Weighted hypersoft configuration model

Ivan Voitalov,[1,2] Pim van der Hoorn,[3] Maksim Kitsak,[1,2,4] Fragkiskos Papadopoulos,[5] and Dmitri Krioukov[1,2,6,7]

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

[2]*Network Science Institute, Northeastern University, Boston, Massachusetts 02115, USA*

[3]*Department of Mathematics and Computer Science, Eindhoven University of Technology, Postbus 513, 5600 MB Eindhoven, Netherlands*

[4]*Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD, Delft, Netherlands*

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

[6]*Department of Mathematics, Northeastern University, Boston, Massachusetts 02115, USA*

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

(Received 30 June 2020; accepted 27 September 2020; published 29 October 2020)

Maximum entropy null models of networks come in different flavors that depend on the type of constraints under which entropy is maximized. If the constraints are on degree sequences or distributions, we are dealing with configuration models. If the degree sequence is constrained exactly, the corresponding microcanonical ensemble of random graphs with a given degree sequence is the configuration model *per se*. If the degree sequence is constrained only on average, the corresponding grand-canonical ensemble of random graphs with a given expected degree sequence is the soft configuration model. If the degree sequence is not fixed at all but randomly drawn from a fixed distribution, the corresponding hypercanonical ensemble of random graphs with a given degree distribution is the hypersoft configuration model, a more adequate description of dynamic real-world networks in which degree sequences are never fixed but degree distributions often stay stable. Here, we introduce the hypersoft configuration model of weighted networks. The main contribution is a particular version of the model with power-law degree and strength distributions, and superlinear scaling of strengths with degrees, mimicking the properties of some real-world networks. As a byproduct, we generalize the notions of sparse graphons and their entropy to weighted networks.

DOI: 10.1103/PhysRevResearch.2.043157

I. INTRODUCTION

Many real-world complex systems that can be represented as networks [1,2] require weighted representations in which connections between nodes are characterized by positive weights [3]. For example, in modeling the global spread of an epidemic using an air transportation network as a backbone, it is important to know not only that there exists a flight from airport i to airport j, but also the volume of the passenger flow between the two airports. This volume is usually encoded as the link weight w_{ij} [4,5].

Within the plethora of weighted and unweighted network models developed in network science to study the structure and function of real-world networks, *maximum-entropy models* [6–16] play a special role. They serve as null models that are indispensable in studying the intricate interdependencies between different network properties [17–27]. Within this class of models, perhaps the most studied are classical random graphs [28–31] that maximize ensemble entropy with the average degree constrained to a given value. They do not reproduce heterogeneous degree distributions observed in many real-world networks [32], which motivated the development of the configuration model.

The *configuration model (CM)* [33,34] is a microcanonical ensemble of random graphs with *sharp constraints* on the degree sequence, meaning that every graph in this ensemble has exactly the same degree sequence, e.g., the one observed in a real-world network. In the CM, every graph satisfying the constraint has the same probability in the ensemble. All other graphs are excluded. The *soft configuration model (SCM)* [6,8,10,35,36] is a grand-canonical ensemble of random graphs with *soft constraints* on the degree sequence. This means that the *expected* degree sequence in the ensemble is equal to a given degree sequence.

In both CM and SCM, a fixed degree sequence is the constraint under which the ensemble entropy is maximized, either micro- or grand-canonically. This constraint, however, does not properly reflect the dynamic nature of node degrees observed in many real networks, where the degrees of all nodes may constantly change, while the shape of the degree distribution stays stable [37,38]. These observations motivated the development of the hypersoft configuration model.

The *hypersoft configuration model (HSCM)* [14,39–41] is a hypercanonical ensemble of random graphs whose entropy is maximized under the constraint that the *degree distribution*

TABLE I. Configuration models of unweighted and weighted networks. The subject of this paper is the WHSCM.

Maximum entropy constraints	Unweighted models	Weighted models
Exact degree (and strength) sequence	CM	WCM
Expected degree (and strength) sequence	SCM	WSCM
Degree (and strength) distribution	HSCM	WHSCM

has a given shape. The HSCM belongs to the class of models with hidden variables [40], meaning that each node has a latent parameter sampled from a fixed distribution. This latent distribution defines the degree distribution, so that by tuning the former, one can reproduce any shape of the latter.

The ((H)S)CM story outlined above for unweighted networks finds an incomplete and somewhat distorted reflection in weighted networks where the constraints under which entropy is maximized are on both degrees and strengths of nodes [12]. The *weighted configuration model (WCM)* is a microcanonical ensemble of networks with sharp constraints on both the degree and strength sequences. Every weighted graph in this ensemble has the same degree and strength sequences, and the same probability in the ensemble. However, we note that several models that come under the WCM name [12,42,43] are not really as defined above. The *weighted soft configuration model (WSCM)* [9,12,13,15] is a grand-canonical ensemble of networks with soft constraints on the degree and strength sequences, meaning that the expected degree and strength sequences in the ensemble are equal to given degree and strength sequences.

In this paper, we introduce the *weighted hypersoft configuration model (WHSCM)* which is a hypercanonical ensemble of networks with a fixed *joint distribution* of degrees and strengths. Similar to the HSCM, the WHSCM is a hidden variable model where each node has two latent parameters sampled from a fixed joint distribution which defines the joint distribution of degrees and strengths. We summarize the taxonomy of the (((W)H)S)CM models in Table I.

Besides introducing the WHSCM in general, the main focus of this paper is a much more involved task, which is to identify the joint distribution of latent parameters that reproduces several features of degree and strength distributions observed in many real weighted networks [3,44–48]. Specifically, these features are: (1) *power-law degree distribution*, (2) *superlinear scaling between strengths and degrees*, and (3) *sparsity*. The last one means that the average degree is constant as a function of the network size.

We proceed by first providing all the necessary motivation and background information in Sec. II that ends with the introduction of the most general form of the WHSCM. In Sec. III, we document in detail the real-world-network-dictated properties, mentioned above, that we want our particular power-law version of the WHSCM to reproduce. To reproduce those properties, we need some experimental input from the WSCM as defined in Ref. [15], a subject of Sec. IV. Based on this input, we derive in Sec. V the WHSCM latent parameter distribution that satisfies the requirements of Sec. III. We check in simulations that these requirements are indeed satisfied in Sec. VI. To demonstrate how the constructed model can be used in the analysis of real-world networks, we juxtapose several real-world networks and their WHSCM counterparts in Sec. VII. We conclude in Sec. VIII with the discussion of obvious and less obvious limitations, caveats, wishful thoughts, and abstract remarks. We release our implementation of the WHSCM graph generator in a software package available at GitHub [49].

II. MOTIVATION, BACKGROUND, AND GENERAL WHSCM

A. Maximum entropy models

Understanding mechanisms that drive formation of complex networks and dynamical processes running in them is a crucial task in network science [50]. It is commonly believed that many real complex systems are *self-organizing*, i.e., adjusting their structure to optimize their function [51]. Because of that, a lot of past research was dedicated to finding structural network properties that may be indicative of yet unknown optimization mechanisms behind network evolution and function [51–53]. However, due to potentially strong interdependencies between different structural properties of networks [17–20], it is important to ensure that a property of interest is indeed a salient feature, and not a mere consequence of some of its other structural properties. A method that is often used to make this check is to compare the significance of the structural property present in a given network with respect to the same property in benchmark null model networks [21–24]. This step should be taken with care as choosing an inappropriate null model for a given network may lead to wrong conclusions about its functional and structural features [25–27].

The maximum entropy network null models [6,8,9,12,13,16,54] have proven to be an indispensable tool in avoiding possible statistical biases caused by interdependencies of structural network properties. These models are ensembles of networks that reproduce given structural properties, and that are maximally random in all other respects. This maximal randomness is important for a great variety of tasks [17–27]. For a basic example, if a maximum entropy null model is defined by property X observed in a real-world network, and if the model also reproduces some other property Y of the network, then we know right away that Y is not a salient independent feature, but a statistical consequence of X [20].

Maximum entropy network models are usually formulated for a given observed network G^* of size n in terms of sets of *constraints*. These constraints are usually some properties $C(G^*)$ of network G^* that the model is required to reproduce. We note that the constraints do certainly *not* have to be properties of any real network, they can be any set of (artificial) network properties, but what we describe is the most common application scenario.

Given constraints $C(G^*)$, the model is then an ensemble of graphs $\mathcal{G}$ of the same size n as the original graph G^*, with each graph $G \in \mathcal{G}$ appearing in the ensemble with probability $P(G)$, known as the ensemble distribution. The ensemble distribution in a maximum entropy model defined by constraints

$C(G^*)$ is the unique unbiased probability distribution $P(G)$ that maximizes Gibbs/Shannon entropy

$$S = -\sum_{G\in\mathcal{G}} P(G)\ln P(G), \tag{1}$$

and that satisfies the constraints $C(G^*)$ and the normalization condition $\sum_{G\in\mathcal{G}} P(G) = 1$ [16,55].

B. Unweighted configuration models

In the simplest case of *undirected* and *unweighted* networks, the degrees of all nodes in a given network are frequently used as constraints in maximum entropy null models. The simplest example is the configuration model.

1. Configuration model (CM)

The configuration model (CM) [33,34] is a microcanonical ensemble of graphs with the same degree sequence as observed in a real network. That is, given the degree sequence $\{k_1^*,\ldots,k_n^*\} = \boldsymbol{k}^*$ observed in a graph G^*, the CM ensemble consists of all graphs G with exactly the same degree sequence, i.e., for each degree sequence $\boldsymbol{k}(G)$ of an ensemble graph G, the following holds:

$$\boldsymbol{k}(G) = \boldsymbol{k}^*. \tag{2}$$

The distribution $P(G)$ that maximizes Shannon entropy in Eq. (1) is the uniform distribution over the set of all graphs whose degree sequence is $\boldsymbol{k}^*$, meaning that for these graphs $P(G) = 1/\mathcal{N}_{\boldsymbol{k}^*}$ and $S = \ln\mathcal{N}_{\boldsymbol{k}^*}$, where $\mathcal{N}_{\boldsymbol{k}^*}$ is the number of graphs with the degree sequence $\boldsymbol{k}^*$. For all other graphs, $P(G) = 0$.

2. Soft configuration model (SCM)

The Soft configuration model (SCM) [6,8,10,35,36] is a grand-canonical ensemble of graphs whose expected degree sequence is constrained to a given (observed) sequence. Specifically, given a degree sequence $\boldsymbol{k}^*$, the SCM constraint is

$$\sum_{G\in\mathcal{G}} \boldsymbol{k}(G)P(G) = \boldsymbol{k}^*, \tag{3}$$

where $\boldsymbol{k}(G)$ is the degree sequence in an ensemble graph G. Since the degree sequence used as a constraint is reproduced only in expectation, it does not have to consist of integers only; the expected degrees can be any positive real numbers.

As in any (grand)canonical ensemble, the Shannon entropy in the SCM is usually maximized using the method of Lagrange multipliers [6], yielding the familiar Gibbs/Boltzmann exponential family distribution

$$P(G) = \frac{e^{-H_{\rm SCM}(G)}}{Z} \tag{4}$$

with Hamiltonian

$$H_{\rm SCM}(G) = \sum_{(i,j)\in\mathcal{P}} (\nu_i+\nu_j)a_{ij} = \sum_i \nu_i k_i(G), \tag{5}$$

where ν_i is the Lagrange multiplier coupled to node i, $k_i(G)$ i's degree in G, $\mathcal{P}$ the set of all node pairs, a_{ij} the adjacency matrix of G, and $Z = \sum_{G\in\mathcal{G}} e^{-H_{\rm SCM}(G)}$ the partition function. In statistical terms, these equations say that the degree sequence is the sufficient statistics in the ensemble, so that all graphs with the same degree sequence have the same probability in the ensemble.

Graphs G can be sampled from the ensemble distribution $P(G)$ constructively by walking over all node pairs i, j and linking them with the Fermi-Dirac *connection probability*

$$p_{ij} = p(\nu_i,\nu_j) = \frac{1}{1+e^{\nu_i+\nu_j}}. \tag{6}$$

The expected degree κ_i of node i in the ensemble is thus

$$\kappa_i = \sum_j p_{ij}, \tag{7}$$

so that the values of the Lagrange multipliers ν_i for a given $\boldsymbol{k}^*$ are found as the solution of the system of n equations

$$\kappa_i = k_i^*. \tag{8}$$

3. Hypersoft configuration model (HSCM)

The Hypersoft configuration model (HSCM) [14,39–41] is a hypercanonical ensemble of graphs defined by the constraint that the expected degree distribution has a given form. The HSCM can be viewed as a hyperparametrization of the SCM, in that the Lagrange multipliers ν are not fixed by any degree sequence as solutions of (8), but random, sampled independently for each node i from a fixed distribution $\rho(\nu)$:

$$\nu_i \leftarrow \rho(\nu). \tag{9}$$

Having a sampled sequence of Lagrange multipliers ν_i, the nodes are then connected as in the SCM, with the Fermi-Dirac connection probability in Eq. (6).

The expected degree $\kappa(\nu)$ of a node with Lagrange multiplier ν in the ensemble is

$$\kappa(\nu) = n\int p(\nu,\nu')\rho(\nu')\,d\nu', \tag{10}$$

where $p(\nu,\nu')$ is from Eq. (6). It is convenient to abuse the notations by defining the expected degree random variable κ via

$$\kappa = \kappa(\nu), \tag{11}$$

where the left-hand side (l.h.s.) κ is a random variable, but the right-hand side (r.h.s.) $\kappa(\nu)$ is a function, defined in Eq. (10), of the random variable ν whose distribution is $\rho(\nu)$. That is, the last equation is a change of latent variables from ν to κ. One can show [14] that in sparse graphs the $\kappa(\nu)$ function can be well approximated as

$$\kappa(\nu) = \kappa_0 e^{R-\nu} = \sqrt{\bar{k}n}\,e^{-\nu}, \tag{12}$$

where R and κ_0 are the interchangeable parameters that control the expected average degree $\bar{k} = \kappa_0^2 e^{2R}/n$ in the ensemble. With this approximation, the Fermi-Dirac connection probability in Eq. (6) can be rewritten in terms of the κ variables as

$$p(\kappa_i,\kappa_j) = \frac{1}{1+\frac{\bar{k}n}{\kappa_i\kappa_j}}. \tag{13}$$

As proven in [56], the classical limit (or Chung-Lu [35,36]) approximation

$$p_{cl}(\kappa_i, \kappa_j) = \min\left(1, \frac{\kappa_i \kappa_j}{\bar{k}n}\right) \tag{14}$$

to the connection probability (13) "almost always works," in the sense that the HSCM ensembles of random graphs defined by the connections probabilities (13) and (14) are asymptotically equivalent under very mild assumptions on the distribution of κ.

Since Eq. (11) defines the relation between the two random variables κ and ν, it also defines the relation between their distributions $\rho(\nu)$ and $\rho(\kappa)$ via the standard formula

$$\rho_\kappa(\kappa) = \rho_\nu[\nu(\kappa)]|\nu'(\kappa)|, \tag{15}$$

where $\nu(\kappa)$ is the inverse function of $\kappa(\nu)$ and $\nu'(\kappa)$ its derivative. By the definition of κ, its distribution $\rho(\kappa)$ is the distribution of expected degrees in the HSCM, the analogy of the expected degree sequence κ_i (7) in the SCM. Therefore the HSCM analogy of the SCM constraints (8) is

$$\rho(\kappa) = \rho^*(\kappa), \tag{16}$$

where $\rho^*(\kappa)$ is any desired expected degree distribution. For example, it can be a pure power law, i.e., the continuous Pareto distribution

$$\rho(\kappa) = (\gamma - 1)\kappa_0^{\gamma-1}\kappa^{-\gamma}, \quad \kappa > \kappa_0 > 0. \tag{17}$$

In view of Eq. (12), the distribution of the Lagrange multipliers ν is exponential in this case:

$$\rho(\nu) = (\gamma - 1)e^{(\gamma-1)(\nu-R)}, \quad \nu \in (-\infty, R]. \tag{18}$$

The general exact expression for the expected average degree in the ensemble is

$$\bar{k} = \bar{\kappa} = \int \kappa\rho(\kappa)\,d\kappa = \int \kappa(\nu)\rho(\nu)\,d\nu. \tag{19}$$

In sparse graphs, the expected degree distribution $\rho(\kappa)$ defines the degree distribution $P(k)$ via

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

where $\mathrm{Pois}(k|\kappa)$ is the Poisson distribution with mean κ [40,57,58]. The Poisson distribution appears here as the $n \to \infty$ limit of the distributions of sums of n Bernoullis with random rates p_{ij} whose sum $\sum_j p_{ij}$ converges to κ_i. The distributions $P(k)$ in the form (20) are called mixed Poisson distributions with κ the mixing parameter [59]. A short proof that the degree distribution in the HSCM is a mixed Poisson distribution can be found in Th. 3.1 in Ref. [57], for instance. The proof relies on the observation that the generating function of the degree distribution is the generating function of the mixed Poisson distribution. The shape of a mixed Poisson distribution $P(k)$ follows the shape of the distribution of its mixing parameter $\rho(\kappa)$. For example, if $\rho(\kappa)$ is Pareto with exponent γ, then the Pareto-mixed Poisson distribution is also a power law, albeit impure, with the same exponent, $P(k) \sim k^{-\gamma}$, or in stricter terms, it is a regularly varying distribution with exponent γ [32].

