m} \equiv \overline{m}(100,100,\Delta\theta)$. $\overline{m}$ is calculated numerically for different values of β using Eq. (30). Higher values of β favor connections at smaller angular distances, i.e., between more similar top-bottom nodes. Shown in the inset, is the same plot in log-log scale. $\overline{m}$ decreases asymptotically as a power-law function of $\Delta\theta$, $\overline{m} \sim \Delta\theta^{-\beta}$. (b) Conditional distribution of the number of common neighbors, $P(m|\kappa_1,\kappa_2,\Delta\theta)$, for two top nodes with $\kappa_1 = \kappa_2 = 400$ separated by different angular distances $\Delta\theta$. From left to right: $\Delta\theta = \pi/32$, $\Delta\theta = \pi/64$, $\Delta\theta = \pi/128$, $\Delta\theta = 0$. The inset shows $P(m|\kappa_1,\kappa_2,\Delta\theta)$ for two top nodes with different values of κ_1, κ_2, separated by $\Delta\theta = \pi/32$. Solid lines are corresponding analytical results using (32). (c) The distribution of the number of common neighbors, $P(m)$, for different values of β, calculated across all pairs of top nodes. (d) Average bipartite clustering coefficients, $\overline{c}_B^{t,b}$, as a function of the network size N for sf-ps modeled networks (squares) and for their degree-preserving randomized counterparts (circles). All modeled networks have the same number of top and bottom nodes, $N = M$. The network parameters γ_t (power-law exponent of the top domain) and β are marked on the plot. The average degrees of the top and bottom domains are $\overline{k} = \overline{\ell} = 11$ for $\gamma_t = 2.1$, and $\overline{k} = \overline{\ell} = 3$ for $\gamma_t = 2.5$. A maximum degree cutoff, $k_{\max} = N^{1/2}$, was imposed in all generated networks in order to avoid structural correlations. $\overline{c}_B^{t,b}$ in the original networks are large and independent of their size, while in their randomized counterparts are small and vanish as $\overline{c}_B^{t,b} \sim N^{-\delta}$, with $\delta = \frac{\gamma_{t,b}-1}{2}$. The plot also shows top and bottom clustering $\overline{c}_B^{t,b}$ for the real bipartite networks (IMDb network, the Condmat and HEP collaboration networks, Wikipedia, and the metabolic network) and for their degree-preserving randomized counterparts. We use the same symbols and colors for each real network and for its randomized counterpart, and we observe a similar behavior as in the modeled networks. Clustering in the original real networks is always orders of magnitude larger than clustering in the corresponding randomized networks.

which aims at quantifying the density of 4-loops adjacent to a node i,

$$c_B(i) = \frac{\sum_{j\neq l}(m_{jl} - 1)}{\sum_{j\neq l}[k_j + k_l - m_{jl} - 1]}. \tag{35}$$

The summation $\sum_{j\neq l}$ goes over all pairs of neighbors j,l of node i, m_{jl} is the number of common neighbors between

j and l, and k_j and k_l are the degrees of j and l. As seen from Eq. (35), $c_B(i)$ is essentially a normalized measure of the density of common neighbors in the vicinity of node i.

$c_B(i)$ has also the following simple and intuitive similarity-based interpretation. Let A_j and A_l be the sets of neighbors of nodes j and l excluding node i. Then $m_{jl} - 1$ is the size of the intersection of A_j and A_l, $m_{jl} - 1 = \|A_j \bigcap A_l\|$, while $k_j + k_l - m_{jl} - 1 = \|A_j \bigcup A_l\|$ is their union. Therefore,

Eq. (35) can be written as

$$c_B(i) = \frac{\sum_{j \neq l} \|A_j \bigcap A_l\|}{\sum_{j \neq l} \|A_j \bigcup A_l\|}.$$ (36)

The ratio of the intersection and union of two sets is known as the Jaccard similarity coefficient [53]. $c_B(i)$ is given by the ratio of the sums of intersections and unions for all pairs of i's neighbors [Eq. (36)]. Therefore, $c_B(i)$ can be interpreted as a combined or effective Jaccard similarity of i's neighbors.

The average bipartite clustering coefficients $\overline{c}_B^t, \overline{c}_B^b$ for the top and bottom node domains can be written as

$$\overline{c}_B^t = \frac{1}{N} \sum_{i \in \Omega_t} c_B(i),$$ (37)

$$\overline{c}_B^b = \frac{1}{M} \sum_{j \in \Omega_b} c_B(j),$$ (38)

where Ω_t and Ω_b are the sets of all top and bottom nodes. Expressions for the expected bipartite clustering coefficients in the $\mathbb{S}^1 \times \mathbb{S}^1$ model are derived in Appendix F. Qualitatively, $\overline{c}_B^t, \overline{c}_B^b$ are large in the $\mathbb{S}^1 \times \mathbb{S}^1$ model and independent of the number of top and bottom nodes N, M, see Fig. 4(d). This result follows from the fact that the expected number of common neighbors $\overline{m}(\kappa_i, \kappa_j, \Delta\theta_{ij})$ between two nodes i, j is independent of the network size (Appendix E 1). In contrast, in the degree-preserving randomized counterparts of the modeled networks, the bipartite clustering coefficient is orders of magnitude smaller, and vanishes with the network size as $\overline{c}_B^{t,b} \sim N^{-\delta_{t,b}}$, with $\delta_{t,b} = (\gamma_{t,b} - 1)/2$ [see Fig. 4(d)], which is the expected behavior for uncorrelated bipartite networks [50]. A similar behavior holds for the real bipartite networks we consider [Fig. 4(d)].

Another important property of bipartite clustering coefficient in sf-sf networks is its self-similarity with respect to a degree-thresholding renormalization procedure [49]. Non-iterative removal of top and bottom nodes with degrees smaller than certain thresholds (k_T, ℓ_T) does not affect the functional form of degree-dependent bipartite clustering coefficients, which follow the same master-curve when plotted as a function of the node degree normalized by the average degree of the corresponding domain (see Fig. 9 and Section F).

Taken together, our results in this section indicate that the $\mathbb{S}^1 \times \mathbb{S}^1$ model can generate a variety of bipartite network topologies, whose main characteristics are consistent with those of real bipartite systems. A natural question then is whether it is possible to reverse the synthesis, and given a bipartite network, to infer the geometric coordinates and hidden variables of its nodes in a way congruent with the $\mathbb{S}^1 \times \mathbb{S}^1$ model. A tempting approach would be to first project the bipartite network onto one of its node domains, apply existing maximum-likelihood estimation techniques [54–56] to map the resulting one-mode projection, and then use the obtained unipartite map to infer the node coordinates of the other domain [48]. A necessary condition for this approach to work is that the geometry of the bipartite network is properly preserved in its one-mode projections. We next examine to what extend this is the case.

IV. ONE-MODE PROJECTIONS

In one-mode projections we project a bipartite network onto one of its node domains, such that nodes of the domain are connected if they have at least one common neighbor in the bipartite network. Even though one-mode projections allow one to study bipartite networks using tools developed for unipartite networks, projections can lead to significant loss of information and artificial inflation of the projected network with fully connected subgraphs. Historically, different approaches have been proposed to deal with the loss of information. One approach, for instance, is to weigh projected links using common neighbor statistics in the original network [57–60]. Another approach to quantify the extent at which information is lost in one-mode projections and to identify circumstances under which one-mode projections are still acceptable is to reduce noise by identifying and removing insignificant links in the projected network [61–64].

Here, we analyze the effects of one-mode projections in the context of the $\mathbb{S}^1 \times \mathbb{S}^1$ model. Specifically, we ask if the latent geometry of a bipartite network is preserved in its one-mode projections. Answering this question is important, as it can shed light on how well latent geometry beneath bipartite networks can be inferred using algorithms developed for unipartite networks such as those in Refs. [54–56]. In the following, we analyze the projections onto the top node domain. As before, the results for the bottom domain can be obtained by swapping the corresponding top and bottom domain variables.

The probability $r_u(i, j)$ that two top domain nodes i and j are connected in the one-mode projection is the probability that the nodes have at least one common neighbor in the bottom domain,

$$r_u(i, j) = 1 - \prod_{1 \leqslant k \leqslant M} [1 - r(\kappa_i, \theta_i; \lambda_k, \phi_k) r(\kappa_j, \theta_j; \lambda_k, \phi_k)],$$ (39)

where $k = 1 \ldots M$ enumerates the nodes of the bottom domain and $r(\kappa, \theta; \lambda, \phi)$ is the connection probability in Eq. (4).

We say that the latent geometry of the bipartite network is preserved in its one-mode projection if $r_u(i, j)$ preserves the functional form prescribed by Eq. (2):

$$r_u(i, j) = f\left(\frac{d_{ij}}{d_u(\kappa_i, \kappa_j)}\right),$$
$$d_{ij} \propto \Delta\theta_{ij},$$
$$d_u(\kappa_i, \kappa_j) \propto \overline{k}_u(\kappa_i)\overline{k}_u(\kappa_j) \propto \kappa_i\kappa_j,$$ (40)

where $f(x)$ is a monotone decreasing function of x, which may or may not coincide with our choice for $r(x)$ in Eq. (3), and $d_u(\kappa_i, \kappa_j)$ is the characteristic distance scale for a pair of nodes with κ_i and κ_j in the projected network. In the case $r_u(i, j)$ takes the form of Eq. (40) one could map projections of real bipartite networks to latent spaces using methods developed for unipartite networks. If, on the other hand, $r_u(i, j)$ is not in the form of Eq. (40), these techniques may not map correctly bipartite networks, and they either need to be adjusted or different techniques need to be developed.

To test if latent geometry is preserved in one-mode projections, we compute $r_u(i, j)$ below. We first note that since $r(\kappa_i, \theta_i; \lambda_k, \phi_k)$ and $r(\kappa_j, \theta_j; \lambda_k, \phi_k)$ depend on κ_i, θ_i and

$\kappa_j, \theta_j, r_u(i,j)$ also depends on $\kappa_i, \theta_i, \kappa_j, \theta_j$. Due to the uniform distribution of the $\phi_k, r_u(i,j)$ does not depend on the individual values of θ_i and θ_j per se but on the angular distance between the nodes, $\Delta\theta_{ij} = \pi - |\pi - |\theta_i - \theta_j||$. Thus, we can set, without loss of generality, $\theta_i = 0$ and $\theta_j = \Delta\theta_{ij}$. Assuming a sufficiently large number of bottom nodes M, we can rewrite Eq. (39) as

$$\ln[1 - r_u(i,j)] = \frac{M}{2\pi} \iint d\phi\, d\lambda\, \rho(\lambda)$$
$$\times \ln[1 - r(\kappa_i, 0; \lambda, \phi) r(\kappa_j, \Delta\theta_{ij}; \lambda, \phi)].$$
(41)

Then we replace the logarithm on the right-hand side of Eq. (41) with its Taylor series expansion,

$$-\ln[1 - r_u(i,j)] = \sum_{n=1}^{\infty} \frac{M}{2\pi n} \int d\lambda\, \rho(\lambda)$$
$$\times \int d\phi\, [r(\kappa_i, 0; \lambda, \phi) r(\kappa_j, \Delta\theta_{ij}; \lambda, \phi)]^n.$$
(42)