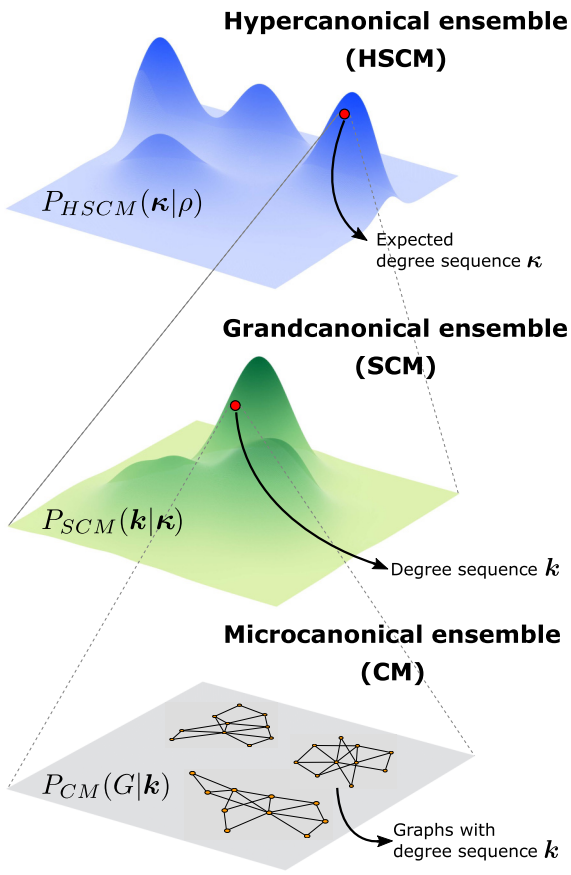

FIG. 1. The nested hierarchy of the configuration models. Sampling a graph G from the hypercanonical HSCM ensemble can be done in three steps: (1) sample a sequence $\boldsymbol{\kappa} = \{\kappa_i\}$ of expected degrees independently from the distribution $\rho(\kappa)$, i.e., from $P_{\mathrm{HSCM}}(\boldsymbol{\kappa}|\rho) = \prod_i \rho(\kappa_i)$; (2) sample a degree sequence $\boldsymbol{k} = \{k_i\}$ from the SCM degree sequence distribution $P_{\mathrm{SCM}}(\boldsymbol{k}|\boldsymbol{\kappa})$; and (3) sample graph G from the uniform CM distribution $P_{\mathrm{CM}}(G|\boldsymbol{k})$. The HSCM probability distribution can thus be written as a chain $P_{\mathrm{HSCM}}(G|\rho) = \int_{\boldsymbol{\kappa}} \sum_{\boldsymbol{k}} P_{\mathrm{CM}}(G|\boldsymbol{k}) P_{\mathrm{SCM}}(\boldsymbol{k}|\boldsymbol{\kappa}) P_{\mathrm{HSCM}}(\boldsymbol{\kappa}|\rho)\,d\boldsymbol{\kappa}$ showing that the ensemble at the next level of the hierarchy is a probabilistic mixture of the ensembles at the previous level. In the graphon theory [60], one has to consider the fourth highest level (not shown) where ρ is also random. Moving up in the hierarchy adds new sources of randomness, thus increasing entropy. Since $P_{\mathrm{SCM}}(\boldsymbol{k}|\boldsymbol{\kappa})$ is intractable—it is a mixture of mixed Poisson distributions—in practice it is much easier to sample graphs as described in the text. The same picture applies to the weighted case upon the addition of strengths $\boldsymbol{s}$ and their expectations $\boldsymbol{\sigma}$.

As shown in Ref. [14] and discussed in Appendix A, the entropy of HSCM graphs is maximized across all graphs whose degree distribution converges to a given distribution.

The nested hierarchy of the described configuration models is visualized in Fig. 1.

C. Weighted configuration models

For *weighted* networks, the configuration models are analogous to those for unweighted networks discussed above. They are formulated in terms of constraints on both degrees and strengths of nodes.

1. Weighted configuration model (WCM)

The WCM is a microcanonical ensemble of weighted networks with sharp constraints on the degree and strength sequences $\boldsymbol{k}^*$ and $\boldsymbol{s}^*$. The ensemble consists of weighted graphs that have exactly the same degree and strength sequences as in the observed graph:

$$\boldsymbol{k}(G) = \boldsymbol{k}^*, \tag{21}$$

$$\boldsymbol{s}(G) = \boldsymbol{s}^*. \tag{22}$$

Analogously to its unweighted version, the distribution maximizing Shannon entropy is the uniform distribution over all graphs with the joint degree-strength sequence equal to $(\boldsymbol{k}^*, \boldsymbol{s}^*)$. This ensemble is well defined and such a uniform distribution always exists because the space of weight matrices $\{w_{ij}\}$ representing graphs satisfying the constraints (21) and (22) is always of a finite volume (Lebesgue measure) if weights are real, or of a finite cardinality if they are integer. Indeed, since all weights are positive, they all are bounded by the minimum strength of the two incident nodes: $0 < w_{ij} \leqslant \min(s_i^*, s_j^*)$.

We note that there exist several network models introduced under the WCM name in the past that are different from the WCM definition above. Specifically, in Ref. [42], the authors consider the unweighted *multigraph* CM with degree sequences following power-law distributions with $\gamma < 2$. In these settings, multiple links between the same pairs of nodes are present with high probability. These multiple links are treated as weights in Ref. [42]. In Ref. [12], the WCM is a model in which only the strength sequence is fixed, but the degree sequence is not fixed. In Ref. [43], the WCM is a model in which the degree sequence is fixed, while the weights of links attached to a node are sampled from a distribution that is allowed to depend on the node degree. To the best of our knowledge, the WCM as we defined it above has been introduced only in [61], albeit under a different name.

2. Weighted soft configuration model (WSCM)

The WSCM [9,12,13,15] is a grand-canonical ensemble of networks whose expected degree and strength sequences are constrained to given (observed) sequences. For any given (observed) degree and strength sequences $\boldsymbol{k}^*$ and $\boldsymbol{s}^*$, the WSCM constraints are

$$\sum_{G\in\mathcal{G}} \boldsymbol{k}(G)P(G) = \boldsymbol{k}^*, \tag{23}$$

$$\sum_{G\in\mathcal{G}} \boldsymbol{s}(G)P(G) = \boldsymbol{s}^*, \tag{24}$$

where $\boldsymbol{k}(G)$ and $\boldsymbol{s}(G)$ are the degree and strength sequences of an ensemble graph G. As in the unweighted case, Shannon entropy is maximized using the method of Lagrange multipliers leading to the ensemble distribution

$$P(G) = \frac{e^{-H_{\mathrm{WSCM}}(G)}}{Z} \tag{25}$$

with Hamiltonian

$$H_{\mathrm{WSCM}}(G) = \sum_{(i,j)\in\mathcal{P}} (\nu_i + \nu_j)a_{ij} + \sum_{(i,j)\in\mathcal{E}} (\mu_i + \mu_j)w_{ij} = \sum_i \nu_i k_i(G) + \mu_i s_i(G), \tag{26}$$

where ν_i and μ_i are the Lagrange multipliers coupled to node i, constraining its degree $k_i(G)$ and strength $s_i(G)$, a_{ij} and w_{ij} are the adjacency and weight matrices of G, $\mathcal{P}$ and $\mathcal{E}$ are the sets of node pairs and connected node pairs, and $Z = \sum_{G\in\mathcal{G}} e^{-H_{\mathrm{WSCM}}(G)}$ the partition function. The sufficient statistics are thus the degree and strength sequences, so that all graphs with the same degree and strength sequences have the same probability in the ensemble.

The WSCM was defined and studied both for positive integer- [9,12,13] and real-valued weights [15]. In the latter case the graphs can be sampled constructively from the ensemble distribution $P(G)$ as follows. First, every pair of nodes i and j is connected with the connection probability

$$p_{ij} = p(\nu_i, \mu_i, \nu_j, \mu_j) = \frac{1}{1 + e^{\nu_i+\nu_j}(\mu_i + \mu_j)}. \tag{27}$$

Second, every established link i, j is weighted by a random weight w_{ij} sampled from the exponential distribution with rate $\mu_i + \mu_j$:

$$w_{ij} \leftarrow \exp(w|\mu_i + \mu_j) = (\mu_i + \mu_j)e^{-(\mu_i+\mu_j)w}. \tag{28}$$

The expected weight of link i, j is then

$$\omega_{ij} = \omega(\nu_i, \mu_i, \nu_j, \mu_j) = \frac{p(\nu_i, \mu_i, \nu_j, \mu_j)}{\mu_i + \mu_j} = \frac{p_{ij}}{\mu_i + \mu_j}. \tag{29}$$

The expected degree κ_i and strength σ_i of node i in the ensemble are thus

$$\kappa_i = \sum_j p_{ij}, \tag{30}$$

$$\sigma_i = \sum_j \omega_{ij}, \tag{31}$$

so that the Lagrange multipliers ν_i, μ_i are found for given $\boldsymbol{k}^*, \boldsymbol{s}^*$ as the solution of the system of the $2n$ equations

$$\kappa_i = k_i^*, \tag{32}$$

$$\sigma_i = s_i^*. \tag{33}$$

3. Weighted hypersoft configuration model (WHSCM)

We introduce the WHSCM here as a hypercanonical ensemble of weighted networks with positive real-valued weights. The maximum entropy constraint of the model is that the joint distribution of expected degrees and strengths has a given form. Similar to the HSCM, which is a hyperparametrization of the SCM, the WHSCM is a hyperparametrization of the WSCM, meaning that the Lagrange multipliers ν_i, μ_i of node i are not fixed by any fixed degree and strength sequences. Instead, ν_i, μ_i are random, sampled independently for each node i from a fixed joint probability distribution $\rho(\nu, \mu)$:

$$(\nu_i, \mu_i) \leftarrow \rho(\nu, \mu). \tag{34}$$

Having a joint sampled sequence of Lagrange multipliers ν_i, μ_i, the nodes are then connected as in the WSCM, with the connection probability in Eq. (27), and the established links are weighted by random weights in Eq. (28).

The expected degree $\kappa(\nu, \mu)$ and strength $\sigma(\nu, \mu)$ of a node with Lagrange multipliers ν and μ in the ensemble are

$$\kappa(\nu, \mu) = n \iint p(\nu, \mu, \nu', \mu')\rho(\nu', \mu')\, d\nu'\, d\mu', \tag{35}$$

$$\sigma(\nu, \mu) = n \iint \omega(\nu, \mu, \nu', \mu')\rho(\nu', \mu')\, d\nu'\, d\mu', \tag{36}$$

where p, ω are from Eqs. (27) and (29). As in the HSCM, it is convenient to abuse the notations by introducing the expected degree and strength random variables κ and σ via

$$\kappa = \kappa(\nu, \mu), \tag{37}$$

$$\sigma = \sigma(\nu, \mu), \tag{38}$$

thus changing latent random variables from ν, μ to κ, σ. The joint distribution of the latter is given by the standard formula

$$\rho_{\kappa,\sigma}(\kappa, \sigma) = \rho_{\nu,\mu}[\nu(\kappa, \sigma), \mu(\kappa, \sigma)] \left| \frac{\partial(\nu, \mu)}{\partial(\kappa, \sigma)} \right|, \tag{39}$$

where $\nu(\kappa, \sigma), \mu(\kappa, \sigma)$ are the inverse functions of $\kappa(\nu, \mu), \sigma(\nu, \mu)$, and $|\partial(\nu, \mu)/\partial(\kappa, \sigma)|$ is the absolute value of the determinant of the Jacobian:

$$\left| \frac{\partial(\nu, \mu)}{\partial(\kappa, \sigma)} \right| = \left| \frac{\partial\nu}{\partial\kappa}\frac{\partial\mu}{\partial\sigma} - \frac{\partial\nu}{\partial\sigma}\frac{\partial\mu}{\partial\kappa} \right|. \tag{40}$$

By the definition of κ and σ, their joint distribution $\rho(\kappa, \sigma)$ is the joint distribution of expected degrees and strengths in the WHSCM, the analogy of the joint expected degree sequence κ_i, σ_i [(30) and (31)]—joint via node index i—in the WSCM. Therefore the WHSCM analogy of the WSCM constraints [(32) and (33)] is

$$\rho(\kappa, \sigma) = \rho^*(\kappa, \sigma), \tag{41}$$

where $\rho^*(\kappa, \sigma)$ is any desired joint distribution of expected degrees and strengths.

The marginal distributions $\rho(\kappa)$ and $\rho(\sigma)$ of the joint distribution $\rho(\kappa, \sigma)$ are the distributions of expected degrees and strengths in the ensemble. Therefore the expected average degree and strengths in the model are given by

$$\bar{k} = \bar{\kappa} = \int \kappa\rho(\kappa)\, d\kappa = \iint \kappa(\nu, \mu)\rho(\nu, \mu)\, d\nu\, d\mu, \tag{42}$$

$$\bar{s} = \bar{\sigma} = \int \sigma\rho(\sigma)\, d\sigma = \iint \sigma(\nu, \mu)\rho(\nu, \mu)\, d\nu\, d\mu. \tag{43}$$

The joint distribution $\rho(\kappa, \sigma)$ of expected degrees κ and strengths σ defines the joint distribution of actual degrees k and strengths s via

$$P(k, s) = \iint P(k, s|\kappa, \sigma)\rho(\kappa, \sigma)\, d\kappa\, d\sigma. \tag{44}$$

Unfortunately, the conditional joint distribution $P(k, s|\kappa, \sigma)$ of degrees and strengths k, s of nodes of a given expected degree and strength κ, σ is in general unknown. It is not even known, in general, what the closed form expression is for the conditional distribution $P(s|\sigma)$ of strengths s of nodes of a given expected strength σ, which appears in the expression for the strength distribution

$$P(s) = \int P(s|\sigma)\rho(\sigma)\, d\sigma. \tag{45}$$

In view of Eq. (28), $P(s|\sigma)$ is the distribution of a sum of exponential random variables with different *random* rates. The best what is known about such distributions—that are called *hypoexponential distributions*—are some bounds on their tails, but only for fixed, not random rates [62]. These distributions $P(s|\sigma)$ are definitely not as simple and well-studied as Poisson distributions $P(k|\kappa)$, so that very little appears to be known about mixtures of the former *a la* in (45). However, it is known [40] that for sparse graphs the conditional distribution $P(k|\kappa)$ of degrees k of nodes with a given expected degree κ in the WHSCM is still Poisson, as in the HSCM, so that Eq. (20) holds in the WHSCM as well.

In Appendix A, we generalize the notions of sparse graphons and their entropy to weighted networks. This generalization allows us to discuss how to extend the HSCM maximum entropy proof to the WHSCM case, thus showing that the entropy of graphs in the WHSCM ensemble defined above is maximized across all graphs whose joint distribution of strengths and degrees converges to a given joint distribution.

The main focus of the rest of the paper is to identify the latent parameter distribution $\rho(\nu, \mu)$ that leads to joint degree-strength distributions $P(k, s)$ observed in real-world weighted networks. In what follows, we first formulate in the next section this real-world-inspired form of $P(k, s)$ that we want our specific version of the general WHSCM to reproduce, and then derive the $\rho(\nu, \mu)$ that leads to this $P(k, s)$.

III. SPECIFIC WHSCM REQUIREMENTS

According to our general definition of the WHSCM model in the previous section, a specific version of the model is fixed by a particular choice of the joint degree-strength distribution $P(k, s)$. Here we document a specific set of properties of this joint distribution, dictated by the properties of many real-world weighted networks [3,45–48], that we want our specific version of the general WHSCM introduced above to reproduce.

First, we require the degree distribution $P(k)$, a marginal of $P(k, s)$, to be a power law

$$P(k) \sim k^{-\gamma} \tag{46}$$

with $\gamma > 2$. Here by "power law" and the "$\sim$" sign, we mean that $P(k)$ is a *regularly varying* distribution [32] which is a distribution whose complementary CDF satisfies

$$\bar{F}(k) = \ell(k)k^{-(\gamma-1)}, \tag{47}$$

where $\ell(k)$ is a *slowly varying function*. Our power laws will be Pareto-mixed Poisson distributions whose $\ell(k)$s converge to constants, $\lim_{k\to\infty} \ell(k) = c$.

Second, we require the strength of nodes to grow superlinearly with their degrees, as observed in many real weighted networks [3,47,48]. This observation is often expressed as

$$\bar{s}(k) \sim k^{\eta}, \tag{48}$$

where $\eta \geqslant 1$ and $\bar{s}(k)$ is the average strength of nodes of degree k. We interpret this relation to mean that

$$\lim_{k\to\infty} \frac{\bar{s}(k)}{k^{\eta}} = s_0 \tag{49}$$

for some constant s_0.

If the distributions $P(s|k)$ of strengths s of nodes of degree k are concentrated around their expected values, and the degree distribution $P(k)$ is a power law with exponent γ, then the resulting strength distribution $P(s)$ is a power law with exponent δ:

$$P(s) \sim s^{-\delta}, \tag{50}$$

$$\delta = 1 + \frac{\gamma - 1}{\eta}. \tag{51}$$

If $\eta > \gamma - 1$, then the strength distribution $P(s)$ has exponent $\delta < 2$, meaning that for such combinations of γ and η, the strength distribution $P(s)$ has an infinite first moment, so that the average strength $\bar{s}$ diverges. We do not want to exclude this possibility from our model, since such combinations of the values of γ and η can be found in real networks [48].

Third, we want our model to produce networks whose average degree $\bar{k}$, given by Eq. (42), is independent of the network size n. A model satisfying this condition is said to produce *sparse* networks.

To simplify the problem significantly, we next observe that if the actual degrees k and strengths s are concentrated around their expected values κ and σ—and this is indeed the case as we show in Appendix B—then the requirements discussed above can be formulated in terms of κ, σ instead of k, s.

We are thus looking for a model in which the distribution $\rho(\kappa)$ of expected degrees κ follows a power law with exponent $\gamma > 2$,

$$\rho(\kappa) \sim \kappa^{-\gamma}, \tag{52}$$

and the expected strengths σ grow super-linearly with expected degrees κ. For further simplicity, we want expected strengths and degrees to be deterministically related via

$$\sigma = \sigma_0 \kappa^{\eta}, \tag{53}$$

where $\sigma_0 > 0$. This choice instantly fixes the joint expected degree-strength distribution to

$$\rho(\kappa, \sigma) = \rho(\kappa)\delta(\sigma - \sigma_0\kappa^{\eta}), \tag{54}$$

where $\delta()$ is the Dirac delta function, while the expected strengths are distributed as

$$\rho(\sigma) \sim \sigma^{-\delta}, \text{ where} \tag{55}$$

$$\delta = 1 + \frac{\gamma - 1}{\eta}. \tag{56}$$

The parameters of a WHSCM model satisfying the discussed requirements are thus

$$\gamma, \delta, \bar{k}, \sigma_0. \tag{57}$$

The reason for having σ_0 as a parameter instead of the more natural average strength $\bar{s}$, given by Eq. (43), is that the latter is actually infinite if $\delta < 2$ as discussed above. However, σ_0 is well defined even in this case, and controls the baseline of the scaling of strength as a function of degree. If $\delta > 2$, the expected average strength $\bar{s}$ is finite and in one-to-one relation to σ_0 via

$$\bar{s} = \sigma_0 \int \kappa^{\eta} \rho(\kappa)\, d\kappa. \tag{58}$$

IV. NUMERICAL EXPERIMENTS IN SEARCH OF A SOLUTION

We are to find the latent parameter distribution $\rho(\nu, \mu)$ that yields the distribution $\rho(\kappa, \sigma)$ of expected degrees and strengths that we want in Eq. (54). Unfortunately, the accomplishment of this task using brute force does not appear possible since it involves solving a system of nonlinear integral equations, as we will see below. Therefore, in this section, we retreat to numeric experiments and describe a workaround that relies on getting some experimental hints from the WSCM.

First, for convenience and consistency with the power-law HSCM described in Sec. II, we change variables from ν to λ via

$$\lambda = e^{R-\nu}. \tag{59}$$

The support of ν is $(-\infty, R]$, so that the support of λ is $[1, \infty)$. We will see that R is the parameter that controls the average degree $\bar{k}$ and constant σ_0, and that it grows logarithmically with n, Appendix C. With this change of variables, the connection probability and expected weight change from (27) and (29) to

$$p(\lambda_i, \mu_i, \lambda_j, \mu_j) = \frac{1}{1 + e^{2R} \cdot \frac{\mu_i + \mu_j}{\lambda_i \lambda_j}}, \tag{60}$$

$$\omega(\lambda_i, \mu_i, \lambda_j, \mu_j) = \frac{1}{1 + e^{2R} \cdot \frac{\mu_i + \mu_j}{\lambda_i \lambda_j}} \cdot \frac{1}{\mu_i + \mu_j}, \tag{61}$$

while the expected degree and strength as functions of the latent variables change from (35) and (36) to

$$\kappa(\lambda, \mu) = n \int_1^{\infty} d\lambda' \int_0^{\infty} d\mu' \, \frac{\rho(\lambda', \mu')}{1 + e^{2R} \cdot \frac{\mu + \mu'}{\lambda\lambda'}}, \tag{62}$$

$$\sigma(\lambda, \mu) = n \int_1^{\infty} d\lambda' \int_0^{\infty} d\mu' \, \frac{\rho(\lambda', \mu')}{\left(1 + e^{2R} \cdot \frac{\mu + \mu'}{\lambda\lambda'}\right)(\mu + \mu')}. \tag{63}$$

Observe that since weights are positive, the support of μ is $(0, \infty)$.

With this change of variables our task becomes to find $\rho(\lambda, \mu)$ producing $\kappa = \kappa(\lambda, \mu)$ and $\sigma = \sigma(\lambda, \mu)$ such that (1) κ is a power-law-distributed random variable (52) and (2) σ is a superlinear function of κ (53). The brute-force solution of this task is the solution of the system of the two two-dimensional nonlinear integral Eqs. (62) and (63) with respect to $\rho(\lambda, \mu)$, so that the resulting $\rho(\kappa, \sigma)$ is as required in (54). The required amount of brute force for this task is well beyond our analytical strength, so that we have to retreat to numeric investigations.

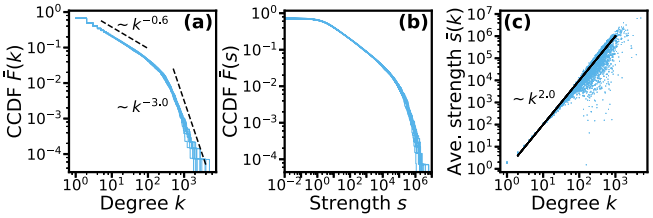

FIG. 2. Basic structural properties of networks in the WHSCM with power-law hidden-variable distributions in Eqs. (64)–(66). Ten networks of size $n = 20\,000$ are generated with $\alpha = 2.5$, $\beta = 2.0$, $a = 1.0$, and $R = 6.0$. The resulting average degree is about 20. (a) and (b) show the complementary cumulative distribution functions (CCDF) of degrees and strengths, while (c) shows the average strengths of nodes of degree k.

A. WHSCM with power-law hidden-variable distributions

The most reasonable choice of $\rho(\lambda, \mu)$ appears to be a clean power-law one. Indeed, as recalled in Sec. II B, such a clean power-law choice of the distribution of latent parameters results in power-law degree distributions in the HSCM, so that one may expect the situation to be similar in the WHSCM. Unfortunately, this expectation is not correct as we show next. To see this, let us set the distribution $\rho(\lambda)$ to be a pure power law, i.e., the Pareto distribution,

$$\rho(\lambda) = (\alpha - 1)\lambda^{-\alpha},\ \alpha > 2,\ \lambda > 1, \tag{64}$$

and couple λ and μ deterministically via

$$\rho(\lambda, \mu) = \rho(\lambda)\delta(\mu - f(\lambda)), \tag{65}$$

$$f(\lambda) = a\lambda^{-\beta},\ \beta \geqslant 0, \tag{66}$$

where $\delta()$ is the Dirac delta function, so that $\rho(\mu) \sim \mu^{(\alpha-1)/\beta-1}$ is another Pareto if $\beta > \alpha - 1$. We then generate random networks in this version of the WHSCM, and report their basic structural properties in Fig. 2. We see that contrary to our expectations, the degree and strength distributions do not look like clean power laws, and that the degree distribution appears to be a double power law. Appendix D contains some analytical arguments explaining this behavior.

B. Experimental hints from the WSCM

Given that the simple Pareto choice of the joint probability distribution $\rho(\lambda, \mu)$ does not lead to weighted networks with desired properties, we devise a workaround, looking for an ansatz for $\rho(\lambda, \mu)$ based on hints from the WSCM. Specifically, we attempt to "reverse engineer" the distribution $\rho(\lambda, \mu)$ by inferring the values of variables λ_i, μ_i in the WSCM for synthetically generated degree and strength sequences that satisfy our desired WHSCM constraints in Sec. III.

The inference is done using maximum likelihood estimation (MLE) by finding a set of values λ_i, μ_i that maximize the log-likelihood $L = \ln P(G)$ of a weighted graph G in the WSCM. The probability $P(G)$ of G in the WSCM ensemble is given by Eq. (25). The partition function Z in that equation is calculated in [15] to yield the following explicit expression for $P(G)$:

$$P(G) = \prod_{(i,j)\in\mathcal{P}} \frac{e^{-(\nu_i+\nu_j+\ln(\mu_i+\mu_j))a_{ij}}}{1 + e^{-(\nu_i+\nu_j+\ln(\mu_i+\mu_j))}} \cdot \prod_{(i,j)\in\mathcal{E}} (\mu_i + \mu_j)e^{-(\mu_i+\mu_j)w_{ij}}, \tag{67}$$

where a_{ij} and w_{ij} denote the entries of the adjacency and weight matrices of G, and $\mathcal{P}$ and $\mathcal{E}$ denote the sets of node pairs and connected node pairs in G. Using

$$\sum_{(i,j)\in\mathcal{P}} (\nu_i + \nu_j)a_{ij} = \sum_{i=1}^{n} k_i\nu_i, \tag{68}$$

$$\sum_{(i,j)\in\mathcal{P}} (\mu_i + \mu_j)w_{ij} = \sum_{i=1}^{n} s_i\mu_i, \tag{69}$$

where k_i, s_i are the degree and strength of node i in graph G, the expression for the logarithm-likelihood simplifies to

$$\ln P(G) = \sum_{i=1}^{n}[k_i\nu_i + s_i\mu_i] - \sum_{(i,j)\in\mathcal{P}} \ln\left[1 + \frac{1}{e^{\nu_i+\nu_j}(\mu_i+\mu_j)}\right]. \tag{70}$$

Observe that $\ln P(G)$ depends on graph G only via its degree $\boldsymbol{k}$ and strength $\boldsymbol{s}$ sequences, because they are the sufficient statistics in this grand-canonical ensemble. Furthermore, the probability $P(\boldsymbol{k}, \boldsymbol{s}|\boldsymbol{\kappa}, \boldsymbol{\sigma})$ of the degree-strengths sequence $\boldsymbol{k}, \boldsymbol{s}$ in the WSCM ensemble defined by the expected degree-strength sequence $\boldsymbol{\kappa}, \boldsymbol{\sigma}$ is at its maximum for $\boldsymbol{k} = \boldsymbol{\kappa}$ and $\boldsymbol{s} = \boldsymbol{\sigma}$ [10,16]. Therefore, to execute our MLE program, all we have to do is to generate synthetic degree-strength sequences satisfying the requirements in Sec. III, and then feed them to the MLE applied to the $\ln P(G)$ above.

We do so as follows: for $i = 1, \ldots, n$, we sample real random numbers x_i from the Pareto distribution with exponent $\gamma > 2$ and mean $\bar{x} = 10$, and then round them to the closest integers to yield degrees sequences $k_i = [x_i]$. The exponent γ and mean $\bar{x}$ of the Pareto distribution determine the $x_{\min} = \bar{x}(\gamma - 2)/(\gamma - 1)$ of its support $[x_{\min}, \infty)$, so that upon this rounding the $k_{\min}$ degree is statistically different from the other degrees, but this difference has no effect on the tail exponent of the resulting distribution of k_is, which is always guaranteed to be the same γ [32]. The strengths are then set to $s_i = \sigma_0 k_i^{\eta}$ for some $\sigma_0 > 0$ and $\eta \geqslant 1$. The obtained sequences of degrees $\{k_1, \ldots, k_n\}$ and strengths $\{s_1, \ldots, s_n\}$ are then supplied to the MLE inference of the parameters $\{\nu_1, \ldots, \nu_n\}$ and $\{\mu_1, \ldots, \mu_n\}$ by maximizing the $\ln P(G)$ in Eq. (70) which is done using the simulated annealing algorithm available in the PaGMO package [63]. The inferred $\{\nu_1, \ldots, \nu_n\}$ are then mapped to $\{\lambda_1, \ldots, \lambda_n\}$ via (59) with $R = \max_i \nu_i$. Another way to find λ_i, μ_i is to solve numerically the system of $2n$ equations in Eqs. (32) and (33). As noted in Ref. [12], this way leads to the same results. In our experiments, however, we find that the MLE approach produces more numerically stable results, so that we use this approach instead.

The obtained sequences $\{\lambda_1, \ldots, \lambda_n\}$ and $\{\mu_1, \ldots, \mu_n\}$ for different values of γ and η are shown in Fig. 3. From this figure, we extract several hints suggesting a possible shape of the joint distribution of latent variables $\rho(\lambda, \mu)$:

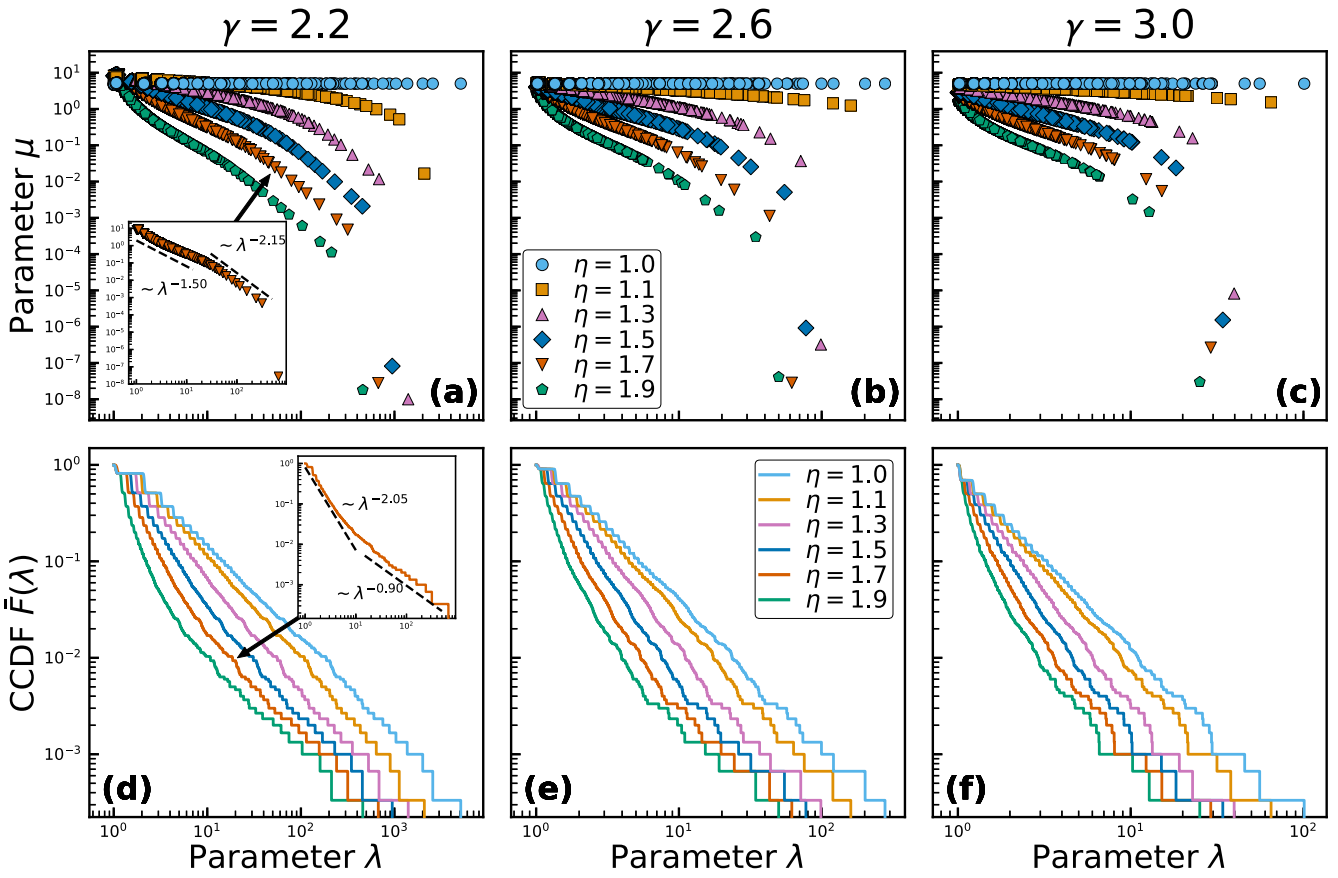

FIG. 3. Maximum likelihood estimation of the latent parameters in the WSCM with power-law degree and strength sequences. Degree sequences of length $n = 3000$ are sampled from the Pareto distribution with varying power-law exponent γ, and then rounded to the closest integer. The corresponding strengths are set to $s_i = \sigma_0 k_i^\eta$. The expected average degree is set to $\bar{k} = 10$ and $\sigma_0 = 0.1$. (a)–(c) show scatter plots of the inferred parameters λ, μ for varying values of η and $\gamma = 2.2,\ 2.6,$ and 3.0. The inset in (a) shows the λ-μ scatter plot for $\eta = 1.7$, $\gamma = 2.2$ along with the power-law fit lines for small-λ and large-λ regions. (d)–(f) show the complementary CDFs of the λ parameter for varying values of η and $\gamma = 2.2,\ 2.6,$ and 3.0. The inset in (d) shows the complementary CDF of the λ parameter for $\eta = 1.7$, $\gamma = 2.2$ along with the power-law fit lines for small-λ and large-λ regions.

(1) The variable μ can be directly coupled to the variable λ via some function $f(\lambda)$ as indicated by the λ-μ scatter plots;

(2) The visual inspection of the λ-μ scatter plots on the log-log scale suggests that this function scales approximately as a power law for small and large values of λ, potentially with two different exponents, i.e., $f(\lambda) \sim \lambda^{-\beta_1}$ when $\lambda \to 1$, and $f(\lambda) \sim \lambda^{-\beta_2}$ when $\lambda \gg 1$;

(3) The complementary CDF $\bar{F}(\lambda)$ behaves approximately as a power law for small and large values of λ, potentially with two different exponents α_1 and α_2.

V. WEIGHTED HYPERSOFT CONFIGURATION MODEL WITH POWER-LAW DEGREE AND STRENGTH DISTRIBUTIONS

Here we rely on the observations at the end of the previous section to specify a particular version of the WHSCM model that satisfies the requirements in Sec. III. These observations instruct us to set the joint distribution of latent parameters λ, μ to

$$\rho(\lambda, \mu) = \rho(\lambda)\delta(\mu - f(\lambda)), \tag{71}$$

where $\delta()$ is the Dirac delta function. This setting fixes the WHSCM-definitive distribution $\rho(\nu, \mu)$ via the change of variables (59), but $\rho(\lambda)$ and $f(\lambda)$ are still to be specified.

We specify them, again following the hints from the end of the previous section, as follows. The distribution of λ is set to

$$\rho(\lambda) = \begin{cases} A_1 \lambda^{-\alpha_1}, & 1 < \lambda \leqslant \lambda_c, \\ A_2 \lambda^{-\alpha_2}, & \lambda_c < \lambda < \infty, \end{cases} \tag{72}$$

where $\alpha_1 > 1$, $\alpha_2 > 1$, and λ_c is a crossover point between the two power laws with exponents α_1 and α_2, whereas A_1, A_2 are the normalization constants given by

$$A_1 = \frac{(\alpha_1 - 1)(\alpha_2 - 1)}{\lambda_c^{1-\alpha_1}(\alpha_1 - \alpha_2) + (\alpha_2 - 1)}, \tag{73}$$

$$A_2 = A_1 \lambda_c^{\alpha_2 - \alpha_1}, \tag{74}$$

while the function $f(\lambda)$ is set to

$$f(\lambda) = \begin{cases} a\lambda^{-\beta_1}, & 1 < \lambda \leqslant \lambda_c, \\ a\lambda_c^{\beta_2 - \beta_1}\lambda^{-\beta_2}, & \lambda_c < \lambda < \infty, \end{cases} \tag{75}$$

where $a > 0$, $\beta_1 \geqslant 0$, $\beta_2 \geqslant 0$. With these settings, our model is fully specified by the following set of parameters:

(1) exponents $\alpha_1,\ \alpha_2,\ \beta_1,\ \beta_2$, and (2) parameters R and a, that we move on to specify below. The double power law crossover point λ_c may also appear as a free parameter, but we set it to be a function of other parameters as follows.