We note that the first term of the sum in the above relation, i.e., the term corresponding to $n = 1$, is the expected number of common neighbors between nodes i and j, $\overline{m}(\kappa_i, \kappa_j, \Delta\theta_{ij})$. Second, we perform the change of integration variable $x \equiv \frac{R\phi}{\mu\sqrt{\kappa_i\kappa_j}\lambda}$, to obtain

$$-\ln[1 - r_u(i,j)] \propto \sqrt{\kappa_i\kappa_j} \int \lambda\, d\lambda\, \rho(\lambda)$$
$$\times \sum_{n=1}^{\infty} \frac{1}{n} \int_{-\infty}^{\infty} dx \left[r\left(\sqrt{\frac{\kappa_j}{\kappa_i}}|x|\right) r\left(\sqrt{\frac{\kappa_i}{\kappa_j}}\left|x - \frac{R\Delta\theta_{ij}}{\mu\sqrt{\kappa_i\kappa_j}\lambda}\right|\right) \right]^n.$$
(43)

The leading contributions to the inner integrals in Eq. (43) come from the two maxima of each integrand at $x_1 = 0$ and $x_2 = \Delta\widetilde{\theta}_{ij} \equiv \frac{R\Delta\theta_{ij}}{\mu\sqrt{\kappa_i\kappa_j}\lambda}$. Specifically, for large $\Delta\widetilde{\theta}_{ij}$, connection probability $r_u(i,j)$ can be approximated, to the leading order, as

$$-\ln[1 - r_u(i,j)] \sim \frac{\overline{\lambda^{1+\beta}}}{M^\beta} \left[\frac{\Delta\theta_{ij}}{d_u(\kappa_i, \kappa_j)} \right]^{-\beta},$$
$$d_u(\kappa_i, \kappa_j) = \left(\kappa_i^\beta \kappa_j + \kappa_i \kappa_j^\beta \right)^{\frac{1}{\beta}},$$
(44)

where $\overline{\lambda^{1+\beta}} \equiv \int \lambda^{1+\beta} \rho(\lambda)\, d\lambda$ (Appendix G). The functional form of $r_u(i,j)$ in Eq. (44) clearly differs from that in Eq. (40) since the characteristic scale $d_u(\kappa_i, \kappa_j)$ differs, indicating that latent geometry is not preserved in one-mode projections, as one could intuitively expect.

At the same time, it is important to note certain similarity between one-mode projection connection probability $r_u(i,j)$ and original bipartite connection probability $r(x)$. Both are decreasing functions of the angular distance $\Delta\theta_{ij}$ normalized by characteristic scales $d_u(\kappa_i, \kappa_j)$, albeit these scales differ in the two cases. Yet since in both cases $d_u(\kappa_i, \kappa_j)$ is larger for pairs of nodes with larger hidden variables, nodes characterized by larger κ values are more likely to connect over large distances and, therefore, are expected to have larger degrees not only in the bipartite network but also in its one-mode projection,

consistent with our findings in Ref. [50]. Furthermore, as seen from Eq. (44), for sufficiently large $\frac{\Delta\theta_{ij}}{d_u(\kappa_i\kappa_j)}$ values $r_u(i,j)$ as a function of $\Delta\theta$ has the same asymptotic behavior as $r(x)$,

$$r_u(i,j) \approx -\ln[1 - r_u(i,j)] \sim \Delta\theta_{ij}^{-\beta}.$$
(45)

Our observation that latent geometry is not exactly preserved in one-mode projections is not specific to our choice of $r(x)$ as the connection probability function in the $\mathbb{S}^1 \times \mathbb{S}^1$ model. We show below that the latent geometry cannot be fully preserved in one-mode projections, regardless of the functional form of $r(x)$ in Eq. (3). Indeed, assuming that $r_u(i,j)$ is given by Eq. (40), we can write

$$-\ln[1 - r_u(i,j)] = g\left(\frac{\Delta\theta_{ij}}{\kappa_i\kappa_j} \right),$$
(46)

where $g(x) \equiv -\ln[1 - f(x)]$. Next, we observe that the right-hand side of Eq. (43) is a sum of convolutions and can be transformed into products of Fourier transforms, yielding

$$g(w) \propto \int d\lambda\, \lambda^2 \rho(\lambda) \sum_{n=1}^{\infty} \frac{1}{n} \left[r_n\left(\frac{\mu\omega\lambda}{R\kappa_i}\right) r_n\left(\frac{\mu\omega\lambda}{R\kappa_j}\right) \right], \quad (47)$$

where $g(w) \equiv \int_{-\infty}^{\infty} dx\, g(x) e^{iwx}$ and $r_n(w) \equiv \int_{-\infty}^{\infty} dx \times [r(|x|)]^n e^{iwx}$. Since the left-hand side of Eq. (47) does not depend on κ_i and κ_j while the right-hand side does, the only admissible solution is $r_n(w) = \text{const}$. This solution corresponds to $r(x) = \delta(x)$, where $\delta(x)$ is the Dirac δ function and cannot be interpreted as a connection probability function.

We thus find that latent geometry cannot be fully preserved for any functional form of the connection probability function $r(x)$. At the same time, our results indicate that connection probability in one-mode projections behaves similarly to that in the original network for large angular distances. This result implies that it may be possible to infer approximately the latent geometry of real bipartite networks from their one-mode projections. Yet it remains unclear how accurate such inferences can be, especially in small bipartite networks whose one-mode projections are overinflated with cliques of sizes comparable to the network size. Such problems can render geometry inference using one-mode projections highly inaccurate, especially in sf-sf networks with power-law exponents γ close to 2, as discussed at the end of Appendix G.

V. INFERRING LATENT GEOMETRY

We have seen that the $\mathbb{S}^1 \times \mathbb{S}^1$ model can construct synthetic bipartite networks that resemble real networks across a range of nontrivial structural characteristics, which include: (i) heterogeneity in distributions of node degrees for at least one of the two domains of nodes; (ii) power-law distribution of the number of common neighbors shared between pairs of nodes; and (iii) strong bipartite clustering. These results imply that we should be able to reverse the synthesis, and given a bipartite network, to infer the hidden variables of its nodes as well as their latent distances. Below, we show that hidden variables and latent distances can be estimated from the observed node degrees and the common neighbors shared by nodes, respectively.

Recall from Sec. III B that the resulting node degree k_i corresponding to a particular hidden variable κ_i is Poisson
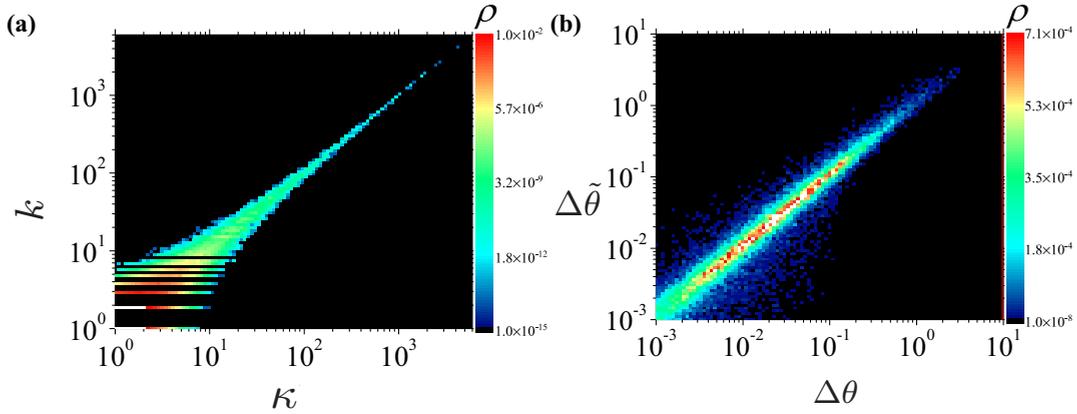
FIG. 5. Inferring hidden variables and latent distances from the network topology. The plots correspond to an sf-ps modeled network with $N = M = 10^5$ nodes, $\rho(\kappa) \sim \kappa^{-2.1}$, $\overline{\kappa} = 11$, $\rho(\lambda) = \delta(\lambda - 11.0)$, and $\beta = 2$. (a) The heat map displays the correlation between node degrees and hidden variables. The colors reflect the density of the nodes as shown in the legend. (b) The angular distances between the top nodes can be inferred by solving Eq. (52). The accuracy of the inference is higher for pairs of nodes that share a larger number of common neighbors. The figure shows the inferred angular distances $\Delta\tilde{\theta}$ as a function of the true angular distances $\Delta\theta$ for pairs of top nodes with $\kappa \geqslant 10$ and number of common neighbors $m \geqslant 5$. The colors in the heat map reflect the density of the nodes as shown in the legend.

distributed with expected value $\overline{k}_i(\kappa_i) = \kappa_i$. Thus, the node hidden variables can be estimated by the observed node degrees as

$$\kappa_i = \overline{k}_i(\kappa_i) \approx k_i. \tag{48}$$

Since the variance of the Poisson distribution is equal to its mean, this estimation works better for higher-degree nodes and can be used in the case of a scale-free degree distribution in the domain, see Fig. 5(a). In the case of a Poisson degree distribution in the domain, all nodes have identical hidden variables that can be estimated as

$$\kappa_i = \overline{\kappa} = \overline{k}, \tag{49}$$

where $\overline{k}$ is the observed average degree in the domain.

The angular distance separating two nodes can be estimated using the observed number of common neighbors between the nodes. As shown in Sec. III F, the number of common neighbors m_{12} shared by two nodes 1,2 is Poisson distributed with an expected value $\overline{m}_{12}(\kappa_1, \kappa_2, \Delta\theta_{12})$ given by Eq. (30), which depends on the nodes' hidden variables κ_1, κ_2 and their angular distance $\Delta\theta_{12}$. If the observed number of common neighbors m_{12} is sufficiently large, we can approximate $\overline{m}_{12}(\kappa_1, \kappa_2, \Delta\theta_{12})$ as

$$\overline{m}_{12}(\kappa_1, \kappa_2, \Delta\theta_{12}) \approx m_{12}. \tag{50}$$

The angular distance $\Delta\theta_{12}$ can be estimated using Eq. (50), which can be solved analytically for integer values of the model parameter $\beta \in (1, \infty)$ or numerically otherwise.

To test the accuracy of the proposed estimation we consider an sf-ps modeled network with $\beta = 2$. In this case, $\overline{m}_{12}$ for the top domain is given by

$$\overline{m}_{12}(\kappa_1, \kappa_2, \Delta\theta_{12}) \approx \frac{\kappa_1\kappa_2(\kappa_1 + \kappa_2)}{(\kappa_1 + \kappa_2)^2 + \left(\frac{M\Delta\theta_{12}}{2}\right)^2} \tag{51}$$

(see Appendix E 1), allowing us to estimate $\Delta\theta_{12}$ as

$$\Delta\theta_{12} \approx \frac{2}{M}\sqrt{\frac{k_1 k_2 (k_1 + k_2)}{m_{12}} - (k_1 + k_2)^2}. \tag{52}$$

We note that the above relation is an approximation and may yield angular distances outside the expected range $\Delta\theta_{12} \in [0, \pi]$ if m_{12} is too large or too small. This estimation procedure works well for pairs of nodes with large m_{12} values, see Fig. 5(b), and it can be used for a fast estimation of the pairwise latent similarity distances between such nodes, e.g., in recommender systems.

VI. DISCUSSION

Understanding the organizing principles determining the structure and evolution of real bipartite networks can lead to significant advances in many challenging problems including community detection [65–72], understanding signaling pathways in gene regulatory networks [73], multicast search [74], and construction of efficient recommender systems [75–78].

We have shown that three common properties of many real bipartite networks—heterogeneous degree distributions, power-law distributions of the number of common neighbors, and strong bipartite clustering—appear as natural reflections of latent geometric spaces underlying these networks, where nodes are points in these spaces, while connections preferentially occur at smaller distances. The distances between nodes in the latent space can be regarded as generalized similarity measures, arising from projections of properly weighted combinations of node attributes, controlling the appearance of links between node pairs.

For our analysis, we have used the simplest possible bipartite network model with latent geometry ($\mathbb{S}^1 \times \mathbb{S}^1$ model) [48]. Within the model, both the power-law distribution of the number of common neighbors and the strong bipartite clustering emerge naturally as reflections of the metric property, i.e., the triangle inequality, of the latent space. To achieve heterogeneous degree distributions we have assigned hidden variables to both top and bottom node domains, so nodes with larger hidden variables connect over larger distances with higher probability and, as a result, establish more connections than nodes with smaller hidden variables.

Although not fully geometric, the $\mathbb{S}^1 \times \mathbb{S}^1$ model is equivalent to the $\mathbb{H}^2 \times \mathbb{H}^2$ model in Appendix D, which is fully geometric, not using any hidden variables (other than node coordinates). In the $\mathbb{H}^2 \times \mathbb{H}^2$ model, heterogeneous degree distributions are consequences of the exponential expansion of space in $\mathbb{H}^2$, coupled with proper boundary conditions.

As with unipartite networks, a particularly pertinent question is the possibility to infer latent geometries underlying real bipartite systems. Through the analysis of one-mode projections, we have shown that latent geometry cannot be fully preserved but can be approximately preserved in one-mode projections of both sf-ps and sf-sf bipartite networks in the $\mathbb{S}^1 \times \mathbb{S}^1$ model. This result supports the possibility of inferring latent geometries underlying real bipartite systems by inferring the geometries of their one-mode projections using existing techniques [54–56], as in Ref. [48]. However, since geometry is not preserved exactly but only approximately, using one-mode projections can render geometry inference inaccurate, especially in smaller networks with weaker bipartite clustering. Such inaccuracies are particularly high in sf-sf networks with power-law exponents γ close to 2, calling for the development of proper methods to infer latent coordinates that do not use one-mode projections. We have shown that if instead of coordinates only pairwise latent distances between nodes with large numbers of common neighbors are to be inferred, e.g., in recommender systems [19–22] or in soft community detection [47,79], then such inferences can be made quickly and reliably based on the common neighbor statistics.

ACKNOWLEDGMENTS

APPENDIX A: REAL-WORLD BIPARTITE NETWORK DATA

Here we provide details on the considered real-wold bipartite networks.

(a) *Actor-film network (IMDb).* The bipartite actor-film network is constructed from the Internet Movie Database (IMDb) [7]. The network consists of two sets of nodes: actors (top nodes) and films (bottom nodes). Every actor is connected to all films in which she or he performed. We have excluded all films whose genre is labeled by IMDb as "Adult." The excluded "Adult" films represent a largely isolated subset of the original actor-film network. The considered network corresponds to the period from 1960 to 2010.

(b) *Condensed matter (Condmat) collaboration network.* The Condensed matter (*Condmat*) collaboration network is constructed from the arXiv e-Print archive [8]. The top and bottom nodes in the network are respectively authors and manuscripts published in the cond-mat section of the arXiv repository. Every manuscript is connected to all of its authors. The network includes all manuscripts published in the cond-mat section of the arXiv as of 2009.

(c) *HEP collaboration network.* The HEP collaboration network is also constructed from the arXiv e-Print archive [8]. The top and bottom nodes in the network are respectively authors and manuscripts published in the HEP section of the arXiv repository. Every manuscript is connected to all of its authors. The network includes all manuscripts published in the high-energy physics section of the arXiv as of 2009.

(d) *Metabolic Network.* The Metabolic network is based on the dataset of metabolic reactions of 107 organisms constructed by H. Ma and A.-P. Zeng [10]. The bipartite network consists of metabolic reactions (bottom nodes) and metabolites (top nodes) that are connected through these reactions.

(e) *Wikipedia.* The constructed bipartite network of Wikipedia [9] consists of users (top nodes) and articles (bottom nodes). Users are connected to all articles they have edited. In order to eliminate random edits, we connect users to articles which were edited by them at least 3 times. We consider users of the English Wikipedia who have created an account and have a user discussion page as of April 2, 2007. The data have been collected and processed in Ref. [80].

The basic topological properties of the considered networks are outlined in Table I and their degree distributions are shown in Fig. 6.

APPENDIX B: DEGREE-PRESERVING RANDOMIZATION

To randomize a bipartite network while preserving the node degrees, we employ a modification of the degree-preserving randomization algorithm for unipartite networks [2]. Specifically, we first remove all connections in the original network. Then, we assign to every node a number of open connections equal to its degree in the original network. Finally, we add links to the network one by one by connecting a randomly selected top-bottom node pair with open connections. Every time a new link is established, the number of open connections of the two nodes connected by the link are decreased by one. We note that the probability to select a particular node is

TABLE I. Basic properties of the considered real bipartite networks. N is the number of top nodes, M is the number of bottom nodes, E is the number of edges, $\bar{k}$ is the average top-node degree, and $\bar{\ell}$ is the average bottom-node degree.

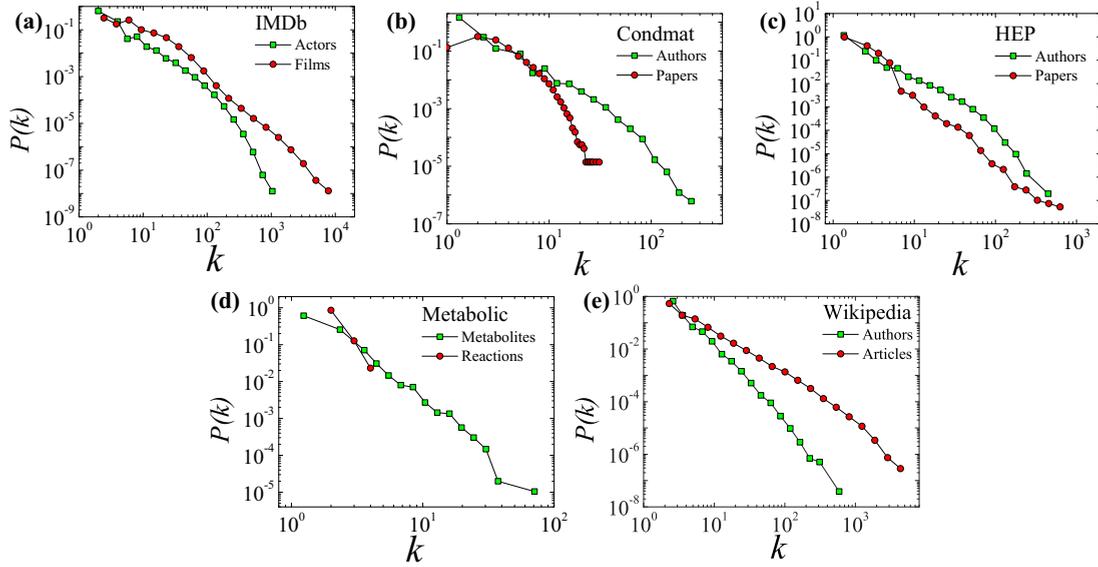| Network name	Type	N	M	E	$\bar{k}$	$\bar{\ell}$
Actor-film network (IMDb)	sf-sf	1 602 914	418 696	6 368 717	4.0	15.2
Condensed matter (Condmat) collaboration network	sf-ps	63 799	79 081	246 351	3.9	3.1
High Energy Physics (HEP) collaboration network	sf-sf	44 267	108 907	255 306	5.8	2.3
Metabolic network	sf-ps	2732	3568	7750	2.8	2.2
Wikipedia	sf-sf	45 875	407 543	874 942	19.1	2.1

FIG. 6. Degree distributions of top and bottom nodes of (a) IMDb, (b) Condmat collaboration network, (c) HEP collaboration network, (d) metabolic network, and (e) Wikipedia. Data have been binned logarithmically to reduce noise.

proportional to the number of its open connections. As a result, the randomized network is characterized by the same degree distribution as the original network and random degree-degree correlations.

APPENDIX C: BIPARTITE NETWORKS IN SPACES WITH BROKEN METRIC STRUCTURE

The key property of any metric space is the triangle inequality: for any triplet of points (A,B,C)

$$d(A,C) \leqslant d(A,B) + d(B,C), \tag{C1}$$

where $d(A,B)$ is the distance between points A and B. As explained in the main text, the triangle inequality in the underlying space leads to high clustering and power-law distribution of the number of common neighbors in the observed topology of the bipartite network.

To highlight the key role of the triangle inequality, here we consider bipartite networks built in spaces with broken metric structure. To this end, we construct a bipartite network following steps 1–3 of the $\mathbb{S}^1 \times \mathbb{S}^1$ model (Sec. III A of the main text), with the difference that before connecting top-bottom nodes we randomize their distances. Specifically, we first assign node hidden variables and coordinates following steps 1 and 2 of the $\mathbb{S}^1 \times \mathbb{S}^1$ model. Then we construct the distance matrix $\{d_{ij}\} = d(\theta_i, \phi_j)$ and randomize its values. To this end, we remove all distance values from the matrix and arrange them into a list of values. We then fill in the randomized matrix by drawing distance values from this list at random without replacement. Finally, we connect top-bottom nodes as in step 3 of the $\mathbb{S}^1 \times \mathbb{S}^1$ model using the randomized distance matrix $\{d_{ij}\}$.

The above modified model preserves the values of the node hidden variables and the distribution of distances between the nodes but breaks the triangle inequality. As seen from Fig. 7(a), the degree distribution in bipartite networks
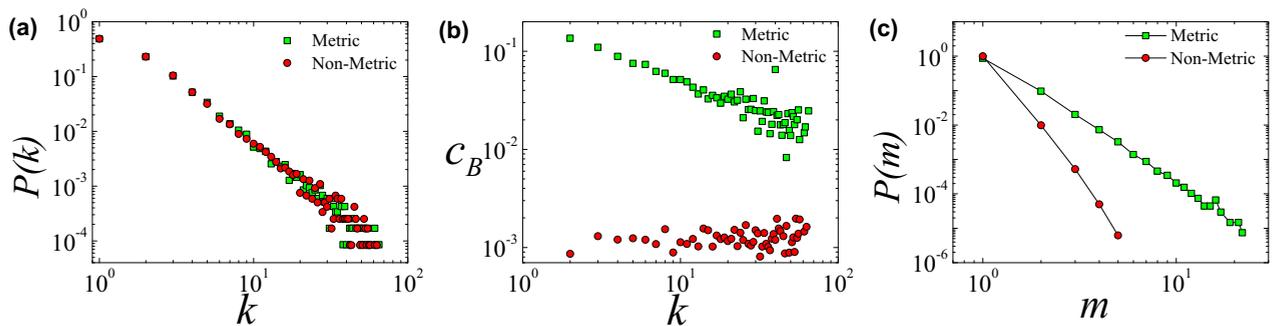

FIG. 7. Bipartite networks in metric spaces vs. spaces with broken metric structure. (a) Degree distribution; (b) average bipartite clustering coefficient as a function of node degree; and (c) distribution of the number of common neighbors. Both networks are sf-sf bipartite networks with $N = M = 10^5$, $\gamma_1 = \gamma_2 = 2.1$, $\bar{k} = \bar{\ell} = 11$, and $\beta = 2.0$. All plots correspond to the top node domain. The bottom node domain has similar properties.

constructed by the model with broken metric structure (non-metric) is the same as that constructed by the original $\mathbb{S}^1 \times \mathbb{S}^1$ model (metric). At the same time, however, the model with broken metric structure fails to reproduce the strong bipartite clustering coefficient and the fat-tail distribution of the number of common neighbors observed in the original $\mathbb{S}^1 \times \mathbb{S}^1$ model [Figs. 7(b) and 7(c)].