Recall that the connection probability $p(\lambda_i, \mu_i, \lambda_j, \mu_j)$ in the model is given by Eq. (60). The crossover point λ_c is selected in such a way that for λ_i, λ_j below λ_c this connection probability can be approximated by dropping the 1 in the denominator, analogous to the classical limit approximation of the Fermi-Dirac distribution function in statistical physics [14]. This means that the constant λ_c defines the point where the term 1 in the denominator in (60) is comparable to the other term there. Therefore we define λ_c to be the value of λ_i, λ_j such that $e^{2R} \times \frac{\mu_i+\mu_j}{\lambda_i\lambda_j} = 1$, so that

$$\lambda_c = (2ae^{2R})^{\frac{1}{2+\beta_1}}. \tag{76}$$

We show in Appendix E that with the settings above, the expected degree of a node with latent parameter λ can be written as

$$\kappa(\lambda) = \begin{cases} I_1(\lambda) + I_2(\lambda), & 1 < \lambda \leqslant \lambda_c, \\ I_3(\lambda) + I_4(\lambda), & \lambda_c < \lambda < \infty, \end{cases} \tag{77}$$

where the integrals I_1-I_4 are explicitly defined in Eqs. (E1)–(E4). Similarly, the expected strength of a node with latent parameter λ can be written as

$$\sigma(\lambda) = \begin{cases} I_5(\lambda) + I_6(\lambda), & 1 < \lambda \leqslant \lambda_c, \\ I_7(\lambda) + I_8(\lambda), & \lambda_c < \lambda < \infty, \end{cases} \tag{78}$$

where the integrals I_5-I_8 are explicitly defined in Eqs. (E13)–(E16). We also show in Appendix E that the functions $\kappa(\lambda)$ and $\sigma(\lambda)$ can be well approximated by power laws

$$\kappa(\lambda) \sim \begin{cases} \lambda^{\phi_1}, & \text{if } 1 < \lambda \leqslant \lambda_c, \\ \lambda^{\phi_2}, & \text{if } \lambda_c < \lambda < \infty, \end{cases} \tag{79}$$

$$\sigma(\lambda) \sim \begin{cases} \lambda^{\chi_1}, & \text{if } 1 < \lambda \leqslant \lambda_c, \\ \lambda^{\chi_2}, & \text{if } \lambda_c < \lambda < \infty, \end{cases} \tag{80}$$

where $\phi_1 \geqslant 1$, $0 < \phi_2 \leqslant 1$, $\chi_1 \geqslant \phi_1$, $\chi_2 \geqslant \phi_2$, and that with these approximations, the distributions of expected degrees and strengths do indeed exhibit power-law behavior:

$$\rho(\kappa) \sim \begin{cases} \kappa^{-(1+\frac{\alpha_1-1}{\phi_1})}, & \text{if } \kappa(1) < \kappa \leqslant \kappa(\lambda_c), \\ \kappa^{-(1+\frac{\alpha_2-1}{\phi_2})}, & \text{if } \kappa(\lambda_c) < \kappa < \infty, \end{cases} \tag{81}$$

$$\rho(\sigma) \sim \begin{cases} \sigma^{-(1+\frac{\alpha_1-1}{\chi_1})}, & \text{if } \sigma(1) < \sigma \leqslant \sigma(\lambda_c), \\ \sigma^{-(1+\frac{\alpha_2-1}{\chi_2})}, & \text{if } \sigma(\lambda_c) < \sigma < \infty. \end{cases} \tag{82}$$

Now, if we know how the scaling exponents $\phi_1, \phi_2, \chi_1, \chi_2$ behave as functions of the model parameters, we can easily select those parameters to ensure that $\rho(\kappa) \sim \kappa^{-\gamma}$ for all values of λ, and that $\rho(\sigma) \sim \sigma^{-\delta}$ with $\delta = 1 + \frac{\gamma-1}{\eta}$ as required in Sec. III. In Appendix E, we find $\phi_1, \phi_2, \chi_1, \chi_2$ as functions of $\alpha_1, \alpha_2, \beta_1, \beta_2$, and show that the following nontrivial choice of the latter as functions of γ, η produces the desired outcome:

$$\alpha_1 = 1 + \eta(\gamma - 1), \tag{83}$$

$$\beta_1 = \left(\gamma - \frac{\gamma-2}{\gamma}\right)(\eta - 1), \tag{84}$$

$$\alpha_2 = 1 + \frac{\alpha_1 - 1}{1+\beta_1}\left[1 + (\gamma-2)\left(1 - \frac{1}{\eta}\right)\right], \tag{85}$$

$$\beta_2 = \begin{cases} 0, & \text{if } \eta = 1, \\ \alpha_2 - 1 + \frac{\eta(\alpha_2-1)}{\gamma-1}, & \text{if } \eta > 1. \end{cases} \tag{86}$$

The choice of the model parameters above is obtained using a series of approximations outlined in Appendix E. We find in simulations that this choice produces networks with desired properties if $\gamma > 2$ and $\eta \leqslant 2$ as we will see below.

Finally, we have to express the R and a parameters as functions of the average degree $\bar{k}$ and σ_0. In the λ terms, the average degree is equal to

$$\bar{k} = \int_1^\infty \kappa(\lambda)\rho(\lambda)\, d\lambda. \tag{87}$$

Both R and a appear in this integral. The second equation defining these two parameters is $\sigma = \sigma_0\kappa^\eta$. For any given value of the latent parameter $\lambda = \lambda_0$, we can write

$$\sigma_0 = \frac{\sigma(\lambda_0)}{[\kappa(\lambda_0)]^\eta}. \tag{88}$$

For the reasons explained in Appendix E, we set the λ_0 to the value such that $\kappa(\lambda_0) = \bar{k}$. By solving numerically the system of Eqs. (87) and (88) with the λ_0 set to this value, we find the values of the parameters R and a. With these solutions, the model is fully specified.

Weighted hypersoft configuration model with power-law degree and strength distributions

To summarize, the graphs in the described power-law WHSCM can be generated using the following algorithm.

(1) For a given set of input parameters, which are the network size n, the degree distribution power law exponent $\gamma > 2$, the strength-degree scaling exponent $\eta \geqslant 1$, the average degree $\bar{k}$, and the constant σ_0 controlling the baseline of the strength-degree scaling.

(2) Find the model parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, R, a$ using Eqs. (83)–(88).

(3) For each node $i = 1, \ldots, n$, sample λ_i from the PDF in Eq. (72), and set $\mu_i = f(\lambda_i)$ according to Eq. (75).

(4) For each node pair (i, j), draw the link between them with probability given by Eq. (60).

(5) For each node pair (i, j) linked at the previous step, draw the weight of the link from the exponential distribution with rate $\mu_i + \mu_j$, Eq. (28).

We provide an implementation of this algorithm at the GitHub repository [49].

VI. SIMULATION RESULTS FOR SYNTHETIC NETWORKS

Here we check in simulations that the specific version of the general WHSCM model documented at the end of the previous section produces networks satisfying the requirements in Sec. III.

We first generate graphs of size $n = 10^5$ with the average degree $\bar{k} = 10$ and $\sigma_0 = 0.1$ for $\gamma \in \{2.2, 2.6, 3.0\}$ and $\eta \in \{1.0, 1.1, 1.3, 1.5, 1.7, 1.9\}$. For each combination of the parameters, 20 graph instances are generated. The resulting empirical complementary CDFs of degrees k and the average

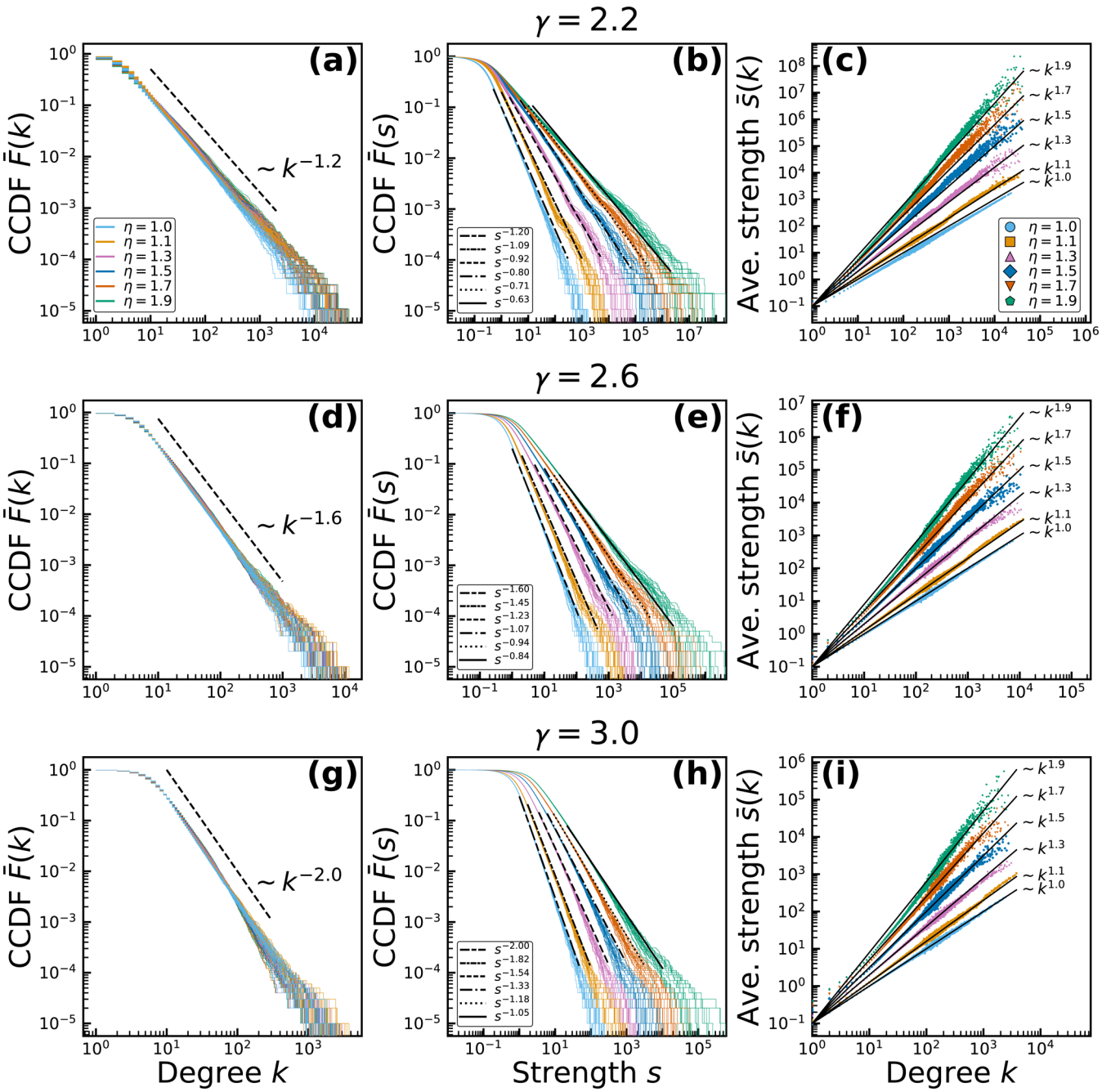

FIG. 4. Structural properties of synthetic networks in the power-law WHSCM model defined in Sec. V. The parameters used to generate the networks are $n = 10^5$, $\bar{k} = 10$, $\sigma_0 = 0.1$, $\gamma \in \{2.2, 2.6, 3.0\}$, and $\eta \in \{1.0, 1.1, 1.3, 1.5, 1.7, 1.9\}$. For each combination of the parameters, 20 network instances are generated. (a), (d), and (g) show the empirical complementary CDFs (CCDFs) of node degrees for $\gamma = 2.2,\ 2.6,$ and 3.0. Each panel shows 20 degree CCDF curves corresponding to the 20 network instances. The curves with the same value of exponent η are of the same color. The black dashed lines show the imposed values of the power-law exponent γ. Panels (b), (e), and (h) show the empirical CCDFs of node strengths for $\gamma = 2.2, 2.6, 3.0$. The black lines denote the power laws with the imposed values of $\delta - 1$ where the strength power-law exponent δ is related to γ and η via $\delta - 1 = (\gamma - 1)/\eta$. Panels (c), (f), (i) show the strength-degree correlations for $\gamma = 2.2, 2.6, 3.0$. The average strength $\bar{s}(k)$ of nodes of degree k is shown for all the 20 network instances. The curves with the same η are of the same color. The black lines show the functions $s(k) = \sigma_0 k^{\eta}$ with the imposed values of η. For clarity, the strength CCDFs are shown for $s \geqslant 10^{-2}$.

strengths $\bar{s}(k)$ of nodes of degree k are shown in Fig. 4. We see that the generated networks indeed have power-law degree distributions with the prescribed values of exponent γ, and that the strength-degree correlations follow the super-linear law with the prescribed values of exponent η. Figure 5 quantifies this further, also showing the convergence of the inferred values of γ and δ to their target values for the networks of growing sizes n. Since the networks are random, so are their inferred γ, δ. We see that the means of the distributions of γ, δ come closer to their target values as n grows, while their variances decrease. The lack of an exact match in certain cases even for the largest considered network size can

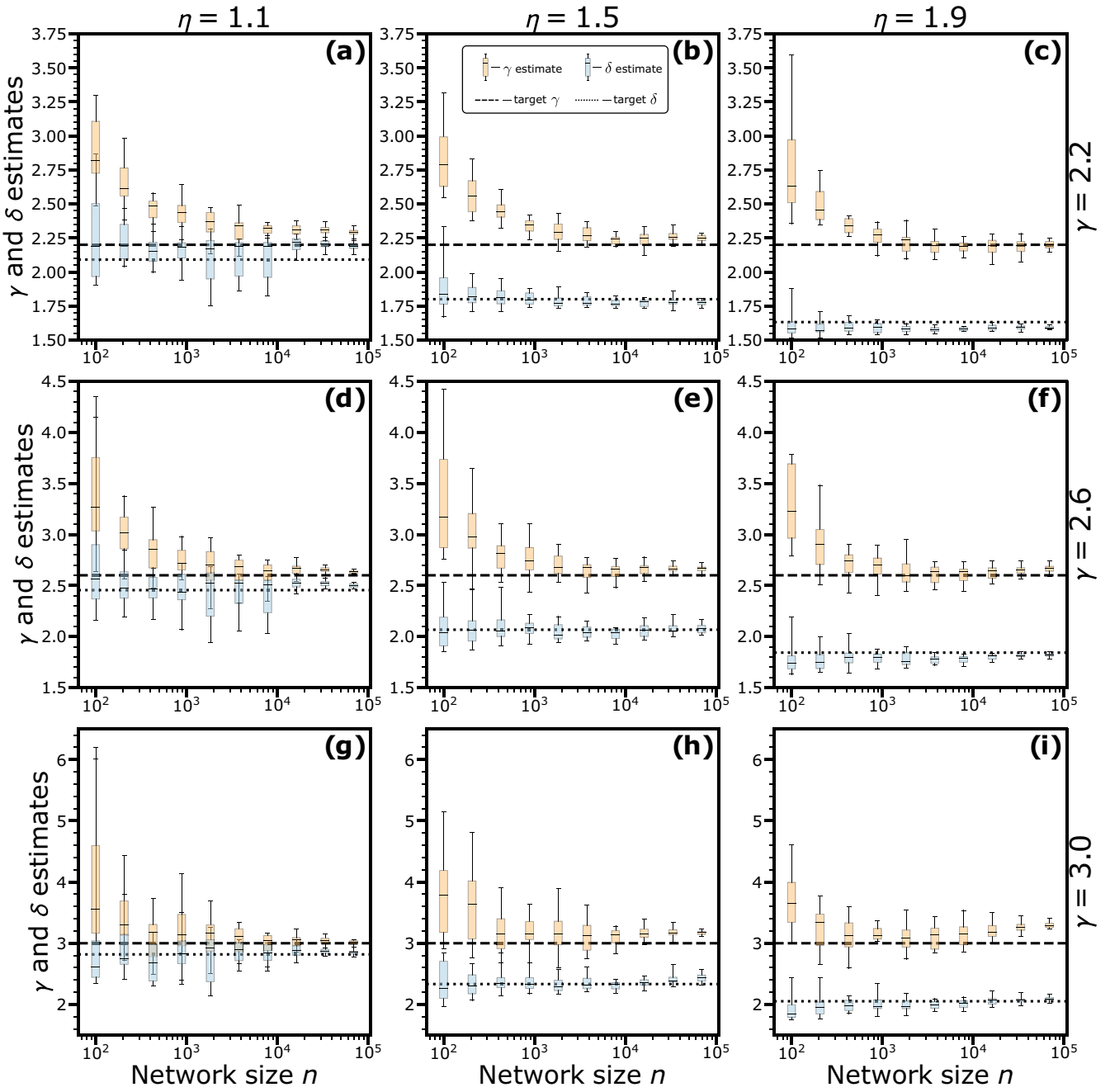

FIG. 5. Inferred power-law exponents of the degree and strength distributions in synthetic WHSCM networks. Synthetic networks of varying sizes $n \in [10^2, 10^5]$ are generated using the power-law WHSCM from Sec. V with the shown matrix of values of parameters γ, η. All networks have $\bar{k} = 10$ and $\sigma_0 = 0.1$. For each combination of parameters and sizes n, 20 random networks are generated. The degree distribution power-law exponent γ and the strength distribution power-law exponent δ are then inferred using the power-law exponent estimation software from [32]. The box plots show the distributions of the inferred values of γ, δ. The box plot settings are standard: the whiskers are the 5th and 95th percentiles, and the box boundaries are the 25th and 75th percentiles of the distributions; the black lines within the boxes are the medians. The dashed and dotted lines show the target values of γ and δ used to generate the networks.

be attributed to that this size is still not sufficiently large, as well as to imperfections of the model and the exponent estimator. The estimation precision of the latter for a given n is unknown [32].

To demonstrate how the degree-strength distributions, constrained in the maximum entropy manner, implicitly constrain some other network properties to some not exactly trivial values, we measure the *weight disparity* [64–66] in the generated networks. The weight disparity Y_i is a quantity that characterizes the heterogeneity of weights of links incident to node i. It is defined as

$$Y_i = \sum_{j \in \mathcal{N}_i} \left(\frac{w_{ij}}{s_i} \right)^2, \tag{89}$$

where $\mathcal{N}_i$ denotes the set of neighbors of node i, w_{ij} the weight of the link between nodes i and j, and s_i the strength

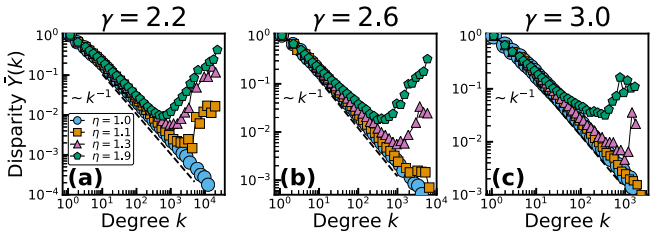

FIG. 6. Weight disparity as a function of node degree in the power-law WHSCM. The figure shows the log-binned average weight disparity $\bar{Y}(k)$ of nodes of degree k in the power-law WHSCM networks with $\gamma = 2.2,\ 2.6,\ 3.0$ and $\eta = 1.0,\ 1.1,\ 1.3,\ 1.9$. The black dashed lines show the $1/k$ scaling. In all the generated networks, the other parameters are set to $n = 10^5$, $\bar{k} = 10$, and $\sigma_0 = 0.1$.

of i. If most of the strength of node i is concentrated in the weights of a few links incident to i, then Y_i is close to 1. If all the links incident to i carry the same weight s_i/k_i, then Y_i is equal to $1/k_i$. To characterize the local distribution of weights in a weighted network as a whole, the average weight disparity $\bar{Y}(k)$ of nodes of degree k is often looked at Ref. [67]. If $\bar{Y}(k) \sim 1/k$ for all degrees k, then the weights are homogeneously distributed among all links and all nodes. An average weight disparity function decaying slower than $1/k$, as observed in many real networks [48,66], indicates that the link weights are distributed more heterogeneously.