APPENDIX D: $\mathbb{H}^2 \times \mathbb{H}^2$ MODEL

In the main text we formulate the $\mathbb{S}^1 \times \mathbb{S}^1$ bipartite model on the Euclidean ring $\mathbb{S}^1$. The choice of $\mathbb{S}^1$ as the latent space is made primarily for simplicity, and generalizations to other spaces are possible. Here we formulate the latent space model using a two-dimensional hyperbolic disk of constant negative curvature $K = -\zeta^2 < 0$ and radius $R_H \gg 1$. In the model, nodes map to points within the hyperbolic disk: angular coordinates of both node types are chosen uniformly at random, while radial coordinates are drawn from corresponding pdf functions $\rho_t(r)$ and $\rho_b(r)$. Connections between top and bottom nodes are established with probabilities depending on distances between the nodes in the hyperbolic disk. The resulting $\mathbb{H}^2 \times \mathbb{H}^2$ model can be summarized as follows:

(1) Sample the angular coordinates of top nodes θ_i, $i = 1, 2, \ldots, N$, uniformly at random from $[0, 2\pi]$, and their radial coordinates r_i, $i = 1, 2, \ldots, N$, from the pdf $\rho_t(r)$ defined in $[0, R_H]$;

(2) Sample the angular coordinates of bottom nodes ϕ_j, $j = 1, 2, \ldots, M$, uniformly at random from $[0, 2\pi]$, and their radial coordinates r_j, $j = 1, 2, \ldots, M$, from the pdf $\rho_b(r)$ defined in $[0, R_H]$;

(3) Connect every top-bottom node pair with probability

$$p(r_i, \theta_i; r_j, \phi_j) = p(x_{ij}), \qquad \text{(D1)}$$

$$\cosh(\zeta x_{ij}) = \cosh(\zeta r_i)\cosh(\zeta r_j)$$
$$- \sinh(\zeta r_i)\sinh(\zeta r_j)\cos(\Delta\theta_{ij}). \quad \text{(D2)}$$

Unlike its counterpart, the $\mathbb{H}^2 \times \mathbb{H}^2$ model is purely geometric and desired degree distributions can be engineered with a proper choice of $\rho_{t,b}(r)$ and $\rho_{t,b}(\theta)$, without introducing external hidden variables. Yet, as we demonstrate below, both models are fully equivalent due to a one-to-one mapping between radial coordinates in the $\mathbb{H}^2 \times \mathbb{H}^2$ model and hidden variables in the $\mathbb{S}^1 \times \mathbb{S}^1$ model:

$$r_i = R_H - \frac{2}{\zeta}\ln(\kappa_i/\kappa_0), \qquad \text{(D3)}$$

$$r_j = R_H - \frac{2}{\zeta}\ln(\lambda_j/\lambda_0). \qquad \text{(D4)}$$

It is straightforward to verify that this mapping transforms the $\mathbb{S}^1 \times \mathbb{S}^1$ connection probability function into that of the $\mathbb{H}^2 \times \mathbb{H}^2$ model,

$$r\left(\frac{R\Delta\theta_{ij}}{\mu\kappa\lambda}\right) = p(\tilde{x}_{ij}), \qquad \text{(D5)}$$

where

$$\tilde{x}_{ij} \equiv r_i + r_j + \frac{2}{\zeta}\ln(\Delta\theta_{ij}/2) \qquad \text{(D6)}$$

is a close approximation for x_{ij} in Eq. (D2) for sufficiently large ζr_i, ζr_j, and $\Delta\theta_{ij} > 2\sqrt{\exp(-\zeta r_i) + \exp(-\zeta r_j)}$. Equation (D5) establishes the equivalence between the two models since any pair of nodes $\{i, j\}$ is equally likely to be connected or disconnected in both models.

Since the two models are equivalent, the $\mathbb{H}^2 \times \mathbb{H}^2$ model produces networks with the same topological properties as in the $\mathbb{S}^1 \times \mathbb{S}^1$ model. In particular, sf-sf bipartite networks can be generated using $\rho_{t,b}(r) \propto \exp(\alpha_{t,b} r)$ with $\alpha_{t,b} = \frac{\zeta}{2}(\gamma_{t,b} - 1)$ and sf-ps bipartite networks can be generated using $\rho_t(r) \propto \exp(\alpha_t r)$, $\alpha_t = \frac{\zeta}{2}(\gamma_t - 1)$, and $\rho_b(r) = \delta(r - R_H)$.

First, it is important to note that the correspondence between the $\mathbb{H}^2 \times \mathbb{H}^2$ and the $\mathbb{S}^1 \times \mathbb{S}^1$ formulations is approximate and holds only for a sufficiently large hyperbolic disk, $R_H \gg 1$. Indeed, when establishing the correspondence between the two models, we relied on the approximation (D6), which works well for large radial node coordinates. This is indeed the case for the majority of nodes in both sf-sf and sf-ps networks. In the case of sf-sf networks $\gamma_{t,b} > 1$, and therefore, $\alpha_{t,b} > 0$, implying that the majority of nodes is located near the periphery of the hyperbolic disk, $r_{i,j} \to R_H \gg 1$. The situation is the same for the top nodes of the sf-ps networks, while bottom nodes of sf-ps are all located at the edge of the hyperbolic disk, $r_j = R_H \gg 1$. Second, radial coordinates of all nodes must be positive, which is satisfied as long as $R_H \geqslant \frac{2}{\zeta}\ln(\kappa_{\max}/\kappa_0)$ and $R_H \geqslant \frac{2}{\zeta}\ln(\lambda_{\max}/\lambda_0)$, where $\kappa_{\max}$ and $\lambda_{\max}$ are the largest hidden variable values, respectively, in the top and bottom domains of the $\mathbb{S}^1 \times \mathbb{S}^1$ model. Since expected degree values in the $\mathbb{S}^1 \times \mathbb{S}^1$ model are proportional to their hidden variables, Eqs. (14) and (16), $\kappa_{\max} \sim N$ and $\lambda_{\max} \sim M$, leading to

$$R_H \sim \ln(N) \sim \ln(M). \qquad \text{(D7)}$$

APPENDIX E: NUMBER OF COMMON NEIGHBORS

Here we derive approximate analytic expressions for the average number of common neighbors and the distribution of the number of common neighbors in the $\mathbb{S}^1 \times \mathbb{S}^1$ model. To ease notation, we sometimes write the node hidden variables and angular coordinates in vectorlike form: $\mathbf{x_i} = \{\kappa_i, \theta_i\}$ and $\mathbf{y_j} = \{\lambda_j, \phi_j\}$. In this case, we define the probability distribution $\rho(\mathbf{x})$ ($\rho(\mathbf{y})$) as the probability that a randomly chosen top (bottom) node is characterized by $\mathbf{x} \equiv \{\kappa, \theta\}$ ($\mathbf{y} \equiv \{\lambda, \phi\}$). Since the hidden variables are independent of the angular coordinates, $\rho(\mathbf{x}) = \rho(\kappa)\rho(\theta)$ and $\rho(\mathbf{y}) = \rho(\lambda)\rho(\phi)$. Thus, the integration (summation) over $\mathbf{x}$ or $\mathbf{y}$ corresponds to integration (summation) over both attributes: $\int d\mathbf{x} \equiv \int d\kappa \int d\theta$, $\int d\mathbf{y} \equiv \int d\lambda \int d\phi$.

1. Average number of common neighbors

The average number of common neighbors between two nodes of the same domain can be calculated from Eq. (30) in the main text. By substituting the connection probability function from Eq. (4) into Eq. (30), we obtain that for large

networks

$$\overline{m}(\kappa_1,\kappa_2,\Delta\theta_{12}) = \frac{\sqrt{\kappa_1\kappa_2}}{I\overline{\lambda}} \int \lambda\,d\lambda\,\rho(\lambda) \int_{-\infty}^{+\infty} dx\, r\left(\sqrt{\frac{\kappa_2}{\kappa_1}}|x|\right) r\left(\sqrt{\frac{\kappa_1}{\kappa_2}}|x-\Delta\widetilde{\theta}_{12}|\right), \tag{E1}$$

where $\Delta\widetilde{\theta}_{12} \equiv NI\overline{k}\Delta\theta_{12}/(\pi\sqrt{\kappa_1\kappa_2}\lambda)$, $\Delta\theta_{12} = \pi - |\pi - |\theta_1 - \theta_2||$, and I is given by Eq. (9). It can be seen from Eq. (E1) that regardless of the functional form of $r(x)$, $\overline{m}(\kappa_1,\kappa_2,\Delta\theta_{12})$ depends on the distance between the nodes $R\Delta\theta_{12} \propto N\Delta\theta_{12}$ and not on domain size N, M per se.

The inner integral of Eq. (E1) can be taken by residues. However, the number of poles of the integrand depends on β. A relatively simple solution exists for $\beta = 2$. Specifically, in this case we get

$$\overline{m}(\kappa_1,\kappa_2,\Delta\theta_{12}) \approx \frac{\kappa_1\kappa_2(\kappa_1+\kappa_2)}{(\kappa_1+\kappa_2)^2 + \left(\frac{M\Delta\theta_{12}}{2}\right)^2} \tag{E2}$$

for sf-ps bipartite networks and

$$\overline{m}(\kappa_1,\kappa_2,\Delta\theta_{12})$$
$$\approx \frac{\kappa_1\kappa_2}{(\kappa_1+\kappa_2)}{}_2F_1\left[1,\frac{\gamma_b-2}{2},\frac{\gamma_b}{2},-\left(\frac{N\overline{k}\Delta\theta_{12}}{2\lambda_0(\kappa_1+\kappa_2)}\right)^2\right] \tag{E3}$$

for sf-sf bipartite networks with $\rho(\lambda) = (\gamma_b - 1)\lambda_0^{\gamma_b-1}\lambda^{-\gamma_b}$, $\lambda \in [\lambda_0,\infty)$, where γ_b is the power-law degree distribution exponent of the bottom domain and λ_0 is the expected minimum bottom domain degree. The function ${}_2F_1$ is the Gauss hypergeometric function.

2. Distribution of the number of common neighbors

We now consider the distribution of the number of common neighbors. First, we calculate the probability that two nodes with hidden variables κ_1 and κ_2 and angular coordinates θ_1 and θ_2 have m common neighbors. To this end, we define the probability $p_{\mathbf{x_1},\mathbf{x_2}}(m|\mathbf{y})$ that two top nodes with vectors $\mathbf{x_1} = \{\kappa_1,\theta_1\}$ and $\mathbf{x_2} = \{\kappa_2,\theta_2\}$ have m common neighbors among bottom nodes with vector $\mathbf{y} = \{\lambda,\phi\}$. Since pairs of nodes are connected independently with probability $r(\mathbf{x};\mathbf{y})$, $p_{\mathbf{x_1},\mathbf{x_2}}(m|\mathbf{y})$ is given by the binomial distribution

$$p_{\mathbf{x_1},\mathbf{x_2}}(m|\mathbf{y}) = C_m^{M(\mathbf{y})}[r(\mathbf{x_1};\mathbf{y})r(\mathbf{x_2};\mathbf{y})]^m$$
$$\times [1 - r(\mathbf{x_1};\mathbf{y})r(\mathbf{x_2};\mathbf{y})]^{M(\mathbf{y})-m}, \tag{E4}$$

where $M(\mathbf{y})$ is the total number of bottom nodes with vector $\mathbf{y}$ and $C_l^k = \binom{k}{l}$ is the binomial coefficient. The probability that two nodes with vectors $\mathbf{x_1}$ and $\mathbf{x_2}$ have m common neighbors can be calculated as

$$P_{\mathbf{x_1},\mathbf{x_2}}(m) = \sum_{\sum m_i = m} \prod_i p_{\mathbf{x_1},\mathbf{x_2}}(m_i|\mathbf{y_i}), \tag{E5}$$

where the product goes over the entire set of the $\mathbf{y_i}$ vectors, and the summation is performed over all possible combinations of m_i such that their sum equals m.

Next, we define the generating functions [81] for $p_{\mathbf{x_1},\mathbf{x_2}}(m_i|\mathbf{y_i})$ and $P_{\mathbf{x_1},\mathbf{x_2}}(m)$,

$$\widehat{p}_{\mathbf{x_1},\mathbf{x_2}}(z|\mathbf{y}) \equiv \sum_m z^m p_{\mathbf{x_1},\mathbf{x_2}}(m|\mathbf{y}), \tag{E6}$$

$$\widehat{P}_{\mathbf{x_1},\mathbf{x_2}}(z) \equiv \sum_m z^m P_{\mathbf{x_1},\mathbf{x_2}}(m). \tag{E7}$$

Since $P_{\mathbf{x_1},\mathbf{x_2}}(m)$ is given by the convolution of $p_{\mathbf{x_1},\mathbf{x_2}}(m_i|\mathbf{y_i})$ [Eq. (E5)], its generating function can be expressed as the product of the generating functions of the $p_{\mathbf{x_1},\mathbf{x_2}}(m_i|\mathbf{y_i})$,

$$\widehat{P}_{\mathbf{x_1},\mathbf{x_2}}(z) = \prod_i \widehat{p}_{\mathbf{x_1},\mathbf{x_2}}(z|\mathbf{y_i}). \tag{E8}$$

It is easy to see that $\widehat{p}_{\mathbf{x_1},\mathbf{x_2}}(z|\mathbf{y}) = [1 - (1-z)r(\mathbf{x_1};\mathbf{y}) r(\mathbf{x_2};\mathbf{y})]^{M(\mathbf{y})}$. By substituting this expression into Eq. (E8) and taking the logarithm of both sides, we get

$$\ln[\widehat{P}_{\mathbf{x_1},\mathbf{x_2}}(z)] = M \int d\mathbf{y}\,\rho(\mathbf{y})\ln[1-(1-z)r(\mathbf{x_1};\mathbf{y})r(\mathbf{x_2};\mathbf{y})]. \tag{E9}$$

One can verify that the expression for the average number of common neighbors $\overline{m}(\mathbf{x_1};\mathbf{x_2})$ [Eq. (30) in the main text] can be obtained by evaluating the derivative of $\widehat{P}_{\mathbf{x_1},\mathbf{x_2}}(z)$ with respect to z at $z = 1$,

$$\overline{m}(\mathbf{x_1};\mathbf{x_2}) = M \int d\mathbf{y}\,\rho(\mathbf{y})r(\mathbf{x_1};\mathbf{y})r(\mathbf{x_2};\mathbf{y}), \tag{E10}$$

as expected. In the case of sparse bipartite networks, Eq. (E9) can be approximated as (see Ref. [50])

$$\ln[\widehat{P}_{\mathbf{x_1},\mathbf{x_2}}(z)] \approx (z-1)\overline{m}(\mathbf{x_1};\mathbf{x_2}), \tag{E11}$$

$$P_{\mathbf{x_1},\mathbf{x_2}}(m) \approx e^{-\overline{m}(\mathbf{x_1};\mathbf{x_2})}[\overline{m}(\mathbf{x_1};\mathbf{x_2})]^m/m!, \tag{E12}$$

proving Eq. (32) in the main text.

The distribution of the number of common neighbors $P(m)$ is obtained by averaging $P_{\mathbf{x_1},\mathbf{x_2}}(m)$ over $\mathbf{x_1}$ and $\mathbf{x_2}$,

$$P(m) = \iint d\mathbf{x_1}\,d\mathbf{x_2}\,\rho(\mathbf{x_1})\rho(\mathbf{x_2})P_{\mathbf{x_1},\mathbf{x_2}}(m). \tag{E13}$$

While in general there is no closed-form solution to Eq. (E13), different closed-form solutions can be obtained for integer values of β. Below, we derive $P(m)$ in the case of sf-ps networks with $\beta = 2$. In this case, $\overline{m}(\mathbf{x_1},\mathbf{x_2}) = \overline{m}(\kappa_1,\kappa_2,\Delta\theta_{12})$ is given by Eq. (E2). The first step towards computing $P(m)$ is to evaluate $P_{\kappa_1,\kappa_2}(m)$ by averaging $P_{\mathbf{x_1},\mathbf{x_2}}(m)$ in Eq. (E12)

over $\Delta\theta_{12}$,

$$P_{\kappa_1,\kappa_2}(m) \equiv \frac{1}{\pi}\int_0^\pi d\Delta\theta_{12}\,P_{\mathbf{x_1},\mathbf{x_2}}(m)$$

$$\propto \frac{(\kappa_1+\kappa_2)}{\sqrt{m}\,m!}e^{-\frac{\kappa_1\kappa_2}{\kappa_1+\kappa_2}}\left[\frac{\kappa_1\kappa_2}{\kappa_1+\kappa_2}\right]^m. \quad (E14)$$

The second step is to average $P_{\kappa_1,\kappa_2}(m)$ over all κ_1,κ_2 values. In the case of sf-ps networks $\rho(\kappa)\sim\kappa^{-\gamma}$, and after a series of straightforward but tedious steps, we obtain

$$P(m) = \iint d\kappa_1\,\rho(\kappa_1)\,d\kappa_2\,\rho(\kappa_2)P_{\kappa_1,\kappa_2}(m)$$

$$\propto \frac{\Gamma[m+3-2\gamma,\kappa_0]}{m^{1/2}\Gamma[m+1]} \sim m^{3/2-2\gamma}, \quad (E15)$$

proving Eq. (34) in the main text.

The analytical evaluation of $P(m)$ for arbitrary β values is intractable due to the nontrivial dependence of $\overline{m}(\mathbf{x_1};\mathbf{x_2})$ on β. However, we can verify numerically that $P(m)$ is power-law distributed for different β values, similarly to real-world bipartite networks, see Fig. 8.

APPENDIX F: BIPARTITE CLUSTERING COEFFICIENT

The expected value of the bipartite clustering coefficient can be derived using the hidden variable formalism for bipartite networks [50]. In particular, the expected bipartite clustering coefficient $\overline{c_B}(\mathbf{x})$ of a top node with vector $\mathbf{x}=\{\kappa,\theta\}$ is given by

$$\overline{c_B}(\mathbf{x}) = \frac{\iint d\mathbf{y_1}\,d\mathbf{y_2}\,P(\mathbf{y_1}|\mathbf{x})P(\mathbf{y_2}|\mathbf{x})\overline{m}(\mathbf{y_1};\mathbf{y_2})}{2\int d\mathbf{y}\,P(\mathbf{y}|\mathbf{x})\overline{\ell}(\mathbf{y}) - \iint d\mathbf{y_1}\,d\mathbf{y_2}\,P(\mathbf{y_1}|\mathbf{x})P(\mathbf{y_2}|\mathbf{x})\overline{m}(\mathbf{y_1};\mathbf{y_2})}, \quad (F1)$$

where $P(\mathbf{y}|\mathbf{x})$ is the conditional probability that a node with vector $\mathbf{x}$ is connected to a node with vector $\mathbf{y}=\{\lambda,\phi\}$; $\overline{m}(\mathbf{y_1};\mathbf{y_2})$ is the average number of common neighbors between nodes with vectors $\mathbf{y_1},\mathbf{y_2}$; and $\overline{\ell}(\mathbf{y})\equiv N\int r(\mathbf{x}';\mathbf{y})\rho(\mathbf{x}')d\mathbf{x}'$ is the expected degree of a bottom node with vector $\mathbf{y}$. The expected number of common neighbors $\overline{m}(\mathbf{y_1};\mathbf{y_2})$ can be obtained from Eq. (30) in the main text by swapping variables (κ,θ) with (λ,ϕ), while $P(\mathbf{y}|\mathbf{x})$ is given by

$$P(\mathbf{y}|\mathbf{x}) = M\frac{\rho(\mathbf{y})r(\mathbf{x};\mathbf{y})}{\bar{k}(\mathbf{x})} = \frac{\rho(\mathbf{y})r(\mathbf{x};\mathbf{y})}{\int r(\mathbf{x};\mathbf{y}')\rho(\mathbf{y}')\,d\mathbf{y}'}, \quad (F2)$$

where $\bar{k}(\mathbf{x})\equiv M\int r(\mathbf{x};\mathbf{y}')\rho(\mathbf{y}')\,d\mathbf{y}'$ is the expected degree of a top node with vector $\mathbf{x}$. We note that, due to the rotational symmetry of the model, $\overline{c_B}(\mathbf{x})$ depends only on the node's hidden variable κ and not on its angular position θ, that is, $\overline{c_B}(\mathbf{x})\equiv\overline{c_B}(\kappa)$. We also note that $\overline{c_B}(\mathbf{x})$ is independent of the network size ($N\propto M$) since neither $P(\mathbf{y}|\mathbf{x})$ nor $\overline{m}(\mathbf{y_1};\mathbf{y_2})$ depend on it. Furthermore, N in $\overline{\ell}(\mathbf{y}) = N\int r(\mathbf{x}';\mathbf{y})\rho(\mathbf{x}')\,d\mathbf{x}'$ cancels out after performing the integration $\int r(\mathbf{x}';\mathbf{y})\rho(\mathbf{x}')\,d\mathbf{x}'$ that yields a factor $1/R\sim 1/N$, so $\overline{\ell}(\mathbf{y})$ in the denominator of Eq. (F1) does not depend on the network size either.

Finally, the average bipartite clustering of the top domain can be obtained by averaging the bipartite clustering coefficient over all nodes in the domain,

$$\bar{c}_B^t = \int \overline{c_B}(\mathbf{x})\rho(\mathbf{x})\,d\mathbf{x}, \quad (F3)$$

and is also independent of network size. Similar results hold for the bottom node domain.

An important property of real sf-sf bipartite networks and the $\mathbb{S}^1\times\mathbb{S}^1$ model is self-similarity of bipartite clustering coefficient with respect to a degree-thresholding renormalization procedure. Noniterative removal of top and bottom nodes with degrees smaller than certain thresholds (k_T,ℓ_T) does not affect bipartite clustering coefficient. In contrast, bipartite clustering coefficient of randomized networks increases as the

threshold increases [Fig. 9(a)]. Moreover degree-dependent bipartite clustering coefficients preserve their functional form, following the same master curve when plotted as a function of the node degree normalized by the average degree of the corresponding domain:

$$c_B^t(k|k_T,\ell_T) = f_t\left(\frac{k}{\bar{k}(k_T,\ell_T)}\right), \quad (F4)$$

$$c_B^b(\ell|k_T,\ell_T) = f_b\left(\frac{\ell}{\bar{\ell}(k_T,\ell_T)}\right), \quad (F5)$$

see Fig. 9(b). Here $f_{t,b}(x)$ are the master curves for the top and bottom domain degree-dependent clustering coefficients, while $\bar{k}(k_T,\ell_T)$ and $\bar{\ell}(k_T,\ell_T)$ are the average degrees of top and bottom domains after degree thresholding. In contrast, as seen from Fig. 9(c), bipartite clustering coefficients of randomized versions of real and modeled bipartite networks are not self-similar.