In Fig. 6, we show the average weight disparity functions $\bar{Y}(k)$ in the generated WHSCM networks. We observe that the weight disparity behaves quite differently for different values of the scaling exponent η. For small values of η, the weight disparity as a function of degree scales roughly as $\bar{Y}(k) \sim 1/k$ in the lower to mid-range degree values. For larger values of γ this behavior persists even for large degrees, indicating that weights are distributed homogeneously for nodes of both low and high degrees. However, if η is larger, the high degree behavior of $\bar{Y}(k)$ changes entirely, and the function starts to increase with the degree k. This indicates that the higher the degree of a node, the more heterogeneous the distribution of weights among its links.

The intuition behind this effect is that the higher the exponent η, the larger the strengths of the high-degree nodes. In order for these nodes to satisfy their demanding strength constraints without disturbing the small strengths of low-degree nodes attached to them, they have to allocate increasingly heavier weights on their links to other high-degree nodes. This creates a weighted connectivity pattern where large portions of the strengths of high-degree nodes are distributed among the links interconnecting the high-degree, high-strength nodes. This weighted rich-club pattern is similar in spirit to the unweighted rich-club effect [25–27]. It is important to reemphasize that this seemingly nontrivial effect is caused purely by simple constraints on low-order network properties—namely, by heterogeneous degree distributions and superlinear scalings of strengths with degrees.

VII. POWER-LAW WHSCM VERSUS REAL-WORLD NETWORKS

Finally, we demonstrate a use case scenario involving the power-law WHSCM to construct null model graphs for the following three real-world weighted networks.

TABLE II. The WHSCM parameters measured in the three real-world networks.

Network name	n	$\bar{k}$	σ_0	γ	η
C. elegans metabolic [68–70]	453	8.94	1.3	2.5	1.154
Comp. geo. collab. [71]	6,158	3.86	1.0	2.6	1.333
Bible proper nouns [70,72]	1,773	10.30	0.66	3.1	1.313

(1) *C. elegans* metabolic network [68–70], where nodes are metabolites and links indicate interactions between them, with the link weight representing the number of such interactions.

(2) Computational geometry collaborations network [71], where nodes are authors publishing works on computational geometry and links between them represent co-authorship, with the link weight representing the number of co-authored works.

(3) Bible proper nouns network [70,72], where nodes are proper nouns of places and names in the King James Version of the Bible and links between them indicate that a pair of nouns are mentioned in the same Bible verse, with link weights representing the number of such noun co-occurrences.

To construct null model graphs for these networks, we first measure the WHSCM input parameters in these real networks. We find the n and $\bar{k}$ parameters directly from the networks, the power-law exponent γ is found using the package from Ref. [32], and the exponent η and σ_0 are found by a linear fit of degrees and strengths in the log-log scale. The resulting values for the three networks are shown in Table II.

Using the parameters from Table II, we generate ten synthetic WHSCM network instances for each of the three real networks, and compare the basic structural properties of the real networks and their WHSCM null model counterparts. Specifically, we look at the following properties: (1) the complementary CDF (CCDF) of degrees, $\bar{F}(k)$; (2) the average strength of nodes as a function of their degrees, $\bar{s}(k)$; (3) the average weight disparity of nodes as a function of their degrees, $\bar{Y}(k)$; (4) the average local clustering coefficient of nodes as a function of their degrees, $\bar{c}(k)$; and (5) the average weighted local clustering coefficient of nodes as a function of their strengths, $\bar{c}_w(s)$.

The last, less familiar property is defined in Ref. [73] for a node i as

$$c_w^{(i)} = \frac{1}{k_i(k_i-1)} \sum_{(j,k)\in\mathcal{N}(i)} (\tilde{w}_{ij}\tilde{w}_{jk}\tilde{w}_{ki})^{1/3}, \qquad (90)$$

where (j,k) are all the pairs of neighbors $\mathcal{N}(i)$ of the node i, and $\tilde{w}$ denotes the link weight rescaled by the maximum weight observed in the network. The function $\bar{c}_w(s)$ is the average of $c_w^{(i)}$ across all nodes i whose strength is s. If weights are real, every node has a unique strength with high probability, so that this function is a scatter plot consisting of n data points, one data point for every node.

Juxtaposing these properties in the considered real networks against their WHSCM counterparts in Fig. 7, we observe that the properties fixed by the WHSCM are well-captured by the null model graphs, as expected. Moreover, even though the weight disparity is not explicitly constrained

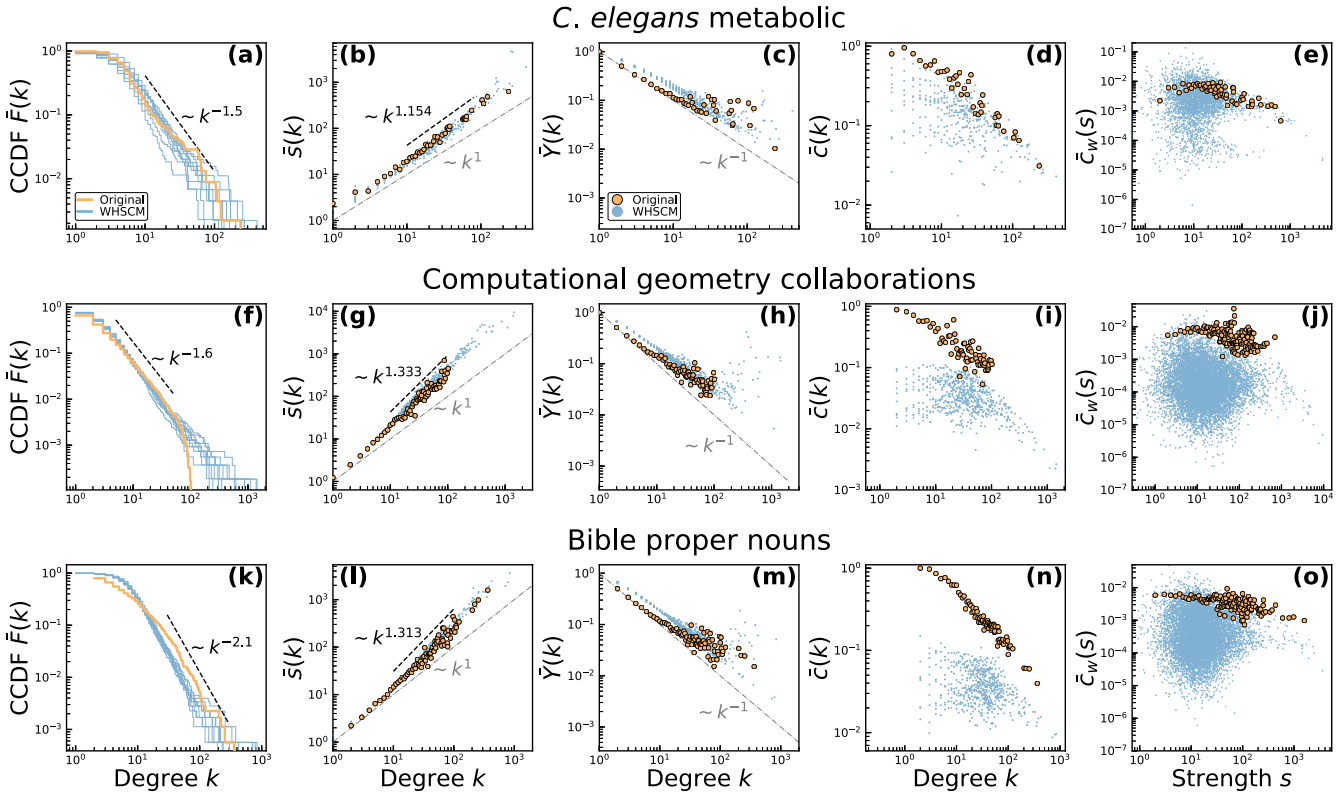

FIG. 7. Real networks vs their WHSCM counterparts. Each row in the figure corresponds to one of the three real weighted networks described in Sec. VII, as indicated by the network names at the top of each row. Each column in the figure shows different structural properties of the real networks: (a), (f), and (k) show the degree CCDFs, with the dashed blacked lines indicating the pure power-law scalings with the exponents γ listed in Table II; (b), (g), and (l) show the average strength as a function of node degree, $\bar{s}(k)$, with the black dashed lines indicating the pure scalings with the exponents η shown in Table II, and the gray dashed lines indicating the linear scaling; (c), (h), and (m) show the average weight disparity, defined in Eq. (89), as a function of node degree, with the gray dashed lines indicating the $1/k$ scaling of the weight disparity corresponding to the perfectly homogeneous distribution of local weights; (d), (i), and (n) show the average local clustering coefficient as a function of node degree; (e), (j), and (o) show the average weighted local clustering coefficient, defined in Eq. (90), as a function of node strength. The data for the real networks are shown in the orange color, while the data for the synthetic WHSCM networks are in blue. For each real networks, ten random WHSCM graphs are generated using the power-law WHSCM with the parameters listed in Table II.

in the WHSCM, it nevertheless behaves in a qualitatively similar manner in the real and null model networks.

However, both the unweighted and weighted clustering coefficients are much lower in the null models than in the real networks. To quantify this difference a bit further, we average the unweighted and weighted clustering coefficients across all nodes in the real networks and across all nodes in all the ten WHSCM replicas of the real networks. The results are shown in Table III. We see that in all the three real networks, both the unweighted and weighted average clustering is much higher, often by orders of magnitude, than in their null model counterparts. This observation suggests that the clustering coefficient is a statistically significant structural feature of these networks that cannot be explained by power-law degree and strength distributions alone. Some other mechanisms, such as latent geometry [48], are thus responsible for the formation of clustering in these real networks.

VIII. DISCUSSION

There exist a plenty of weighted network models with tunable degree and strength distributions [46–48,74–84]. Here, we have contributed to this list by introducing the unique unbiased null model, the WHSCM, that satisfies the maximum

TABLE III. The unweighted and weighted average local clustering coefficients $\bar{c}_{u,r}$, $\bar{c}_{w,r}$ in the considered real weighted networks and in their synthetic WHSCM counterparts, $\bar{c}_{u,s}$, $\bar{c}_{w,s}$.

Network name	$\bar{c}_{u,r}$	$\bar{c}_{u,s}$	$\bar{c}_{w,r}$	$\bar{c}_{w,s}$
C. elegans metabolic	6.5×10^{-1}	2.2×10^{-2}	6.6×10^{-3}	2.8×10^{-3}
Comp. geo. collab.	4.9×10^{-1}	3.2×10^{-2}	4.7×10^{-3}	8.2×10^{-5}
Bible proper nouns	7.2×10^{-1}	4.6×10^{-2}	5.2×10^{-3}	5.0×10^{-4}

entropy requirement. The model produces random graphs whose entropy is maximal across all graphs whose joint distribution of degrees and strengths converges to a given distribution. The outline of the proof that this is indeed so in Appendix A is not as detailed as the proof of the maximum entropy properties of the HSCM in Ref. [14]. The WHSCM proof can thus be improved by filling in all the missing details.

In developing a particular version of the model with power-law degree and strength distributions, we encountered the major challenge in the form of the system of nonlinear integral equations (35), (36) or (62), (63) that need to be solved to find the latent parameter distribution $\rho(\nu, \mu)$ or $\rho(\lambda, \mu)$. These equations appear intractable, so that we devised a workaround that worked well. Yet one may definitely question how good our approach is in general, and how valid, reliable, and accurate the double power-law modeling of the distributions in Fig. 3 is in particular. We dedicated Appendix E to support our double power-law assumption, but the argument there is not very rigorous. Is there a better, more accurate model for these distributions, possibly improving the convergence speeds in Fig. 5? More generally, is there an entirely different, more principled approach to the problem of finding (approximate) solutions of the systems of Eqs. (35), (36), (62), and (63)?

It would be nice to have such an approach because one may wish to constrain the degree-strength distributions not necessarily to power laws but to something else—to truncated power laws, for instance, observed in real weighted networks [3,85–87]. What is the latent parameter distribution in this case? In Fig. 2, we saw that a "clean" power-law choice of the latent parameter distribution $\rho(\lambda, \mu)$ led to truncated power laws for the marginals of $P(k, s)$, but even with this clean choice of $\rho(\lambda, \mu)$, things are difficult to control analytically as Appendix D shows. In any case, for any desired strength-degree distribution $\rho(\kappa, \sigma)$, the most principled solution is the exact solution of the system of integral Eqs. (62) and (63) with respect to the latent parameter distribution $\rho(\lambda, \mu)$.

We emphasize that the introduced null model is what it is, a null model, so that it should be used as such. It should not be confused with or considered as a realistic model of real-world weighted networks. We saw, for instance, that the model does not capture clustering observed in real networks. Latent geometry was proposed in Ref. [48] as a possible mechanism explaining clustering in real weighted networks. It would be interesting to see whether the model in Ref. [48] satisfies the maximum entropy requirements, and if so, then under what constraints.

Related to that, it would be also nice to have a weighted generalization of random hyperbolic graphs [88] whose maximum entropy properties are well understood and whose infinite temperature limit is exactly the HSCM [89,90]. In other words, what is the model of weighted random hyperbolic graphs that have analogous maximum entropy properties and whose infinite temperature limit is the WHSCM?

ACKNOWLEDGMENTS

We thank G. Cimini, D. Garlaschelli, R. Mastrandrea, A. Allard, M. Á. Serrano, and M. Boguñá for useful discussions and suggestions. This work was supported by NSF Grant No. IIS-1741355 and ARO Grants No. W911NF-17-1-0491 and No. W911NF-16-1-0391. M.K. was additionally supported by the NExTWORKx project. F.P. acknowledges support by the TV-HGGs project (OPPORTUNITY/0916/ERC-CoG/0003), funded by the Cyprus Research and Innovation Foundation.

APPENDIX A: (W)HSCM AS ENTROPY MAXIMIZERS

Here we summarize the key points of the proof from Ref. [14] that the HSCM random graphs maximize graph entropy across all graphs whose degree distribution converges to a given distribution, and show how to generalize this proof to weighted graphs in the WHSCM.

1. HSCM as entropy maximizer

For a given distribution $P(k)$ of node degrees k, what is the graph ensemble whose ensemble distribution $P(G)$ of graphs G maximizes Shannon entropy (1) but in such a way that the degree distribution in the ensemble converges to $P(k)$? As proved in Ref. [14], the unique answer is the HSCM.

The proof is not exactly trivial because any brute-force attack at it is doomed to fail since if understood literally, the task is an intractable combinatorial optimization problem, a mission impossible. To circumvent this impasse, a workaround collection of techniques, based on the graphon theory [60], was devised in Ref. [14]. We describe this collection here.

The main idea of this workaround proof is to maximize not the entropy $S[P(G)]$ of the intractable ensemble distribution $P(G)$ but the graphon entropy defined below. Intuitively, the graphon entropy is the entropy of random edges conditioned on given values of the Lagrange multipliers ν_i. In other words, for a given collection of fixed ν_is, the graphon entropy is the SCM entropy. The maximization of the graphon entropy turns out to be a tractable functional analysis problem with an explicit unique solution, but there is another contribution to the HSCM entropy which is the entropy of ν_is that are not fixed but random in the HSCM. The crux of the proof is to show that the entropy of the graphon that maximizes the graphon entropy is the leading term in the graph entropy $S[P(G)]$, while the entropy of ν_is is subleading. Since so, the graphon that maximizes the graphon entropy maximizes also the graph entropy, thus solving the original entropy maximization problem. We provide some key details behind how this works next.

In the HSCM, the graphon is no mystery but just the connection probability function $p(\nu, \nu')$. This function literally says that if the Lagrange multipliers of two nodes happen to be ν and ν', then the link between them is the Bernoulli random variable with the success rate $p(\nu, \nu')$. Observe that the Bernoulli random variable $\mathrm{Be}(p)$ is trivially the random variable that maximizes the Shannon entropy of distributions on $\{0, 1\}$ with mean p. Recall that the Lagrange multipliers ν as random variables can be mapped to the expected degrees κ via (10) and (11) resulting in the graphon $p(\kappa, \kappa')$ expressed as a function of κs. Observe that if κ is now treated as a latent variable, then the expected degree of a node with latent variable κ is κ:

$$\kappa = n \int p(\kappa, \kappa')\rho(\kappa')\, d\kappa'. \tag{A1}$$

The edges a_{ij} in the HSCM are Bernoullis with different success rates $p(\kappa_i, \kappa_j)$ that are random because κs are random. The entropy $S[\mathrm{Be}(p)]$ of the Bernoulli random variable with the success rate p is

$$S[\mathrm{Be}(p)] = -p \ln p - (1-p)\ln(1-p). \tag{A2}$$

These observations justify the definition of the entropy $S[p]$ of the graphon $p()$ in Ref. [60] which is

$$S[p] = \iint S[\mathrm{Be}(p(\kappa, \kappa'))]\rho(\kappa)\rho(\kappa')\, d\kappa\, d\kappa'. \tag{A3}$$

What is the graphon $p()$ that maximizes the graphon entropy in Eq. (A3) while satisfying the constraint in Eq. (A1), where $\rho(\kappa)$ is our desired expected degree distribution, the one that yields the desired $P(k)$? As shown in Ref. [14], it is relatively simple to prove that the unique exact answer is the Fermi-Dirac graphon in Eq. (6) with νs mapped to κs via (10) and (11). As also shown in Ref. [14], in sparse graphs this exact solution is asymptotically equivalent to the approximate expressions in Eqs. (13) and (14) that express the solution graphon explicitly in terms of the κ variables.

Proving that the entropy of this graphon $S[p]$ is the dominating term in the graph entropy $S[P(G)]$, in comparison to entropy of random κs, is a much more delicate endeavor. For this, the following techniques from [60] are properly adjusted in Ref. [14].

First, it is easy to see that the graphon entropy is a trivial lower bound for the HSCM graph entropy divided by $\binom{n}{2}$. Indeed, if all κs are fixed, then the graphon entropy in the HSCM is the entropy of the SCM graphs with this graphon divided by $\binom{n}{2}$. The hard part is thus to find a matching upper bound, and this is where the techniques from Ref. [60] come really useful.

The key point in establishing such an upper bound is to recognize that for any partition of the values of κs into consecutive intervals π_k, $k = 1, \ldots, K$, the entropy of $P(G)$ is upper-bounded by the entropy of the averaged graphon defined below, plus the entropy of the indicator random variables I_{ik} that indicate whether the random expected degree κ_i of node i happened to land in the interval π_k. The averaged graphon is defined as the piecewise constant function $\bar{p}(\kappa, \kappa')$ whose values for the values of κ, κ' belonging to a given rectangle $\pi_k \times \pi_{k'}$ in the $\kappa \times \kappa'$ partition are equal to the average value of the graphon $p(\kappa, \kappa')$ in this rectangle. Observe that the smaller the number of the partition intervals K, the smaller the total entropy of the indicator random variables I_{ik}, just because there are fewer of them, but the larger the sum of the error terms coming from graphon averaging, simply because rectangles $\pi_k \times \pi_{k'}$ are large. The smaller they are, the smaller the total graphon entropy error term, but the larger the total entropy of indicators I_{ik}. The crux of the proof is to find a "sweet spot"—the right number of intervals of the right size guaranteeing the proper balance between these two types of contributions to the upper bound, which we want to be tighter than the difference between the graph and graphon entropies. In sparse graphs, this task turns out to be a blade runner exercise.