Since expected node degrees in the $\mathbb{S}^1\times\mathbb{S}^1$ model are tuned to be equal to their hidden variables, Eqs. (14) and (16), the degree thresholding procedure is equivalent to removal of nodes with hidden variables less than threshold (k_T,ℓ_T). Then, the number of top and bottom nodes in thresholded network $G(k_T,\ell_T)$ decreases as

$$N(k_T) = N\left(\frac{k_T}{\kappa_0}\right)^{1-\gamma_t}, \quad M(\ell_T) = M\left(\frac{\ell_T}{\lambda_0}\right)^{1-\gamma_b}, \quad (F6)$$

where N and M are the number of nodes in original network G. Power-law distributions of hidden variables $\rho(\kappa)$ and $\rho(\lambda)$ of nodes in $G(k_T,\ell_T)$ are unchanged but start at k_T and ℓ_T, respectively,

$$\rho(\kappa|k_T) = (\gamma_t-1)k_T^{\gamma_t-1}\kappa^{-\gamma_t}, \quad (F7)$$

$$\rho(\lambda|\ell_T) = (\gamma_b-1)\ell_T^{\gamma_b-1}\lambda^{-\gamma_b}. \quad (F8)$$
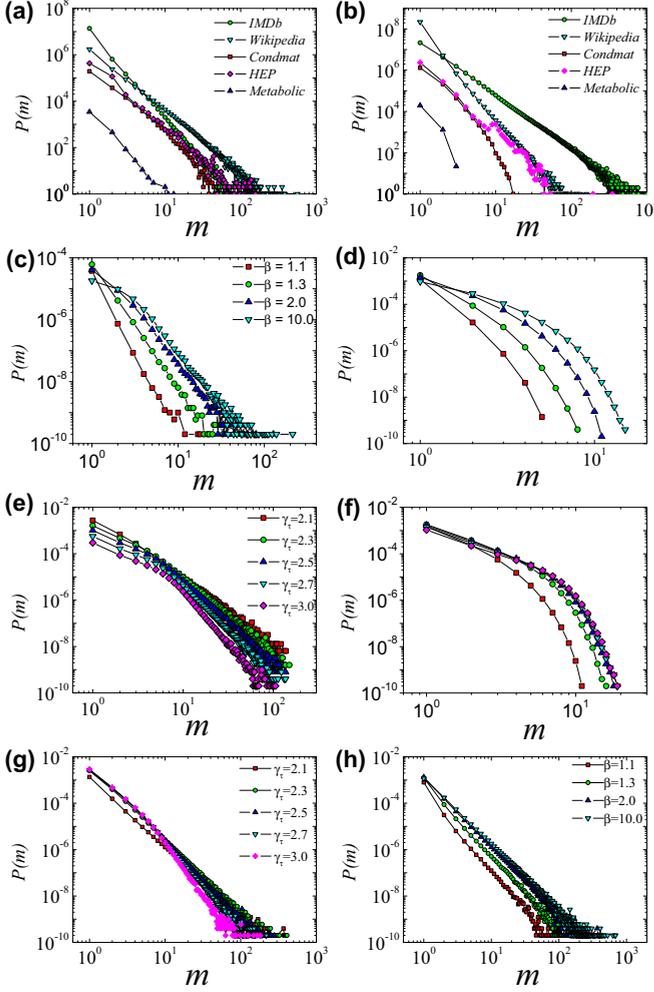
FIG. 8. Distribution of the number of common neighbors $P(m)$ in real [(a) and (b)] and modeled [(c)–(h)] bipartite networks. The properties of the real networks are outlined in Table I. Panels (a), (c), (e), (g), and (h) correspond to pairs of nodes in top domains, while panels (b), (d), and (f) correspond to pairs of nodes in bottom domains. The distributions $P(m)$ are not normalized for the real networks in panels (a) and (b) (i.e., they correspond to number of nodes instead of percentage of nodes) to improve visibility. All modeled networks are characterized by $N = M = 10^5$. Panels (c) and (d) correspond to sf-ps networks with top-domain power-law exponent $\gamma_t = 2.5$, average top-domain degree $\bar{k} = 3.0$, and different values of β. Panels (e) and (f) correspond to sf-ps networks with $\beta = 2.0$ and different values of γ_t, where $\bar{k}$ is chosen as $\bar{k} = \frac{\gamma_t - 1}{\gamma_t - 2}$. Panels (g) and (h) correspond to sf-sf networks. Panel (g) depicts $P(m)$ for the top domain of networks that have different γ_t values, $\bar{k} = 10$, bottom-domain power-law exponent $\gamma_b = 2.5$, and $\beta = 2.0$. Panel (h) depicts $P(m)$ for the top domain of networks that have $\gamma_t = 2.1$, $\bar{k} = 10$, $\gamma_b = 2.5$, and different values of β.

As a result, expected degrees of individual nodes and network domains scale as

$$\bar{k}(\kappa | k_T, \ell_T) = \left(\frac{\ell_T}{\lambda_0} \right)^{2-\gamma_b} \kappa, \tag{F9}$$

$$\bar{\ell}(\lambda | k_T, \ell_T) = \left(\frac{k_T}{\kappa_0} \right)^{2-\gamma_t} \lambda, \tag{F10}$$

$$\bar{k}(k_T, \ell_T) = \left(\frac{k_T}{\kappa_0} \right) \left(\frac{\ell_T}{\lambda_0} \right)^{2-\gamma_b} \bar{k}, \tag{F11}$$

$$\bar{\ell}(k_T, \ell_T) = \left(\frac{\ell_T}{\lambda_0} \right) \left(\frac{k_T}{\kappa_0} \right)^{2-\gamma_t} \bar{\ell}. \tag{F12}$$

Scaling relationships for degree-dependent bipartite clustering coefficients follow from (F1). Through a number of straightforward but tedious steps, which we omit here for brevity, we obtain

$$\bar{c}_B^t(\kappa | k_T, \ell_T) = \widetilde{f}_t(\kappa / k_T) = f_t\left(\frac{k}{\bar{k}(k_T, \ell_T)} \right), \tag{F13}$$

$$\bar{c}_B^b(\lambda | k_T, \ell_T) = \widetilde{f}_b(\lambda / \ell_T) = f_b\left(\frac{\ell}{\bar{\ell}(k_T, \ell_T)} \right), \tag{F14}$$

proving the self-similarity of sf-sf $\mathbb{S}^1 \times \mathbb{S}^1$ models. Finally, it is now easy to check, using Eqs. (F13) and (F14), that the average bipartite clustering coefficients of the top and bottom node domains take finite values independent of the threshold (k_T, ℓ_T):

$$\bar{c}_B^t(k_T, \ell_T) = \int d\kappa \, \rho(\kappa) \bar{c}_B^t(\kappa | k_T, \ell_T) = \bar{c}_B^t, \tag{F15}$$

$$\bar{c}_B^b(k_T, \ell_T) = \int d\lambda \, \rho(\lambda) \bar{c}_B^b(\kappa | k_T, \ell_T) = \bar{c}_B^b. \tag{F16}$$

Unlike in sf-sf networks, bipartite clustering in sf-ps networks is self-similar only with respect to degree thresholding of the scale-free domain, (k_T, λ_0).

APPENDIX G: ONE-MODE PROJECTIONS

Here we fill in the details on the estimation of the effective connection probability between nodes in one-mode projections. We consider the top node domain—similar results hold for the bottom domain. Any two nodes i, j in the top domain are connected in their one-mode projection if they share at least one neighbor in the original bipartite network. Equation (43) in the main text allows us to estimate (up to a proportionality coefficient) the connection probability of the two nodes in the one-mode projection, $r_u(i, j)$, as

$$-\ln[1 - r_u(i,j)] \propto \sqrt{\kappa_i \kappa_j} \int \lambda \, d\lambda \, \rho(\lambda) \sum_{n=1}^{\infty} \frac{1}{n} \int_{-\infty}^{\infty} dx \left[r\left(\sqrt{\frac{\kappa_j}{\kappa_i}} |x| \right) r\left(\sqrt{\frac{\kappa_i}{\kappa_j}} \left| x - \frac{R\Delta\theta_{ij}}{\mu\sqrt{\kappa_i\kappa_j}\lambda} \right| \right) \right]^n. \tag{G1}$$

Note that the first term in the sum of Eq. (G1), i.e., the term with $n = 1$, is proportional to the expected number of common neighbors between nodes i and j. The inner integral in Eq. (G1) has two maxima at $x = 0$ and $x = \Delta\widetilde{\theta}_{ij} \equiv \frac{R\Delta\theta_{ij}}{\mu\sqrt{\kappa_i\kappa_j}}$. Below, we estimate $r_u(i, j)$ for large $\Delta\widetilde{\theta}_{ij}$ (recall that $R \propto N \propto M$). To evaluate Eq. (G1), we split the integration interval into the five
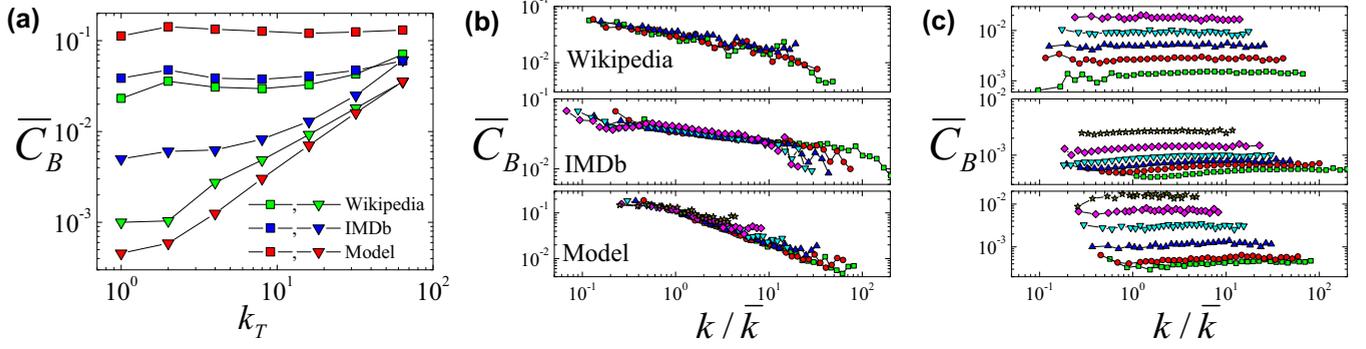
FIG. 9. Bipartite clustering is self-similar under degree-thresholding renormalization. (a) Average bipartite clustering coefficient, $\overline{c}^t_B$, as a function of the threshold parameters $k_T = \ell_T$ for the Wikipedia network (green squares), the IMDb network (blue squares), and an sf-sf modeled network (red squares) with $N = M = 10^5$, $\gamma_t = \gamma_b = 2.5$, $\overline{k} = \overline{\ell} = 3.0$, and $\beta = 2.0$. Shown with triangles are the results for the corresponding degree-preserving randomized versions of these networks. Note that c^t_B is nearly independent of k_T in the real and modeled networks, while in their degree-preserving randomized versions it increases with k_T. (b) Average bipartite clustering of top nodes as a function of their normalized degree for the same real and modeled networks as in plot (a), calculated for different values of thresholds $k_T = \ell_T$. (c) Average bipartite clustering of top nodes as a function of their normalized degree for the degree-preserving randomized versions of the networks in plot (b), calculated for different threshold values $k_T = \ell_T$.

subintervals: $(-\infty, -\delta)$, $(-\delta, \delta)$, $(\delta, \Delta\widetilde{\theta}_{ij} - \delta)$, $(\Delta\widetilde{\theta}_{ij} - \delta, \Delta\widetilde{\theta}_{ij} + \delta)$, and $(\Delta\widetilde{\theta}_{ij} + \delta, \infty)$ by choosing $\delta > 0$ such that $\delta \ll \Delta\widetilde{\theta}_{ij}$ and $\delta \sim \Delta\widetilde{\theta}_{ij}$, i.e.,

$$\delta = c\Delta\widetilde{\theta}_{ij}, \tag{G2}$$

where constant $c \ll 1$.