Notwithstanding these blade runner difficulties, the required partition π_k was found in [14], completing the proof that the HSCM is indeed the unique entropy maximizer across all random graph ensembles whose degree distribution converges to a given $P(k)$.

2. WHSCM as entropy maximizer

The maximum entropy HSCM proof described in the previous section should apply to the WHSCM as well, upon the modifications that we discuss below. The key idea behind these modifications is a proper generalization of graphons and their entropy to weighted networks.

Similar to the unweighted case, in the weighted case it is more convenient to deal with the expected degree and strength variables κ, σ instead of the Lagrange multipliers ν, μ. The map from the latter to the former is given by Eqs. (35) and (36) which, if rewritten in the κ, σ variables, become the system of the self-consistency equations

$$\kappa = n \iint p(\kappa, \sigma, \kappa', \sigma')\rho(\kappa', \sigma')\, d\kappa'\, d\sigma', \tag{A4}$$

$$\sigma = n \iint \omega(\kappa, \sigma, \kappa', \sigma')\rho(\kappa', \sigma')\, d\kappa'\, d\sigma', \tag{A5}$$

analogous to Eq. (A1).

In unweighted networks, graph edges a_{ij} are Bernoullis with different random success rates $p(\kappa_i, \kappa_j)$:

$$a_{ij} = \mathrm{Be}[p(\kappa_i, \kappa_j)]. \tag{A6}$$

In weighted networks, the edges are no longer Bernoullis. Instead they are random variables that we call Bernoulli exponential, or BeExp for short:

$$w_{ij} = \mathrm{BeExp}[p(\kappa_i, \sigma_i, \kappa_j, \sigma_j), \omega(\kappa_i, \sigma_i, \kappa_j, \sigma_j)], \tag{A7}$$

where p and ω are the two parameters of the BeExp.

We define the vanilla BeExp as follows: if $w = \mathrm{BeExp}(p, \omega)$, where $p \in [0, 1]$, $\omega > 0$, and $w \geqslant 0$, then $w = 0$ with probability $1 - p$, while with probability p, w is the exponential random variable with mean ω (or rate $1/\omega$). In other words, the PDF of the BeExp $w = \mathrm{BeExp}(p, \omega)$ is

$$P(w) = \begin{cases} 1-p, & \text{if } w = 0, \\ \frac{p}{\omega} e^{-w/\omega}, & \text{if } w > 0, \end{cases} \tag{A8}$$

so that its entropy is

$$\begin{aligned} S[\mathrm{BeExp}(p, \omega)] &= -\int_0^\infty P(w) \ln P(w)\, dw \\ &= -p \ln p - (1-p)\ln(1-p) \\ &\quad + p(1 + \ln \omega) \\ &= S[\mathrm{Be}(p)] + pS[\mathrm{Exp}(\omega)], \end{aligned} \tag{A9}$$

where $S[\mathrm{Exp}(\omega)] = 1 + \ln \omega$ is the entropy of the exponential distribution with mean ω.

We observe that the $\mathrm{BeExp}(p, \omega)$ can be intuitively thought of as a "smearing" of the probability p of 1 (edge existence) in $\mathrm{Be}(p)$ into the $\mathrm{Exp}(\omega)$, the exponential distribution with mean ω (edge weight). As $\mathrm{Be}(p)$ is trivially the maximum entropy distribution on $\{0, 1\}$ with mean p, so is $\mathrm{BeExp}(p, \omega)$, less trivially, the maximum entropy distribution on $[0, \infty)$ with mean ω and $P(w > 0) = p$. That is, the $\mathrm{BeExp}(p, \omega)$ is the maximum entropy distribution under the constraints that the edge exists with probability p and that its mean weight is ω. It is common knowledge in statistical mechanics that in maximum entropy canonical ensembles of systems

of particles, the distributions of particles over particle states are also maximum entropy (Fermi-Dirac or Bose-Einstein). Since graph edges are analogous to particles in statistical mechanics [6,9,15,16], these observations motivate us to constrain the space of all possible probability distributions on $[0, \infty)$ to the two-parametric maximum entropy BeExp family.

Equation (A7) says that all edges in our weighted networks are BeExp's, albeit with different random parameters which are functions $p(\kappa, \sigma, \kappa', \sigma')$ and $\omega(\kappa, \sigma, \kappa', \sigma')$ of random κ, σ. Similar to the unweighted case, these observations instruct us to define the weighted graphon to be the parameters of our maximum entropy BeExp random variable. That is, we define a weighted graphon as *the pair of functions* $\{p(\kappa, \sigma, \kappa', \sigma'), \omega(\kappa, \sigma, \kappa', \sigma')\}$. Similar to the unweighted graphon entropy in Eq. (A3), we then define the weighted graphon entropy as

$$S[p, \omega] = \iiiint S\{\text{BeExp}[p(\kappa, \sigma, \kappa', \sigma'), \omega(\kappa, \sigma, \kappa', \sigma')]\} \times \rho(\kappa, \sigma)\, \rho(\kappa', \sigma')\, d\kappa\, d\sigma\, d\kappa'\, d\sigma'. \tag{A10}$$

The rest of the proof then proceeds as in the unweighted case, using the same techniques as in Ref. [14]: first show that the unique graphon maximizing the graphon entropy in Eq. (A10) subject to the constraints in Eqs. (A4) and (A5) is given by Eqs. (27) and (29) with ν, μ mapped to κ, σ via Eqs. (35) and (36), and then prove that the entropy of this graphon dominates the graph entropy, while the entropy of random κ, σ is negligible. The latter step could be challenging as it calls for repeating the blade runner partition finding exercise from Ref. [14], this time for the product space of $\kappa \times \sigma$ values. This should be still possible using the same ideas as in Ref. [14]—roughly, the key idea is that the partition is such that all its boxes have the same number of nodes in them on average. However, for the specific power-law WHSCM considered in this paper, or for any other WHSCM version in which strengths σ are set to be a deterministic function of degrees κ, we only need to partition the space of κ values. That is, the settings are exactly as in Ref. [14] in that regard, so that exactly the same partition as in Ref. [14] can be used in these cases.

APPENDIX B: SIMULATION RESULTS FOR THE RELATION BETWEEN EXPECTED AND ACTUAL DEGREES AND STRENGTHS

The WHSCM is formulated in terms of constraints for the joint distribution of *expected* degrees and strengths, while in many cases we are interested in the behavior of *actual* degrees and strengths realized in the corresponding ensemble of graphs. Thus we need to show that the results obtained so far for expected values also hold for actual degrees and strengths. For latent variable graph models, it is known that actual degrees are concentrated around their expected values, and distributed according to Poisson distribution, i.e., $P(k|\kappa) = \text{Pois}(\kappa)$ [40]. Since no similar claims are known for the behavior of strengths, we test the correlation between both κ and k, and σ and s values in our model. To this end, we constructed log-binned scatter plots of actual degrees/strengths versus their expected values in WHSCM graphs with varying parameters γ and η. The results are shown in Fig. 8. From the figure, it is evident that both degrees and strengths are highly correlated with their expected values. The narrow error bars also indicate that the distribution of actual degree/strength values around their expected values are narrow. Thus, the power-law scalings obtained for the expected values should also hold for actual values, as we have already demonstrated in the main text for various values of γ and η.

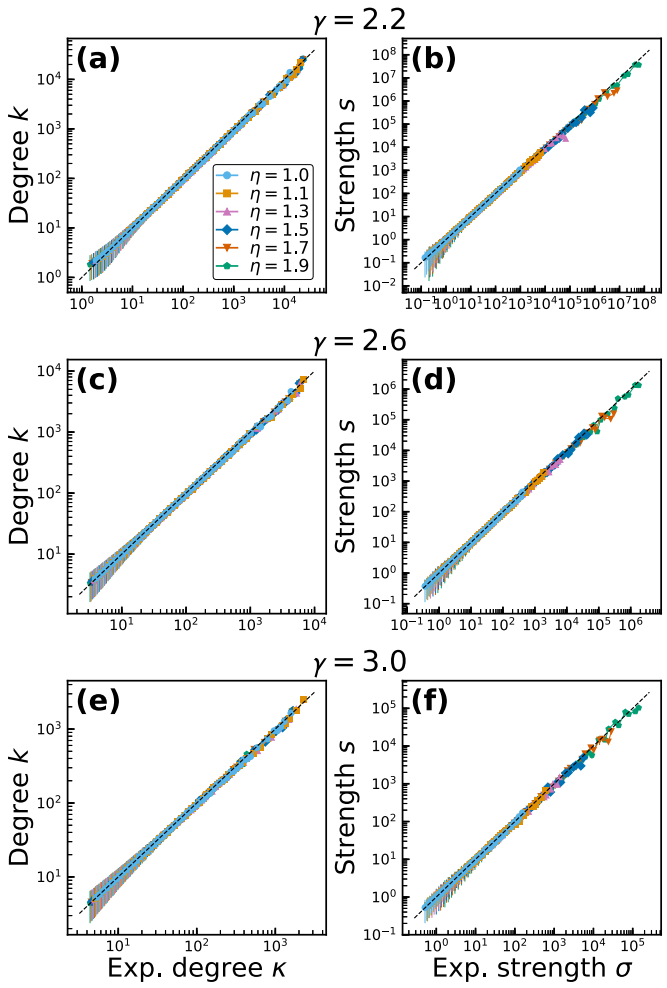

FIG. 8. Correlation of the expected and actual degrees/strengths in the WHSCM. Each row corresponds to the WHSCM input parameter γ indicated on the top. Each color-coded plot corresponds to an η parameter indicated in the legend. The rest of the WHSCM input parameters are $n = 10,000$, $\bar{k} = 10$, and $\sigma_0 = 0.1$. Data is log-binned. Error bars show standard deviation from the bin mean. Black dashed line shows perfect linear correlation.

APPENDIX C: SCALING OF R WITH THE NUMBER OF NODES

In the unweighted HSCM, the parameter R scales as $R \sim \frac{1}{2} \ln n$ with the network size n, which is evident from Eq. (12). This motivated us to assume a similar scaling for the analysis of the WHSCM. In this Appendix, we validate this choice by studying numerically how the parameters R and a scale with the system size in the WHSCM.

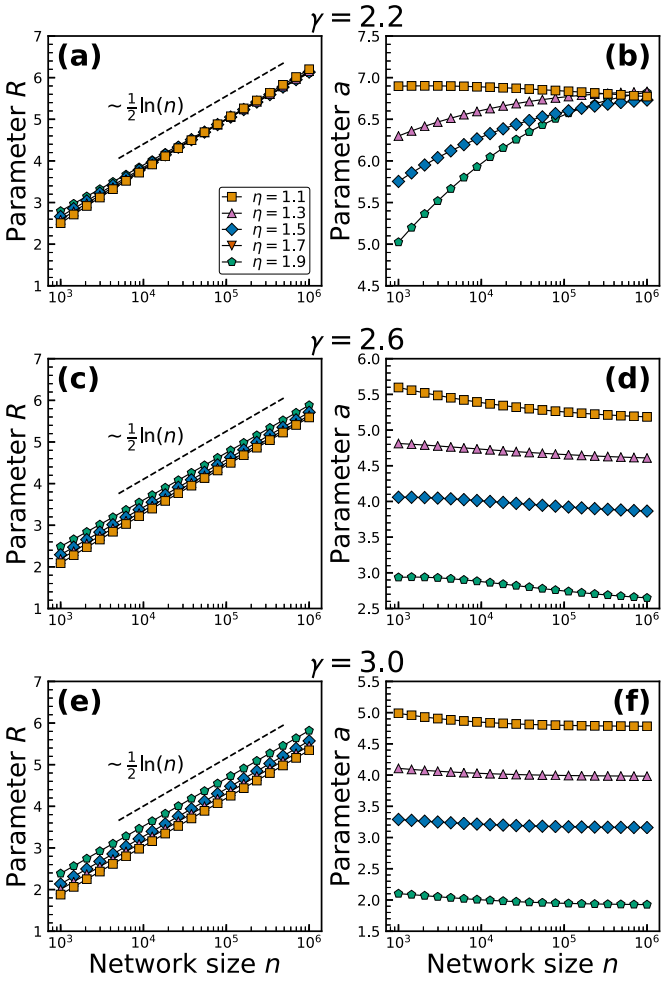

FIG. 9. Scaling of the R and a parameters as a function of network size n for various values of γ, η, and fixed values of $\bar{k} = 10$ and $\sigma_0 = 0.1$.

To this end, we solve Eqs. (87) and (88) as function of network size n for various input exponents γ, η, and fixed values of average degree $\bar{k} = 10$, and expected strength-degree scaling constant $\sigma_0 = 0.1$. The resulting solution curves are shown in Fig. 9. The solutions indicate that the parameter R scales with n in the same way as in the unweighted HSCM case, while the parameter a varies slowly with the network size. We note that since we solve for the R, a numerically with fixed average degree requirement, the resulting networks are guaranteed to have constant average degree independent of network size, thus forming a sparse ensemble of graphs.

APPENDIX D: WHSCM WITH POWER-LAW $\rho(\lambda, \mu)$

Here we show that the WHSCM with the pure power-law joint distribution of latent parameters

$$\rho(\lambda, \mu) = \rho(\lambda)\delta(\mu - f(\lambda)), \tag{D1}$$

$$\rho(\lambda) = (\alpha - 1)\lambda^{-\alpha},\ \alpha > 2, \tag{D2}$$

$$f(\lambda) = a\lambda^{-\beta},\ \beta \geqslant 0 \tag{D3}$$

does *not* produce weighted networks with clean power-law distributions of degrees and strengths. We show this for the degree distribution $P(k)$. Similar arguments apply to the strength distribution $P(s)$.

The expected degree of a node with the latent variable λ is given by

$$\kappa(\lambda) = n\int_1^\infty \frac{(\alpha-1)\lambda'^{-\alpha}}{1 + ae^{2R\left[\frac{\lambda^{-\beta}+\lambda'^{-\beta}}{\lambda\lambda'}\right]}} d\lambda'. \tag{D4}$$

While this integral is not expressible in a closed form in general, it is possible to find its approximations for different values of parameters λ and β. For large β and $\lambda \to 1$, the approximation is

$$n\int_1^\infty \frac{(\alpha-1)\lambda'^{-\alpha}}{ae^{2R\left[\frac{\lambda^{-\beta}+\lambda'^{-\beta}}{\lambda\lambda'}\right]}} d\lambda' = \frac{n}{ae^{2R}}\left(\frac{\alpha-1}{\alpha-2}\right)\lambda^{1+\beta}\cdot {}_2F_1\left(1, \frac{\alpha-2}{\beta}, 1+\frac{\alpha-2}{\beta}, -\lambda^\beta\right). \tag{D5}$$

For large λ, the approximation is

$$n\int_1^\infty \frac{n(\alpha-1)\lambda'^{-\alpha}}{1+\frac{ae^{2R}}{\lambda\lambda'^{1+\beta}}} d\lambda' = n\,{}_2F_1\left(1, \frac{\alpha-1}{1+\beta}, \frac{\alpha+\beta}{1+\beta}, -\frac{ae^{2R}}{\lambda}\right). \tag{D6}$$

These two approximations have different scalings of $\kappa(\lambda)$ with λ that can be approximated by double power laws with two different exponents $\phi_1 \geqslant \phi_2 > 0$. While we are not able to obtain closed-form expressions for the exponents ϕ_1, ϕ_2 in terms of the model parameters α, β in general case, from numerical simulations we observe that networks generated according to this version of the WHSCM have degree distributions with double-power-law-like behavior. At least the degree distribution does definitely not look like a power but as a power law with a power-law cutoff as Fig. 2 demonstrates.

In other words, a single power-law distribution with exponent α of the latent parameter λ results in two different scaling exponents for the resulting degree distribution $P(k)$, that are equal to $-(1 + (\alpha - 1)/\phi_1)$ for small degrees k, and $-(1 + (\alpha - 1)/\phi_2)$ for large degrees k. Such double-scaling behavior is yet another motivation to use double power law $\rho(\lambda)$ that would match the two scalings and result in single scaling of $P(k)$ with single target exponent γ.

APPENDIX E: ANALYSIS OF EXPECTED DEGREES AND STRENGTHS IN THE WHSCM

In the WHSCM, each node is characterized by the two latent parameters λ and μ that are distributed according to the joint probability distribution from Eq. (71). In Sec. V, we stated expressions for the exponent $\alpha_1, \alpha_2, \beta_1$ and β_2 for the densities of λ and μ. They are obtained from analysis of the integral expressions for both κ and σ in Eqs. (62) and (63), respectively.

1. Approximating the integral expressions for expected degrees

By our double power-law choice for the marginal distributions for λ and μ, the expected degree of a node with latent

parameter λ, Eq. (62), may be written as the following four integrals, where each integral represents one of combinations for "small" ($\leqslant\lambda_c$) and "large" ($>\lambda_c$) parameters λ and λ':

$$I_1(\lambda) = n\int_1^{\lambda_c} \frac{A_1\lambda'^{-\alpha_1}}{1+ae^{2R}\left(\frac{\lambda^{-\beta_1}+\lambda'^{-\beta_1}}{\lambda\lambda'}\right)}d\lambda', \tag{E1}$$

$$I_2(\lambda) = n\int_{\lambda_c}^{\infty} \frac{A_2\lambda'^{-\alpha_2}}{1+ae^{2R}\left(\frac{\lambda^{-\beta_1}+\lambda_c^{\beta_2-\beta_1}\lambda'^{-\beta_2}}{\lambda\lambda'}\right)}d\lambda', \tag{E2}$$

$$I_3(\lambda) = n\int_1^{\lambda_c} \frac{A_1\lambda'^{-\alpha_1}}{1+ae^{2R}\left(\frac{\lambda_c^{\beta_2-\beta_1}\lambda^{-\beta_2}+\lambda'^{-\beta_1}}{\lambda\lambda'}\right)}d\lambda', \tag{E3}$$

$$I_4(\lambda) = n\int_{\lambda_c}^{\infty} \frac{A_2\lambda'^{-\alpha_2}}{1+a\lambda_c^{\beta_2-\beta_1}e^{2R}\left(\frac{\lambda^{-\beta_2}+\lambda'^{-\beta_2}}{\lambda\lambda'}\right)}d\lambda'. \tag{E4}$$

While these integrals cannot be computed directly in a closed-form, it is possible to approximate them. For (E1), we use that for both $\lambda, \lambda' < \lambda_c$ the 1 in the denominator can be removed. This cannot be done for the other three integrals, since $\lambda' > \lambda_c$ or $\lambda > \lambda_c$. Therefore, in the second integral (E2), we use that $\lambda_c^{\beta_2-\beta_1}\lambda'^{-\beta_2} \leqslant \lambda_c^{-\beta_1}$ and hence this term is negligible with respect to $\lambda^{-\beta_1}$ when $1 \leqslant \lambda \leqslant \lambda_c$. A similar approach is applied to (E3), with the role of λ and λ' reversed. Finally, for integral (E4), we use that for $\lambda, \lambda' > \lambda_c$,

$$\lambda_c^{\beta_2-\beta_1}\left(\frac{\lambda^{-\beta_2}+\lambda'^{-\beta_2}}{\lambda\lambda'}\right) < 2\lambda_c^{-(2+\beta_1)},$$

and, given that $\lambda_c \approx (2ae^{2R})^{1/(2+\beta_1)}$ and $ae^{2R} \sim n$, as we demonstrate in Appendix C, we have

$$2\lambda_c^{-(2+\beta_1)} \approx \frac{1}{ae^{2R}} \ll 1,$$

so that the 1 is the dominant term in the denominator.

This allows us to obtain the following approximations for the integrals above:

$$I_1(\lambda) \approx n\int_1^{\lambda_c} \frac{A_1\lambda'^{-\alpha_1}}{ae^{2R}\left(\frac{\lambda^{-\beta_1}+\lambda'^{-\beta_1}}{\lambda\lambda'}\right)}d\lambda', \tag{E5}$$

$$I_2(\lambda) \approx n\int_{\lambda_c}^{\infty} \frac{A_2\lambda'^{-\alpha_2}}{1+ae^{2R}\left(\frac{\lambda^{-1-\beta_1}}{\lambda'}\right)}d\lambda', \tag{E6}$$

$$I_3(\lambda) \approx n\int_1^{\lambda_c} \frac{A_1\lambda'^{-\alpha_1}}{1+ae^{2R}\left(\frac{\lambda'^{-1-\beta_1}}{\lambda}\right)}d\lambda', \tag{E7}$$

$$I_4(\lambda) \approx n\int_{\lambda_c}^{\infty} A_2\lambda'^{-\alpha_2}d\lambda'. \tag{E8}$$

The four integrals on the right-hand side can be evaluated exactly:

$$I_1(\lambda) \approx \frac{nA_1\lambda^{1+\beta_1}}{ae^{2R}(\alpha_1-2)}\cdot\left[{}_2F_1\left(1, \frac{\alpha_1-2}{\beta_1}, 1+\frac{\alpha_1-2}{\beta_1}, -\lambda^{\beta_1}\right)\right.$$
$$\left. - \lambda_c^{2-\alpha_1}{}_2F_1\left(1, \frac{\alpha_1-2}{\beta_1}, 1+\frac{\alpha_1-2}{\beta_1}, -\left(\frac{\lambda}{\lambda_c}\right)^{\beta_1}\right)\right], \tag{E9}$$

$$I_2(\lambda) \approx \frac{nA_2\lambda_c^{1-\alpha_2}}{\alpha_2-1}{}_2F_1\left(1, \alpha_2-1, \alpha_2, -\frac{ae^{2R}}{\lambda_c\lambda^{1+\beta_1}}\right), \tag{E10}$$

$$I_3(\lambda) \approx \frac{nA_1}{\alpha_1-1}\left[{}_2F_1\left(1, \frac{\alpha_1-1}{1+\beta_1}, \frac{\alpha_1+\beta_1}{1+\beta_1}, -\frac{ae^{2R}}{\lambda}\right)\right.$$
$$\left. - \lambda_c^{1-\alpha_1}{}_2F_1\left(1, \frac{\alpha_1-1}{1+\beta_1}, \frac{\alpha_1+\beta_1}{1+\beta_1}, -\frac{ae^{2R}}{\lambda_c^{1+\beta_1}\lambda}\right)\right], \tag{E11}$$

$$I_4(\lambda) \approx \frac{nA_2\lambda_c^{1-\alpha_2}}{\alpha_2-1}, \tag{E12}$$

where ${}_2F_1(q_1, q_2, q_3, z)$ is Gauss hypergeometric function.

2. Approximating the integral expressions for expected strengths

As was the case for the expected degree, the expected strength as a function of the latent parameter λ, see Eq. (63), may be split into four integrals as follows:

$$I_5(\lambda) = n\int_1^{\lambda_c} \frac{A_1\lambda'^{-\alpha_1}}{1+ae^{2R}\left(\frac{\lambda^{-\beta_1}+\lambda'^{-\beta_1}}{\lambda\lambda'}\right)} \times \frac{1}{a(\lambda^{-\beta_1}+\lambda'^{-\beta_1})}d\lambda', \tag{E13}$$

$$I_6(\lambda) = n\int_{\lambda_c}^{\infty} \frac{A_2\lambda'^{-\alpha_2}}{1+ae^{2R}\left(\frac{\lambda^{-\beta_1}+\lambda_c^{\beta_2-\beta_1}\lambda'^{-\beta_2}}{\lambda\lambda'}\right)} \times \frac{1}{a\left(\lambda^{-\beta_1}+\lambda_c^{\beta_2-\beta_1}\lambda'^{-\beta_2}\right)}d\lambda', \tag{E14}$$

$$I_7(\lambda) = n\int_1^{\lambda_c} \frac{A_1\lambda'^{-\alpha_1}}{1+ae^{2R}\left(\frac{\lambda_c^{\beta_2-\beta_1}\lambda^{-\beta_2}+\lambda'^{-\beta_1}}{\lambda\lambda'}\right)} \times \frac{1}{a\left(\lambda_c^{\beta_2-\beta_1}\lambda^{-\beta_2}+\lambda'^{-\beta_1}\right)}d\lambda', \tag{E15}$$

$$I_8(\lambda) = n\int_{\lambda_c}^{\infty} \frac{A_2\lambda'^{-\alpha_2}}{1+a\lambda_c^{\beta_2-\beta_1}e^{2R}\left(\frac{\lambda^{-\beta_2}+\lambda'^{-\beta_2}}{\lambda\lambda'}\right)} \times \frac{1}{a\lambda_c^{\beta_2-\beta_1}(\lambda^{-\beta_2}+\lambda'^{-\beta_2})}d\lambda'. \tag{E16}$$

Using arguments similar to those for the expected degree, we may find the following approximations for the integrals above:

$$I_5(\lambda) \approx \frac{nA_1\lambda^{1+\beta_1}}{e^{2R}a^2\beta_1(1+(\lambda_c/\lambda)^{\beta_1})}\left[\frac{\lambda_c^{\beta_1}+\lambda^{\beta_1}}{1+\lambda^{\beta_1}}\right.$$
$$- \lambda_c^{2+\beta_1-\alpha_1} + \frac{(2+\beta_1-\alpha_1)\left(\lambda_c^{\beta_1}+\lambda^{\beta_1}\right)}{\alpha_1-2}$$
$$\times\left\{{}_2F_1\left(1, \frac{\alpha_1-2}{\beta_1}, 1+\frac{\alpha_1-2}{\beta_1}, -\lambda^{\beta_1}\right)\right.$$
$$\left.\left. - \lambda_c^{2-\alpha_1}{}_2F_1\left(1, \frac{\alpha_1-2}{\beta_1}, 1+\frac{\alpha_1-2}{\beta_1}, -\left(\frac{\lambda}{\lambda_c}\right)^{\beta_1}\right)\right\}\right], \tag{E17}$$

$$I_6(\lambda) \approx \frac{nA_2\lambda_c^{1-\alpha_2}\lambda^{\beta_1}}{a(\alpha_2-1)}{}_2F_1\left(1, \alpha_2-1, \alpha_2, -\frac{ae^{2R}}{\lambda_c\lambda^{1+\beta_1}}\right), \tag{E18}$$

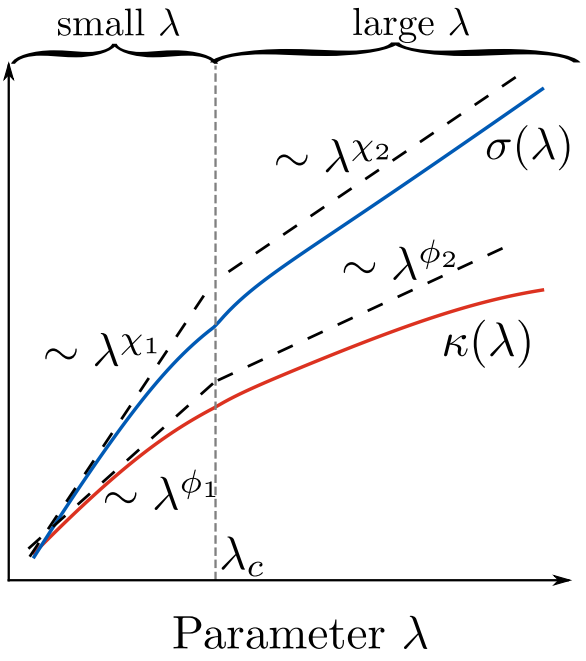

FIG. 10. Illustration of the approximation used to find expression for κ and σ. The scheme is plotted on the log-log scale.

$$I_7(\lambda) \approx \frac{nA_1}{a(\alpha_1-\beta_1-1)} \times \left[{}_2F_1\left(1, \frac{\alpha_1}{1+\beta_1}-1, \frac{\alpha_1}{1+\beta_1}, -\frac{ae^{2R}}{\lambda}\right) - \lambda_c^{1+\beta_1-\alpha_1} {}_2F_1\left(1, \frac{\alpha_1}{1+\beta_1}-1, \frac{\alpha_1}{1+\beta_1}, -\frac{ae^{2R}}{\lambda_c^{1+\beta_1}\lambda}\right)\right], \tag{E19}$$

$$I_8(\lambda) \approx \frac{nA_2\lambda_c^{1+\beta_1-\alpha_2-\beta_2}\lambda^{\beta_2}}{a(\alpha_2-1)} \times {}_2F_1\left(1, \frac{\alpha_2-1}{\beta_2}, 1+\frac{\alpha_2-1}{\beta_2}, -\left(\frac{\lambda}{\lambda_c}\right)^{\beta_2}\right). \tag{E20}$$

3. Approximating the behavior of the integrals with power-law scalings

Unfortunately, it is hard to extract anything about the behavior of the $\kappa(\lambda)$ or $\sigma(\lambda)$ from the approximations that we obtained above. However, numerical evaluation of the expressions above, shows that the expected degree $\kappa(\lambda)$ scales roughly as some power of the latent parameter λ, where the scaling changes its exponent around the λ_c point, as shown in Fig. 10. Similarly, the expected strength $\sigma(\lambda)$ scales roughly as some different power of the latent parameter λ, also changing the scaling exponent around the λ_c. We share the MATHEMATICA notebook `kappa-sigma-approximations.nb` to plot the approximated functions $\kappa(\lambda)$, $\sigma(\lambda)$ for various values of the model parameters at the GitHub repository [49].

We therefore seek for an approximation of the expected degree function in the double power-law form that changes its exponent at the constant λ_c, i.e.,

$$\kappa(\lambda) \sim \begin{cases} \lambda^{\phi_1}, & 1 < \lambda \leqslant \lambda_c, \\ \lambda^{\phi_2}, & \lambda_c < \lambda < \infty. \end{cases} \tag{E21}$$

Similarly, for the expected strength we have

$$\sigma(\lambda) \sim \begin{cases} \lambda^{\chi_1}, & 1 < \lambda \leqslant \lambda_c, \\ \lambda^{\chi_2}, & \lambda_c < \lambda < \infty. \end{cases} \tag{E22}$$

With this ansatz, we now proceed to investigate how the scaling exponents ϕ_1, ϕ_2, χ_1, χ_2 behave as functions of the model parameters: α_1, α_2, β_1, β_2, γ, and η.

First, we note that the scaling exponents ϕ_1, ϕ_2, χ_1, χ_2 along with the model parameters α_1, α_2, β_1, β_2 should be related to the exponents γ and $\delta = 1 + \frac{\gamma-1}{\eta}$ of the degree and strength power-law distributions. Indeed, given the distribution of the latent parameter λ from Eq. (72) and the κ scaling from the Eq. (E21), we find that κ is distributed according to $\rho(\kappa) \sim \kappa^{-(1+(\alpha_1-1)/\phi_1)}$ for $\lambda \leqslant \lambda_c$, and $\rho(\kappa) \sim \kappa^{-(1+(\alpha_2-1)/\phi_2)}$ for $\lambda > \lambda_c$. Similarly, using the Eqs. (72) and (E22), the expected strength σ is distributed according to $\rho(\sigma) \sim \sigma^{-(1+(\alpha_1-1)/\chi_1)}$ for $\lambda \leqslant \lambda_c$, and $\rho(\sigma) \sim \sigma^{-(1+(\alpha_2-1)/\chi_2)}$ for $\lambda > \lambda_c$. As we require that $\rho(\kappa) \sim \kappa^{-\gamma}$ and $\rho(\sigma) \sim \sigma^{-(1+(\gamma-1)/\eta)}$, the model parameters should satisfy:

$$\phi_1 = (\alpha_1 - 1)/(\gamma - 1), \tag{E23}$$

$$\phi_2 = (\alpha_2 - 1)/(\gamma - 1), \tag{E24}$$

$$\chi_1 = \eta\phi_1, \tag{E25}$$

$$\chi_2 = \eta\phi_2. \tag{E26}$$

With these equations, we analyze how the scaling exponents $\phi_1, \phi_2, \chi_1, \chi_2$ depend on the parameters of the model.

4. Analysis of the ϕ_1 scaling exponent

We start by analyzing the scaling of $\kappa \sim \lambda^{\phi_1}$ for the small values of λ, i.e., $\lambda \leqslant \lambda_c$. While it is possible to get the scaling of hypergeometric functions from Eqs. (E9) and (E10), we note that the resulting scaling of κ as a function of λ is given by the sum of λ terms raised to different powers. Moreover, coefficients in front of these terms may change their signs depending on the choice of the parameters α_1, α_2, β_1, β_2. In general, for *any* combination of the model parameters, it is impossible to approximate this sum of power terms as a single power of λ, e.g., by extracting leading order terms.

To circumvent this issue, we find an approximation to the ϕ_1 scaling using the observations from the MLE inference of the WSCM latent parameters introduced in Sec. IV. More precisely, we first infer the nodes' λ and μ parameters given degree and strength sequences with predefined exponents γ and η. Second, we obtain the estimate of the ϕ_1 by linear fitting of the $\kappa(\lambda)$ function on the log-log scale. We observe that for various input values of γ and η, the ϕ_1 scaling exponent is very close to η. We therefore assume that $\phi_1 \approx \eta$, and, given Eq. (E23), the α_1 parameter is set to

$$\alpha_1 \approx 1 + \eta(\gamma - 1). \tag{E27}$$

5. Analysis of the χ_1 scaling exponent

The approximated expression for the expected strength $\sigma(\lambda)$ in the small regime, i.e., $\lambda \leqslant \lambda_c$, is given by Eqs. (E17)

and (E18). Although it is again possible to obtain $\sigma(\lambda)$ scaling in terms of series of various λ power-terms, it is hard to extract a single power scaling that approximates the sum of these power-terms for all parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$. However, it is possible to relate the χ_1 scaling exponent to the β_1 parameter using the following considerations. From the MLE inference of the WSCM latent parameters as in the case of the ϕ_1 scaling exponent, we observe that the resulting empirical model parameter β_1 is linearly related to the $\eta - 1$, and the coefficient between the two depends only on the input parameter γ, i.e., $\beta_1 \approx u(\gamma)(\eta - 1)$, where $u(\gamma)$ is a function of γ. By numerical fitting, we find that $u(\gamma)$ behaves approximately as $u(\gamma) \approx (\gamma - \frac{\gamma-2}{\gamma})$. Thus

$$\beta_1 \approx \left(\gamma - \frac{\gamma-2}{\gamma}\right)(\eta - 1), \tag{E28}$$

which yields $\beta_1(\eta+1) \approx (\gamma - (\gamma-2)/\gamma)(\eta^2 - 1)$, so that $\eta^2 \approx 1 + \frac{(\eta+1)\beta_1}{\gamma-(\gamma-2)/\gamma}$. Moreover, given Eq. (E25) and the fact that ϕ_1 may be approximated as $\phi_1 \approx \eta$, the resulting exponent χ_1 should be approximately equal to $\chi_1 \approx \eta^2$. This means that $\chi_1 \approx 1 + \frac{(\eta+1)\beta_1}{\gamma-(\gamma-2)/\gamma}$.