We start by evaluating the inner integral for the nth term of Eq. (G1) in the second subregion. We denote this integral by $\mathcal{J}_2$. To ease notation, we drop the indices i, j from the angular distance, i.e., $\Delta\theta_{ij} \equiv \Delta\theta$, $\Delta\widetilde{\theta}_{ij} \equiv \Delta\widetilde{\theta}$. When $\Delta\widetilde{\theta} \gg \delta$, within the integration interval $|x - \Delta\widetilde{\theta}| \approx \Delta\widetilde{\theta}$, and $\mathcal{J}_2$ can be approximated as

$$\mathcal{J}_2 = \int \lambda \, d\lambda \, \rho(\lambda) \int_{-\delta}^{\delta} dx \left[\frac{1}{1 + \left(\sqrt{\frac{\kappa_j}{\kappa_i}} |x|\right)^\beta} \right]^n \left[\frac{1}{1 + \left(\sqrt{\frac{\kappa_i}{\kappa_j}} |x - \Delta\widetilde{\theta}|\right)^\beta} \right]^n \approx \frac{2\overline{\lambda^{1+n\beta}} \sqrt{\frac{\kappa_i}{\kappa_j}} I_n\left(\beta; \delta\sqrt{\frac{\kappa_j}{\kappa_i}}\right)}{\left(\frac{R\Delta\theta}{\mu\kappa_j}\right)^{n\beta}}, \tag{G3}$$

where $I_n(\beta; a) \equiv \int_0^a \frac{dx}{(1+x^\beta)^n}$ and $\overline{\lambda^{1+n\beta}} \equiv \int \lambda^{1+n\beta} \rho(\lambda) \, d\lambda$. $I_n(\beta; a)$ increases as a function of a and it is bounded from above as

$$I_n(\beta; a) \leqslant I_n(\beta; \infty) = \frac{\Gamma[n - 1/\beta]\Gamma[1 + 1/\beta]}{\Gamma[n]}, \tag{G4}$$

where $\Gamma[x]$ is the γ function. Depending on $\rho(\lambda)$, $\overline{\lambda^{1+n\beta}}$ may depend on M. Since $R \propto M$ and the other terms in Eq. (G4) are independent of the network size, we can write

$$\mathcal{J}_2 \sim \overline{\lambda^{1+n\beta}} M^{-n\beta}. \tag{G5}$$

The evaluation of the integral in the fourth subregion, $\mathcal{J}_4$, is the same as $\mathcal{J}_2$ but with the κ_i and κ_j variables swapped,

$$\mathcal{J}_4 = \int \lambda\rho(\lambda) \, d\lambda \int_{\Delta\widetilde{\theta}-\delta}^{\Delta\widetilde{\theta}+\delta} dx \left[\frac{1}{1 + \left(\sqrt{\frac{\kappa_j}{\kappa_i}} |x|\right)^\beta} \right]^n \left[\frac{1}{1 + \left(\sqrt{\frac{\kappa_i}{\kappa_j}} |x - \Delta\widetilde{\theta}|\right)^\beta} \right]^n \approx \frac{2\overline{\lambda^{1+n\beta}} \sqrt{\frac{\kappa_j}{\kappa_i}} I_n\left(\beta; \delta\sqrt{\frac{\kappa_i}{\kappa_j}}\right)}{\left(\frac{R\Delta\theta}{\mu\kappa_i}\right)^{n\beta}}. \tag{G6}$$

Therefore, similarly to $\mathcal{J}_2$,

$$\mathcal{J}_4 \sim \overline{\lambda^{1+n\beta}} M^{-n\beta}. \tag{G7}$$

The integrals in the fifth and third subregions, $\mathcal{J}_5$, $\mathcal{J}_3$, are evaluated below

$$\mathcal{J}_5 = \int \lambda\rho(\lambda) \, d\lambda \int_{\Delta\widetilde{\theta}+\delta}^{\infty} dx \left[\frac{1}{1 + \left(\sqrt{\frac{\kappa_j}{\kappa_i}} |x|\right)^\beta} \right]^n \left[\frac{1}{1 + \left(\sqrt{\frac{\kappa_i}{\kappa_j}} |x - \Delta\widetilde{\theta}|\right)^\beta} \right]^n$$

$$\propto \int \lambda\rho(\lambda) \, d\lambda \int_{\Delta\widetilde{\theta}+\delta}^{\infty} \frac{dx}{x^{n\beta}(x - \Delta\widetilde{\theta})^{n\beta}} \leqslant \int \lambda\rho(\lambda) \, d\lambda \int_{\Delta\widetilde{\theta}+\delta}^{\infty} \frac{dx}{x^{n\beta}\delta^{n\beta}} \leqslant \int \lambda\rho(\lambda) \, d\lambda \int_{\Delta\widetilde{\theta}}^{\infty} \frac{dx}{x^{n\beta}\delta^{n\beta}} \sim \overline{\lambda^{2n\beta}} M^{1-2n\beta}, \tag{G8}$$

where the last scaling holds due to Eq. (G2), and

$$\mathcal{J}_3 \propto \int \lambda \rho(\lambda)\, d\lambda \int_\delta^{\Delta\widetilde{\theta}-\delta} \frac{dx}{x^{n\beta}(\Delta\widetilde{\theta}-x)^{n\beta}} \leqslant \int \lambda \rho(\lambda)\, d\lambda \int_\delta^{\Delta\widetilde{\theta}-\delta} \frac{dx}{\delta^{2n\beta}} \sim \overline{\lambda^{2n\beta}} M^{1-2n\beta}. \tag{G9}$$

The integral in the first subregion, $\mathcal{J}_1$ can be approximated similarly to $\mathcal{J}_5$, resulting in

$$\mathcal{J}_1 \sim \overline{\lambda^{2n\beta}} M^{1-2n\beta}. \tag{G10}$$

In the case of sf-ps networks $\rho(\lambda) = \delta(\lambda - \bar{\lambda})$ and therefore all moments of $\rho(\lambda)$ are independent of M. In the case of sf-sf networks $\rho(\lambda) \sim \lambda^{-\gamma}$, $\gamma > 2$, and sufficiently high moments of $\rho(\lambda)$ are size dependent. In a finite sample of randomly drawn λ values, the maximum value $\lambda_{\max}$ can be estimated with the natural cutoff, $\lambda_{\max} \sim M^{\frac{1}{\gamma-1}}$ [82]. The natural cutoff allows one to estimate size-dependent moments of $\rho(\lambda)$ as

$$\overline{\lambda^{1+n\beta}} \propto \lambda_{\max}^{2+n\beta-\gamma} \sim M^{\frac{2+n\beta-\gamma}{\gamma-1}}, \tag{G11}$$

$$\overline{\lambda^{2n\beta}} \propto \lambda_{\max}^{1+2n\beta-\gamma} \sim M^{\frac{1+2n\beta-\gamma}{\gamma-1}}. \tag{G12}$$

It is straightforward to verify using Eqs. (G5), (G6), (G8), (G9), and (G10) that $\mathcal{J}_1$, $\mathcal{J}_3$, and $\mathcal{J}_5$ decrease faster than $\mathcal{J}_2$ and $\mathcal{J}_4$ as M increases, for both sf-ps and sf-sf bipartite networks. Therefore, to the leading order, $r_u(i,j)$ can be approximated as

$$-\ln[1 - r_u(i,j)] \propto \frac{\overline{\lambda^{1+\beta}}}{(M\Delta\theta_{ij})^\beta}\left[\kappa_i(\kappa_j)^\beta I_1\left(\beta; \delta\sqrt{\frac{\kappa_j}{\kappa_i}}\right) + \kappa_j(\kappa_i)^\beta I_1\left(\beta; \delta\sqrt{\frac{\kappa_i}{\kappa_j}}\right)\right], \tag{G13}$$

where $\overline{\lambda^{1+\beta}} \equiv \int \lambda^{1+\beta}\rho(\lambda)\, d\lambda$. It is important to note that since this leading order corresponds to the $n = 1$ term in Eq. (G1), and since this term is proportional to the expected number of common neighbors between the two nodes,

$$-\ln[1 - r_u(i,j)] \propto \overline{m}(\kappa_i, \kappa_j, \Delta\theta_{ij}), \tag{G14}$$

the right-hand side of Eq. (G13) is proportional to the expected number of common neighbors between nodes i and j, proving Eq. (31) in the main text.

Finally, since $\delta \sim \Delta\widetilde{\theta} \gg 1$, $I_1(\beta; \delta\sqrt{\frac{\kappa_j}{\kappa_i}}) \approx I_1(\beta; \delta\sqrt{\frac{\kappa_i}{\kappa_j}}) \approx I_1(\beta; \infty)$, and therefore

$$-\ln[1 - r_u(i,j)] \sim \frac{\overline{\lambda^{1+\beta}}}{M^\beta}\left[\frac{\Delta\theta_{ij}}{d_u(\kappa_i, \kappa_j)}\right]^{-\beta},$$

$$d_u(\kappa_i, \kappa_j) \equiv \left(\kappa_i^\beta \kappa_j + \kappa_i \kappa_j^\beta\right)^{\frac{1}{\beta}}, \tag{G15}$$

proving Eq. (44) in the main text.

Equation (G15) allows us to analyze the behavior of $-\ln[1 - r_u(i,j)]$ as the system size $M \propto N$ increases. In the case of sf-ps networks $\overline{\lambda^{1+\beta}} = \bar{\lambda}^{1+\beta}$ is finite and independent of M, leading to

$$-\ln[1 - r_u(i,j)] \sim M^{-\beta}. \tag{G16}$$

In the case of sf-sf networks, $\overline{\lambda^{1+n\beta}}$ is given by Eq. (G11) and results in

$$-\ln[1 - r_u(i,j)] \sim M^{-\frac{(\gamma-2)(1+\beta)}{\gamma-1}}. \tag{G17}$$

As can be seen from Eqs. (G16) and (G17), the connection probability $r_u(i,j)$ of top nodes in the one-mode projection decreases as M increases, and scales as $M^{-\eta}$, where exponent $\eta = \beta$ in sf-ps networks and $\eta = \frac{(\gamma-2)(1+\beta)}{\gamma-1}$ in sf-sf networks. This result implies that one-mode projections of sufficiently large bipartite networks are sparse. However, we note that $\eta \to 0$ in sf-sf networks as the power-law exponent of the bottom domain $\gamma \to 2$. This result implies that one-mode projections of finite sf-sf networks characterized by γ close to 2 and small β values may be very dense and overinflated with fully connected subgraphs. This effect can render geometry inference using the one-mode projections inaccurate.

[1] M. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).

[2] M. Latapy, C. Magnien, N. Del Vecchio, N. Del Vecchio, and N. Del Vecchio, Basic notions for the analysis of large two-mode networks, Soc. Netw. **30**, 31 (2008).

[3] L. Tian, Y. He, H. Liu, and R. Du, A general evolving model for growing bipartite networks, Phys. Lett. A **376**, 1827 (2012).

[4] C. Xu Zhang, Z. K. Zhang, and C. Liu, An evolving model of online bipartite networks, Phys. A Stat. Mech. Appl. **392**, 6100 (2013).

[5] M. B. Díaz, M. A. Porter, and J. P. Onnela, Competition for popularity in bipartite networks, Chaos **20**, 43101 (2010).

[6] J. C. Nacher, T. Ochiai, M. Hayashida, and T. Akutsu, A mathematical model for generating bipartite graphs and its application to protein networks, J. Phys. A Math. Theor. **42**, 485005 (2009).

[7] IMDb, The Internet Movie Database, http://www.imdb.com.

[8] ArXiv, ArXiv e-Print archive, http://www.arxiv.org.

[9] Wikipedia, Wikipedia, the free encyclopedia that anyone can edit, http://www.wikipedia.org.

[10] H. Ma and A.-P. Zeng, Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms, Bioinformatics **19**, 270 (2003).

[11] E. Burgos, H. Ceva, L. Hernández, R. P. J. Perazzo, M. Devoto, and D. Medan, Two classes of bipartite networks: nested biological and social systems, Phys. Rev. E **78**, 046113 (2008).

[12] A. Iamnitchi, M. Ripeanu, and I. Foster, Small-world file-sharing communities, in *Twenty-third Annu. Jt. Conf. IEEE Comput. Commun. Soc.* (INFOCOM 2004) (IEEE, Los Alamitos, CA, 2004), Vol. 2, pp. 952–963.

[13] V. Zlatić, D. Garlaschelli, and G. Caldarelli, Networks with arbitrary edge multiplicities, Europhys. Lett. **97**, 28005 (2012).

[14] J. Moody, The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999, Am. Sociol. Rev. **69**, 213 (2004).

[15] M. Tomassini and L. Luthi, Empirical analysis of the evolution of a scientific collaboration network, Phys. A Stat. Mech. Appl. **385**, 750 (2007).

[16] B. Palsson, *Systems Biology* (Cambridge University Press, Cambridge, 2015).

[17] A. G. Smart, L. A. N. Amaral, and J. M. Ottino, Cascading failure and robustness in metabolic networks, Proc. Natl. Acad. Sci. USA **105**, 13223 (2008).

[18] Y. Liu, L. Xiao, and L. Ni, Building a scalable bipartite P2P overlay network, IEEE Trans. Parallel Distrib. Syst. **18**, 1296 (2007).

[19] T. Zhou, J. Ren, M. Medo, and Y.-C. Zhang, Bipartite network projection and personal recommendation, Phys. Rev. E **76**, 046115 (2007).

[20] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, Recommender systems, Phys. Rep. **519**, 1 (2012).

[21] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, Recommender systems survey, Knowledge-Based Syst. **46**, 109 (2013).

[22] J. Ben Schafer, J. Konstan, and J. Riedl, Recommender systems in e-commerce, in *Proc. 1st ACM Conf. Electron. Commer.* (ACM Press, New York, 1999), pp. 158–166.

[23] C. Berge and E. Minieka, *Graphs and Hypergraphs* (North-Holland, Amsterdam, 1973), Vol. 7.

[24] G. Ghoshal, V. Zlatić, G. Caldarelli, and M. E. J. Newman, Random hypergraphs and their applications, Phys. Rev. E **79**, 066118 (2009).

[25] C. Liu, C. H. Yeung, and Z.-K. Zhang, Self-organization in social tagging systems, Phys. Rev. E **83**, 066104 (2011).

[26] G. Tibély, P. Pollner, T. Vicsek, and G. Palla, Extracting tag hierarchies, PLoS ONE **8**, e84133 (2013).

[27] S. Ming-Sheng and Z. Zi-Ke, Diffusion-based recommendation in collaborative tagging systems, Chin. Phys. Lett. **26**, 118903 (2009).

[28] Z.-K. Zhang and C. Liu, A hypergraph model of social tagging networks, J. Stat. Mech. (2010) P10005.

[29] M. McPherson, L. Smith-Lovin, and J. M. Cook, Birds of a feather: Homophily in social networks, Annu. Rev. Sociol. **27**, 415 (2001).

[30] D. D. McFarland and D. J. Brown, Social distance as a metric: A systematic introduction to smallest space analysis, Bond. Plur. Form Subst. Urban Soc. Netw. **6**, 213 (1973).

[31] P. Sarkar and A. W. Moore, Dynamic social network analysis using latent space models, ACM SIGKDD Explor. Newsl. **7**, 31 (2005).

[32] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, Models of social networks based on social distance attachment, Phys. Rev. E **70**, 056122 (2004).

[33] Ö. Simek and D. Jensen, Navigating networks by using homophily and degree, Proc. Natl. Acad. Sci. USA **105**, 12758 (2008).

[34] L. Ferretti and M. Cortelezzi, Preferential attachment in growing spatial networks, Phys. Rev. E **84**, 016103 (2011).

[35] M. Barthélemy, Spatial networks, Phys. Rep. **499**, 1, (2011).

[36] P. Sarkar, D. Chakrabarti, and A. W. Moore, Theoretical justification of popular link prediction heuristics, IJCAI Proc-Intl. Joint Conf. Artif. Intell. **22**, 2722 (2011).

[37] M. Penrose, *Random Geometric Graphs* (Oxford University Press, Oxford, 2003).

[38] P. Balister, A. Sarkar, and B. Bollobás, Percolation, connectivity, coverage and colouring of random geometric graphs, in *Handbook of Large-Scale Random Networks* (Springer, Berlin, 2008), pp. 117–142.

[39] M. Franceschetti and R. Meester, *Random Networks for Communication: From Statistical Physics to Information Systems* (Cambridge University Press, Cambridge, 2008).

[40] S. K. Iyer and D. Yogeshwaran, Percolation and connectivity in AB random geometric graphs, Adv. Appl. Probab. **44**, 21 (2012).

[41] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication* (Cambridge University Press, Cambridge, 2005).

[42] M. D. Penrose, Continuum AB percolation and AB random geometric graphs, J. Appl. Probab. **51**, 333 (2014).

[43] G. Caldarelli, A. Capocci, P. De Los Rios, and M. Muñoz, Scale-Free Networks from Varying Vertex Intrinsic Fitness, Phys. Rev. Lett. **89**, 258702 (2002).

[44] M. Boguná and R. Pastor-Satorras, Class of correlated random networks with hidden variables, Phys. Rev. E **68**, 036112 (2003).

[45] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, Hyperbolic geometry of complex networks, Phys. Rev. E **82**, 036106 (2010).

[46] F. Papadopoulos, M. Kitsak, M. A. Serrano, M. Boguna, and D. Krioukov, Popularity versus similarity in growing networks, Nature **489**, 537 (2012).

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

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

[49] M. Serrano, D. Krioukov, and M. Boguñá, Self-Similarity of Complex Networks and Hidden Metric Spaces, Phys. Rev. Lett. **100**, 078701 (2008).

[50] M. Kitsak and D. Krioukov, Hidden variables in bipartite networks, Phys. Rev. E **84**, 026114 (2011).

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

[52] P. Zhang, J. Wang, X. Li, M. Li, Z. Di, and Y. Fan, Clustering coefficient and community structure of bipartite networks, Phys. A Stat. Mech. Appl. **387**, 6869 (2008).

[53] P. Jaccard, Etude comparative de la distribution florale dans une portion des Alpes et du Jura, Bulletin del la Société Vaudoise des Sciences Naturelles **37**, 547 (1901).

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

[55] F. Papadopoulos, C. Psomas, and D. Krioukov, Network mapping by replaying hyperbolic growth, IEEE/ACM Trans. Netw. **23**, 198 (2015).

[56] F. Papadopoulos, R. Aldecoa, and D. Krioukov, Network geometry inference using common neighbors, Phys. Rev. E **92**, 022807 (2015).

[57] S. Battiston and M. Catanzaro, Statistical properties of corporate board and director networks, Eur. Phys. J. B **38**, 345 (2004).

[58] M. E. J. Newman, Who is the best connected scientist? A study of scientific coauthorship networks, Complex Netw. **650**, 337 (2004).

[59] S. Le Blond, J. L. Guillaume, and M. Latapy, Clustering in P2P exchanges and consequences on performances, Lect. Not. Comput. Sci. **3640**, 193 (2005).

[60] S. A. Morris and G. G. Yen, Construction of bipartite and unipartite weighted networks from collections of journal papers, arXiv:physics/0503061 (2005).

[61] M. Tumminello, S. Miccichè, F. Lillo, J. Piilo, and R. N. Mantegna, Statistically validated networks in bipartite complex systems, PLoS ONE **6**(3), e17994 (2011).

[62] S. Gualdi, G. Cimini, K. Primicerio, R. Di Clemente, and D. Challet, Statistically validated network of portfolio overlaps and systemic risk, Sci. Rep. **6**, 39467 (2016).

[63] N. Dianati, A maximum entropy approach to separating noise from signal in bimodal affiliation networks, arXiv:1607.01735.

[64] F. Saracco, M. J. Straka, R. Di Clemente, A. Gabrielli, G. Caldarelli, and T. Squartini, Inferring monopartite projections of bipartite networks: An entropy-based approach, arXiv:1607.02481.

[65] S. Lehmann, M. Schwartz, and L. K. Hansen, Biclique communities, Phys. Rev. E **78**, 016108 (2008).

[66] X. Wang and X. Qin, Asymmetric intimacy and algorithm for detecting communities in bipartite networks, Phys. A Stat. Mech. Appl. **462**, 569 (2016).

[67] X. Liu and T. Murata, Community detection in large-scale bipartite networks, in *IEEE/WIC/ACM Int. Jt. Conf. Web Intell. Intell. Agent Technol. 2009* (WI-IAT'09) (IET, Milan, Italy, 2009), Vol. 1, pp. 50–57.

[68] N. Benchettara, R. Kanawati, and C. Rouveirol, Supervised machine learning applied to link prediction in bipartite social networks, in *2010 Int. Conf. Adv. Soc. Networks Anal. Min.* (ASONAM) (IEEE, Los Alamitos, CA, 2010), pp. 326–330.

[69] X. Liu and T. Murata, Evaluating community structure in bipartite networks, in *2010 IEEE Second Int. Conf. Soc. Comput.* (SocialCom) (IEEE, Los Alamitos, CA, 2010), pp. 576–581.

[70] Z. Li, S. Zhang, and X. Zhang, Mathematical model and algorithm for link community detection in bipartite networks, Am. J. Oper. Res. **5**, 421 (2015).

[71] D. B. Larremore, A. Clauset, and A. Z. Jacobs, Efficiently inferring community structure in bipartite networks, Phys. Rev. E **90**, 012805 (2014).

[72] M. Kheirkhahzadeh, A. Lancichinetti, and M. Rosvall, Efficient community detection of network flows for varying Markov times and bipartite networks, Phys. Rev. E **93**, 032309 (2016).

[73] J. Silvio Gutkind, *Signaling Networks and Cell Cycle Control* (Humana Press, New York, 2000).

[74] Y. Yao, Z. Dalin, W. Qing, and S. Jinglin, A multicast search scheme based on bipartite graph matching model, Intl. J. Signal Process. Image Process. Pattern Recognit. **8**, 397 (2015).

[75] G. Uchyigit and M. Y. Ma, *Personalization Techniques and Recommender Systems* (World Scientific, Singapore, 2008), Vol. 70.

[76] D.-C. Nie, Y.-H. An, Q. Dong, Y. Fu, and T. Zhou, Information filtering via balanced diffusion on bipartite networks, Phys. A Stat. Mech. Appl. **421**, 44 (2015).

[77] Y.-H. An, Q. Dong, C.-J. Sun, D.-C. Nie, and Y. Fu, Diffusion-like recommendation with enhanced similarity of objects, Phys. A Stat. Mech. Appl. **461**, 708 (2016).

[78] X. Pan, G. Deng, and J.-G. Liu, Weighted bipartite network and personalized recommendation, Phys. Proc. **3**, 1867 (2010).

[79] M. E. J. Newman and T. P. Peixoto, Generalized Communities in Networks, Phys. Rev. Lett. **115**, 088701 (2015).

[80] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri, Feedback effects between similarity and social influence in online communities, in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* (ACM, New York, 2008), pp. 160–168.

[81] H. S. Wilf, *Generating Functionology* (Elsevier, Amsterdam, 2013).

[82] S. N. Dorogovtsev and J. F. F. Mendes, Evolution of networks, Adv. Phys. **51**, 1079 (2002).