6. Analysis of the ϕ_2 scaling exponent

The scaling of $\kappa(\lambda) \sim \lambda^{\phi_2}$ for large $\lambda > \lambda_c$ is encoded in the two integrals from Eqs. (E11) and (E12). The approximation in Eq. (E12) does not have any λ-dependent terms, so it does not contribute to the scaling. We may approximate the scaling of the hypergeometric functions appearing in Eq. (E11) for $\lambda \to \lambda_c$. In this case, the argument of the *second* hypergeometric function approaches -1, therefore, the function approaches a constant and does not contribute to the scaling. The *first* hypergeometric function may be approximated as follows, assuming that its argument is large:

$$\begin{aligned} &{}_2F_1\left(1, \frac{\alpha_1-1}{1+\beta_1}, 1 + \frac{\alpha_1-1}{1+\beta_1}, -\frac{ae^{2R}}{\lambda}\right) \\ &\approx \frac{\alpha_1 - 1}{\alpha_1 - 2 - \beta_1}\left(\frac{\lambda}{ae^{2R}}\right) \\ &\quad + \frac{\pi(\alpha_1-1)}{(1+\beta_1)\sin\left(\frac{\pi(\alpha_1-1)}{1+\beta_1}\right)}\left(\frac{\lambda}{ae^{2R}}\right)^{\frac{\alpha_1-1}{1+\beta_1}}. \end{aligned} \tag{E29}$$

Additionally, in the case when $\lambda \to \infty$, the $\kappa(\lambda)$ function should saturate to the maximum possible degree $n-1$, therefore, we only consider the $\kappa(\lambda)$ scaling near the left boundary of this regime, i.e., λ_c. The signs of the coefficients in front of the two λ-dependent terms may be of different signs, so it is again impossible to extract a single-exponent scaling from the expression above. However, from the WSCM MLE-inferred latent parameters as was done in the case of ϕ_1, χ_1 scalings, we observe that for large values of $\eta \gg 1$, the scaling exponent $\phi_2 \to \frac{\alpha_1-1}{1+\beta_1}$. We therefore seek a scaling exponent ϕ_2 in the form $\phi_2 \approx \psi(\gamma,\eta)\frac{\alpha_1-1}{1+\beta_1}$. We know that $\phi_2 \to 1$ when $\eta \to 1$ to be compatible with the unweighted HSCM, and we know that in this limit $\beta_1 \to 0$, $\alpha_1 \to \gamma$. Therefore there should be a prefactor of $\frac{1}{\gamma-1}$ in front of the $\frac{\alpha_1-1}{1+\beta_1}$ to guarantee that $\phi_2 \to 1$ in the HSCM limit. Additionally, as $\phi_2 \to \frac{\alpha_1-1}{1+\beta_1}$ in the large η limit, we need to compensate for this prefactor. Given these requirements, we find from the numerical fitting procedure that the following form of $\psi(\gamma,\eta)$ fits well the ϕ_2 exponent for various γ, η parameters: $\psi(\gamma,\eta) = \frac{1}{\gamma-1}[1 + (\gamma-2)(1-1/\eta)]$. Then the ϕ_2 scaling exponent is: $\phi_2 = \frac{\alpha_1-1}{1+\beta_1}\frac{1}{\gamma-1}[1 + (\gamma-2)(1-1/\eta)]$. Therefore the model parameter α_2 should be selected as follows:

$$\alpha_2 \approx 1 + \frac{\alpha_1 - 1}{1+\beta_1}(1 + (\gamma-2)(1-1/\eta)). \tag{E30}$$

7. Analysis of the χ_2 scaling exponent

The scaling of $\sigma(\lambda)$ is given in Eq. (E19) and (E20). The hypergeometric functions from Eq. (E19) will not contribute to the scaling in the large λ limit, as their arguments would approach -1, so we only expect to see an effect of these functions near the left boundary of the region, λ_c. Conversely, the hypergeometric function from Eq. (E20) is the main contributor to the scaling for large λ. We note that we may not neglect the behavior of the $\sigma(\lambda)$ for large λ, as, in general, strengths are not bounded for a weighted network, unlike degrees that have to be at most $n-1$. Using the large argument approximation for hypergeometric functions as before, we obtain for Eq. (E20):

$$\begin{aligned} &{}_2F_1\left(1, \frac{\alpha_2-1}{\beta_2}, 1 + \frac{\alpha_2-1}{\beta_2}, -\left(\frac{\lambda}{\lambda_c}\right)^{\beta_2}\right) \\ &\approx \frac{\alpha_2-1}{\alpha_2-1-\beta_2}\left(\frac{\lambda}{\lambda_c}\right)^{-\beta_2} + \frac{\pi(\alpha_2-1)}{\beta_2\sin\left(\frac{\pi(\alpha_2-1)}{\beta_2}\right)}\left(\frac{\lambda}{\lambda_c}\right)^{1-\alpha_2}. \end{aligned} \tag{E31}$$

Given the λ^{β_2} prefactor in Eq. (E20), we observe that the only possible resulting scaling may be $\sim\lambda^{1+\beta_2-\alpha_2}$. However, we note that the large-argument approximation used above is only valid for $\beta_1 > 0$ and $\beta_2 > 0$. Instead, when $\eta \to 1$, we expect to recover the unweighted HSCM behavior, effectively giving us the same scaling for the χ_2 as for the ϕ_2 exponent. Thus, we will set β_2 to 0 in this limit, and use the above approximation otherwise. This defines our choice for the β_2 parameter:

$$\beta_2 \approx \begin{cases} 0, & \eta = 1, \\ \alpha_2 - 1 + \frac{\eta(\alpha_2-1)}{\gamma-1}, & \eta > 1. \end{cases} \tag{E32}$$

8. Finding the (R, a) solution

As explained in Sec. V, we find the R, a parameters of the WHSCM by numerically solving Eqs. (87) and (88) for a fixed λ_0. For the solver from our code package [49], we set λ_0 such that $\kappa(\lambda_0) = \bar{k}$. This is done to prevent the solver from finding an (R, a) solution that corresponds to a low-degree region ($k < \bar{k}$) of the $\bar{s}(k)$ scaling curve that may not exhibit a clean power-law scaling and distort the resulting baseline for the strength-degree correlation curve. For each solver round, we iteratively update the λ_0 value corresponding to $\kappa(\lambda_0) = \bar{k}$ with the current solver's guess of (R, a).

APPENDIX F: Behavior of the WHSCM in the $\eta \to 1$ limit

With the choice of model parameters above and setting $\eta = 1$, we have the following distribution of the latent parameter λ:

$$\rho(\lambda) = (\alpha - 1)\lambda^{-\alpha}, \tag{F1}$$

where $\lambda \in (1, \infty)$ and $\alpha = \alpha_1 = \alpha_2 = \gamma$. This gives the expressions for $\kappa(\lambda)$ and $\sigma(\lambda)$:

$$\kappa(\lambda) = n \int_1^\infty \frac{(\gamma - 1)\lambda'^{-\gamma}}{1 + \frac{2ae^{2R}}{\lambda\lambda'}} d\lambda', \tag{F2}$$

$$\sigma(\lambda) = n \int_1^\infty \frac{(\gamma - 1)\lambda'^{-\gamma}}{1 + \frac{2ae^{2R}}{\lambda\lambda'}} \frac{1}{2a} d\lambda'. \tag{F3}$$

The integral for $\kappa(\lambda)$ may be evaluated directly:

$$\kappa(\lambda) = n\, {}_2F_1\left(1, \gamma - 1, \gamma, -\frac{2ae^{2R}}{\lambda}\right). \tag{F4}$$

This function scales linearly with λ for large enough $ae^{2R} \gg 1$, therefore, the expected degree $\kappa(\lambda) \sim \lambda$. From the above expressions, we also see that $\sigma(\lambda) = \frac{1}{2a}\kappa(\lambda)$. Thus, the constant a may be easily found given the model input parameter σ_0, $a = \frac{1}{2\sigma_0}$. Moreover, the average degree $\bar{k}$ defined by Eq. (87) now reads

$$\bar{k} = n(\gamma - 1)^2 \Phi(-2ae^{2R}, 2, \gamma - 1), \tag{F5}$$

where $\Phi(z, q_1, q_2)$ is the Lerch transcendent function. These equations allows to solve for the model parameter R numerically, given a target average degree $\bar{k}$. The considered limiting behavior is included as a special case in the code package for generation of WHSCM networks [49].

[1] M. E. J. Newman, *Networks*, 2nd ed. (Oxford University Press, Oxford, 2018).

[2] A.-L. Barabási, *Network Science* (Cambridge University Press, Cambridge, 2016).

[3] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, The architecture of complex weighted networks, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004).

[4] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, The role of the airline transportation network in the prediction and predictability of global epidemics, Proc. Natl. Acad. Sci. U.S.A. **103**, 2015 (2006).

[5] D. Brockmann and D. Helbing, The hidden geometry of complex, network-driven contagion phenomena, Science **342**, 1337 (2013).

[6] J. Park and M. E. J. Newman, Statistical mechanics of networks, Phys. Rev. E **70**, 066117 (2004).

[7] G. Bianconi, The entropy of randomized network ensembles, Europhys. Lett. **81**, 28005 (2007).

[8] D. Garlaschelli and M. I. Loffredo, Maximum likelihood: Extracting unbiased information from complex networks, Phys. Rev. E **78**, 015101(R) (2008).

[9] D. Garlaschelli and M. I. Loffredo, Generalized Bose-Fermi Statistics and Structural Correlations in Weighted Networks, Phys. Rev. Lett. **102**, 038701 (2009).

[10] T. Squartini and D. Garlaschelli, Analytical maximum-likelihood method to detect patterns in real networks, New J. Phys. **13**, 083001 (2011).

[11] O. Sagarra, C. J. P. Vicente, and A. Díaz-Guilera, Statistical mechanics of multiedge networks, Phys. Rev. E **88**, 062806 (2013).

[12] R. Mastrandrea, T. Squartini, G. Fagiolo, and D. Garlaschelli, Enhanced reconstruction of weighted networks from strengths and degrees, New J. Phys. **16**, 043022 (2014).

[13] T. Squartini, R. Mastrandrea, and D. Garlaschelli, Unbiased sampling of network ensembles, New J. Phys. **17**, 023052 (2015).

[14] P. van der Hoorn, G. Lippner, and D. Krioukov, Sparse maximum-entropy random graphs with a given power-law degree distribution, J. Stat. Phys. **173**, 806 (2018).

[15] A. Gabrielli, R. Mastrandrea, G. Caldarelli, and G. Cimini, Grand canonical ensemble of weighted networks, Phys. Rev. E **99**, 030301(R) (2019).

[16] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, The statistical physics of real-world networks, Nat. Rev. Phys. **1**, 58 (2019).

[17] A. Vazquez, R. Dobrin, D. Sergi, J.-P. Eckmann, Z. N. Oltvai, and A.-L. Barabási, The topological relationship between the large-scale attributes and local interaction patterns of complex networks, Proc. Natl. Acad. Sci. U.S.A. **101**, 17940 (2004).

[18] D. V. Foster, J. G. Foster, P. Grassberger, and M. Paczuski, Clustering drives assortativity and community structure in ensembles of networks, Phys. Rev. E **84**, 066117 (2011).

[19] P. Colomer-de Simón, M. A. Serrano, M. G. Beiró, J. I. Alvarez-Hamelin, and M. Boguná, Deciphering the global organization of clustering in real complex networks, Sci. Rep. **3**, 2517 (2013).

[20] C. Orsini, M. M. Dankulov, P. Colomer-de Simón, A. Jamakovic, P. Mahadevan, A. Vahdat, K. E. Bassler, Z. Toroczkai, M. Boguñá, G. Caldarelli *et al.*, Quantifying randomness in real networks, Nat. Commun. **6**, 8627 (2015).

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

[22] J. Park and M. E. J. Newman, Origin of degree correlations in the Internet and other networks, Phys. Rev. E **68**, 026112 (2003).

[23] R. Guimera, M. Sales-Pardo, and L. A. Amaral, Classes of complex networks defined by role-to-role connectivity profiles, Nat. Phys. **3**, 63 (2007).

[24] T. Opsahl, V. Colizza, P. Panzarasa, and J. J. Ramasco, Prominence and Control: The Weighted Rich-Club Effect, Phys. Rev. Lett. **101**, 168702 (2008).

[25] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, Detecting rich-club ordering in complex networks, Nat. Phys. **2**, 110 (2006).

[26] L. A. N. Amaral and R. Guimera, Lies, damned lies and statistics, Nat. Phys. **2**, 75 (2006).

[27] S. Zhou and R. J. Mondragón, The rich-club phenomenon in the Internet topology, IEEE Commun. Lett. **8**, 180 (2004).

[28] R. Solomonoff and A. Rapoport, Connectivity of random nets, Bull. Math. Biophys. **13**, 107 (1951).

[29] E. N. Gilbert, Random graphs, Ann. Math. Stat. **30**, 1141 (1959).

[30] P. Erdős and A. Rényi, On random graphs, Publicationes Mathematicae Debrecen **6**, 290 (1959).

[31] P. Erdős and A. Rényi, On the evolution of random graphs, Publications of the Mathematical Institute of the Hungarian Academy of Sciences **5**, 17 (1960).

[32] I. Voitalov, P. van der Hoorn, R. van der Hofstad, and D. Krioukov, Scale-free networks well done, Phys. Rev. Res. **1**, 033034 (2019).

[33] E. A. Bender and E. R. Canfield, The asymptotic number of labeled graphs with given degree sequences, J. Comb. Theory, Ser. A **24**, 296 (1978).

[34] M. Molloy and B. Reed, A critical point for random graphs with a given degree sequence, Random Struct. Algorithms **6**, 161 (1995).

[35] F. Chung and L. Lu, Connected components in random graphs with given expected degree sequences, Ann. Comb. **6**, 125 (2002).

[36] F. Chung and L. Lu, The average distances in random graphs with given expected degrees, Proc. Natl. Acad. Sci. U.S.A. **99**, 15879 (2002).

[37] M. E. J. Newman, Clustering and preferential attachment in growing networks, Phys. Rev. E **64**, 025102(R) (2001).

[38] A. Dhamdhere and C. Dovrolis, Twelve years in the evolution of the Internet ecosystem, IEEE/ACM Trans. Netw. **19**, 1420 (2011).

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

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

[41] K. Anand, D. Krioukov, and G. Bianconi, Entropy distribution and condensation in random networks with a given degree distribution, Phys. Rev. E **89**, 062807 (2014).

[42] M. Á. Serrano and M. Boguñá, in *AIP Conference Proceedings* (American Institute of Physics, Aveiro, Portugal, 2005), Vol. 776, pp. 101–107.

[43] T. Britton, M. Deijfen, and F. Liljeros, A weighted configuration model and inhomogeneous epidemics, J. Stat. Phys. **145**, 1368 (2011).

[44] D. Garlaschelli, S. Battiston, M. Castri, V. D. Servedio, and G. Caldarelli, The scale-free topology of market investments, Physica A **350**, 491 (2005).

[45] G. Bagler, Analysis of the airport network of India as a complex weighted network, Physica A **387**, 2972 (2008).

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

[47] M. Popović, H. Štefančić, and V. Zlatić, Geometric Origin of Scaling in Large Traffic Networks, Phys. Rev. Lett. **109**, 208701 (2012).

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

[49] I. Voitalov, Weighted hypersoft configuration model, https://github.com/ivanvoitalov/whscm (2020).

[50] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).

[51] M. E. J. Newman, The structure and function of complex networks, SIAM Rev. **45**, 167 (2003).

[52] A. E. Motter and Z. Toroczkai, Introduction: Optimization in networks, Chaos **17**, 026101 (2007).

[53] Z. Burda, A. Krzywicki, O. Martin, and M. Zagorski, Motifs emerge from function in model gene regulatory networks, Proc. Natl. Acad. Sci. U.S.A. **108**, 17263 (2011).

[54] K. Anand and G. Bianconi, Entropy measures for networks: Toward an information theory of complex topologies, Phys. Rev. E **80**, 045102(R) (2009).

[55] E. T. Jaynes, Information theory and statistical mechanics, Phys. Rev. **106**, 620 (1957).

[56] S. Janson, Asymptotic equivalence and contiguity of some random graphs, Random Struct. Algor. **36**, 26 (2010).

[57] T. Britton, M. Deijfen, and A. Martin-Löf, Generating simple random graphs with prescribed degree distribution, J. Stat. Phys. **124**, 1377 (2006).

[58] B. Bollobás, S. Janson, and O. Riordan, The phase transition in inhomogeneous random graphs, Random Struct. Algor. **31**, 3 (2007).

[59] J. Grandell, *Mixed Poisson Processes* (Chapman and Hall, London, 1997).

[60] S. Janson, Graphons, cut norm and distance, couplings and rearrangements, NYJM Monogr. **4** (2013).

[61] G. Bianconi, Entropy of network ensembles, Phys. Rev. E **79**, 036114 (2009).

[62] S. Janson, Tail bounds for sums of geometric and exponential variables, Stat. Probab. Lett. **135**, 1 (2018).

[63] F. Biscani and D. Izzo, A parallel global multiobjective framework for optimization: pagmo, J. Open Source Softw. **5**, 2338 (2019).

[64] M. Barthélemy, B. Gondran, and E. Guichard, Spatial structure of the Internet traffic, Physica A **319**, 633 (2003).

[65] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, Characterization and modeling of weighted networks, Physica A **346**, 34 (2005).

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

[67] V. Latora, V. Nicosia, and G. Russo, *Complex Networks: Principles, Methods and Applications* (Cambridge University Press, Cambridge, UK, 2017).

[68] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, The large-scale organization of metabolic networks, Nature (London) **407**, 651 (2000).

[69] J. Duch and A. Arenas, Community detection in complex networks using extremal optimization, Phys. Rev. E **72**, 027104 (2005).

[70] J. Kunegis, in *Proceedings of the 22nd International Conference on World Wide Web* (Association for Computing Machinery (ACM), New York, NY, 2013), pp. 1343–1350.

[71] V. Batagelj and A. Mrvar, Pajek data: Collaboration network in computational geometry, http://vlado.fmf.uni-lj.si/pub/networks/data/collab/geom.htm.

[72] C. Harrison and C. Römhild, Bible cross-references, http://www.chrisharrison.net/index.php/Visualizations/BibleViz.

[73] J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, and J. Kertesz, Generalizations of the clustering coefficient to weighted complex networks, Phys. Rev. E **75**, 027105 (2007).

[74] S.-H. Yook, H. Jeong, A.-L. Barabási, and Y. Tu, Weighted Evolving Networks, Phys. Rev. Lett. **86**, 5835 (2001).

[75] A. Barrat, M. Barthélemy, and A. Vespignani, Weighted Evolving Networks: Coupling Topology and Weight Dynamics, Phys. Rev. Lett. **92**, 228701 (2004).

[76] A. Barrat, M. Barthélemy, and A. Vespignani, Modeling the evolution of weighted networks, Phys. Rev. E **70**, 066149 (2004).

[77] G. Bianconi, Emergence of weight-topology correlations in complex scale-free networks, Europhys. Lett. **71**, 1029 (2005).

[78] T. Antal and P. L. Krapivsky, Weight-driven growing networks, Phys. Rev. E **71**, 026103 (2005).

[79] W.-X. Wang, B. Hu, T. Zhou, B.-H. Wang, and Y.-B. Xie, Mutual selection model for weighted networks, Phys. Rev. E **72**, 046140 (2005).

[80] Q. Ou, Y.-D. Jin, T. Zhou, B.-H. Wang, and B.-Q. Yin, Power-law strength-degree correlation from resource-allocation dynamics on weighted networks, Phys. Rev. E **75**, 021102 (2007).

[81] T. Tanaka and T. Aoyagi, Weighted scale-free networks with variable power-law exponents, Physica D **237**, 898 (2008).

[82] G. Krings, F. Calabrese, C. Ratti, and V. D. Blondel, Urban gravity: A model for inter-city telecommunication flows, J. Stat. Mech.: Theory Exp. (2009) L07003.

[83] G. Fagiolo, The international-trade network: Gravity equations and topological properties, J. Econ. Interact. Coord. **5**, 1 (2010).

[84] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, A universal model for mobility and migration patterns, Nature **484**, 96 (2012).

[85] M. E. Newman, The structure of scientific collaboration networks, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).

[86] L. E. Da Rocha, Structural evolution of the Brazilian airport network, J. Stat. Mech.: Theory Exp. (2009) P04020.

[87] J. Zhang, X.-B. Cao, W.-B. Du, and K.-Q. Cai, Evolution of Chinese airport network, Physica A **389**, 3922 (2010).

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

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

[90] R. Aldecoa, C. Orsini, and D. Krioukov, Hyperbolic graph generator, Comput. Phys. Commun. **196**, 492 (2015).